

Dokumentvorlage, Version vom 20.01.2011

**Dossier zur Nutzenbewertung
gemäß § 35a SGB V**

Teriflunomid (AUBAGIO®)

sanofi-aventis groupe

Modul 4 A

Schubförmig-remittierende Multiple Sklerose

Medizinischer Nutzen und
medizinischer Zusatznutzen,
Patientengruppen mit therapeutisch
bedeutsamem Zusatznutzen

Stand: 26.09.2013

Inhaltsverzeichnis

	Seite
Tabellenverzeichnis	7
Abbildungsverzeichnis	17
Abkürzungsverzeichnis	20
4 Modul 4 – allgemeine Informationen	22
4.1 Zusammenfassung der Inhalte von Modul 4.....	23
4.1.1 Hintergrund.....	23
4.1.2 Fragestellung.....	23
4.1.3 Datenquellen.....	24
4.1.4 Ein-/Ausschlusskriterien für Studien.....	24
4.1.5 Methoden zur Bewertung der Aussagekraft der Nachweise und zur Synthese von Ergebnissen.....	25
4.1.6 Vorteil einer oralen gegenüber einer parenteralen Applikationsform.....	26
4.1.7 Ergebnisse zum medizinischen Nutzen.....	26
4.1.7.1 Mortalität.....	26
4.1.7.2 Morbidität.....	27
4.1.7.3 Lebensqualität.....	27
4.1.7.4 Patientenzufriedenheit.....	27
4.1.7.5 Unerwünschte Ereignisse.....	27
4.1.7.6 Fazit zum medizinischen Nutzen von 14 mg Teriflunomid.....	28
4.1.8 Medizinischer Zusatznutzen von 14 mg Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie Interferon beta-1a, 44 µg, s. c., (Rebif® 44 µg).....	28
4.1.8.1 TENERE.....	28
4.1.8.1.1 Mortalität.....	30
4.1.8.1.2 Morbidität.....	30
4.1.8.1.3 Lebensqualität.....	30
4.1.8.1.4 Patientenzufriedenheit.....	30
4.1.8.1.5 Unerwünschte Ereignisse.....	31
4.1.8.2 Indirekte Vergleiche.....	32
4.1.8.2.1 Mortalität.....	35
4.1.8.2.2 Morbidität.....	35
4.1.8.2.3 Lebensqualität.....	36
4.1.8.2.4 Patientenzufriedenheit.....	36
4.1.8.2.5 Unerwünschte Ereignisse.....	36
4.1.8.3 Fazit zum medizinischen Zusatznutzen von 14 mg Teriflunomid.....	37
4.1.9 Ausmaß des Zusatznutzens und Beleglage.....	37
4.1.9.1 Beschreibung des Zusatznutzens auf Endpunktebene.....	37
4.1.9.1.1 Mortalität.....	37
4.1.9.1.2 Morbidität.....	37
4.1.9.1.3 Lebensqualität.....	38
4.1.9.1.4 Patientenzufriedenheit.....	38
4.1.9.1.5 Unerwünschte Ereignisse.....	38
4.1.9.2 Ausmaß des Zusatznutzens - Fazit.....	39

4.1.9.3	Schlussfolgerungen zum Zusatznutzen und zum therapeutisch bedeutsamen Zusatznutzen	40
4.2	Methodik	41
4.2.1	Fragestellung	41
4.2.2	Kriterien für den Einschluss von Studien in die Nutzenbewertung	42
4.2.2.1	Grundlagen zur Festlegung von Ein- und Ausschlusskriterien	42
4.2.2.1.1	Patientenpopulation	43
4.2.2.1.1.1	Diagnosekriterien	43
4.2.2.1.1.2	Krankheitsschwere	43
4.2.2.1.1.3	Verlaufsform	43
4.2.2.1.1.4	Alter	43
4.2.2.1.2	Intervention	43
4.2.2.1.3	Vergleichstherapie (Komparator/Comparison)	44
4.2.2.1.4	Endpunkte (Zielgrößen/Outcomes)	44
4.2.2.1.5	Studientypen	45
4.2.2.1.6	Studiendauer	45
4.2.2.2	Ein- und Ausschlusskriterien für randomisierte kontrollierte Studien mit dem zu bewertenden Arzneimittel und den indirekten Vergleich	46
4.2.2.2.1	Ein- und Ausschlusskriterien für Studien mit Teriflunomid als Intervention	46
4.2.2.2.2	Ein- und Ausschlusskriterien für Studien mit Interferon beta-1a, 44 µg, subkutan, 3 x wöchentlich, (Rebif®) als Intervention (für indirekte Vergleiche)	48
4.2.3	Informationsbeschaffung	51
4.2.3.1	Studien des pharmazeutischen Unternehmers	51
4.2.3.2	Bibliografische Literaturrecherche	51
4.2.3.3	Suche in Studienregistern	52
4.2.3.4	Selektion relevanter Studien	53
4.2.4	Bewertung der Aussagekraft der Nachweise	53
4.2.5	Informationssynthese und -analyse	55
4.2.5.1	Beschreibung des Designs und der Methodik der eingeschlossenen Studien	55
4.2.5.2	Gegenüberstellung der Ergebnisse der Einzelstudien	55
4.2.5.2.1	Patientencharakteristika	56
4.2.5.2.2	Endpunkte	56
4.2.5.2.2.1	Mortalität	57
4.2.5.2.2.2	Morbidität	57
4.2.5.2.2.3	Gesundheitsbezogene Lebensqualität	59
4.2.5.2.2.4	Patientenzufriedenheit	59
4.2.5.2.2.5	Nebenwirkungen (Unerwünschte Ereignisse)	60
4.2.5.2.3	Metaanalysen	63
4.2.5.2.4	Sensitivitätsanalysen	66
4.2.5.2.4.1	Als sekundäre Analysen der patientenrelevanten Endpunkte	66
4.2.5.2.4.2	Im Rahmen von Metaanalysen und indirekten Vergleichen	67
4.2.5.2.5	Subgruppenmerkmale und andere Effektmodifikatoren	67
4.2.5.2.6	Indirekte Vergleiche	70
4.3	Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen	74

4.3.1	Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel.....	74
4.3.1.1	Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel	74
4.3.1.1.1	Studien des pharmazeutischen Unternehmers.....	74
4.3.1.1.2	Studien aus der bibliografischen Literaturrecherche.....	78
4.3.1.1.3	Studien aus der Suche in Studienregistern	81
4.3.1.1.4	Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel.....	82
4.3.1.2	Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel.....	83
4.3.1.2.1	Studiendesign und Studienpopulationen	83
4.3.1.2.2	Verzerrungspotenzial auf Studienebene.....	89
4.3.1.3	Ergebnisse aus randomisierten kontrollierten Studien.....	90
4.3.1.3.1	Schübe	97
4.3.1.3.1.1	Jährliche Schubrate – RCT	98
4.3.1.3.1.2	Zeit bis zum Schub und Patienten mit bestätigtem Schub - RCT	105
4.3.1.3.1.3	Schubfreie Patienten - RCT.....	112
4.3.1.3.1.4	Zusätzliche Auswertungen bezüglich der Schübe, die für die TEMSO-Studie (und teilweise für die TOWER-Studie) durchgeführt worden sind.....	116
4.3.1.3.1.5	Metaanalysen der schubbezogenen Endpunkte	119
4.3.1.3.1.6	Gesamtergebnis zu schubbezogenen Endpunkten.....	121
4.3.1.3.2	Behinderungsprogression	122
4.3.1.3.2.1	Zeit bis zur Behinderungsprogression (inklusive Anteil Pntien mit Behinderungsprogression zu einem bestimmten Zeitpunkt) - RCT	122
4.3.1.3.2.2	Mittlere Änderung des MSFC-Z-Scores - RCT	132
4.3.1.3.2.3	Mittlere Änderung des EDSS-Scores - RCT	136
4.3.1.3.2.4	Metaanalysen zur Endpunkten, die sich auf die Behinderungsprogression beziehen	141
4.3.1.3.2.5	Zusammenfassung der Ergebnisse zu Endpunkten, die sich auf die Behinderungsprogression beziehen	144
4.3.1.3.3	Therapieversagen (Time to Failure) – RCT	144
4.3.1.3.4	Fatigue – RCT	150
4.3.1.3.5	Krankheitslast (Burden of Disease (MRT)) – RCT	160
4.3.1.3.5.1	Krankheitslast (Burden of Disease).....	163
4.3.1.3.5.2	Übersicht weiterer MRT-Parameter	164
4.3.1.3.5.3	Aktive Läsionen („unique active lesions“).....	167
4.3.1.3.6	Patientenzufriedenheit – RCT	168
4.3.1.3.7	Lebensqualität (SF-36)– RCT	174
4.3.1.3.8	Lebensqualität (EQ-5D) – RCT	178
4.3.1.3.9	Mortalität – RCT	181
4.3.1.3.10	Unerwünschte Ereignisse (Gesamtraten AEs, SAEs, Behandlungsabbrüche) – RCT	185
4.3.1.3.10.1	Gesamtrate unerwünschter Ereignisse.....	191
4.3.1.3.10.2	Gesamtrate schwerwiegender unerwünschter Ereignisse.....	192

4.3.1.3.10.3 Gesamtrate unerwünschter Ereignisse, die zum Behandlungsabbruch führten	193
4.3.1.3.10.4 Diskussion der Gesamtraten unerwünschter Ereignisse	194
4.3.1.3.10.5 Metaanalysen zu den Gesamtraten unerwünschter Ereignisse	195
4.3.1.3.11 Unerwünschte Ereignisse (Unerwünschte Ereignisse von besonderem Interesse) – RCT	197
4.3.1.3.11.1 Reaktionen an der Injektionsstelle.....	203
4.3.1.3.11.2 Grippeähnliche Symptome	204
4.3.1.3.11.3 Neutropenie	208
4.3.1.3.11.4 Alopezie	211
4.3.1.3.11.5 Überempfindlichkeit/Hautreaktionen	214
4.3.1.3.11.6 Hepatische unerwünschte Ereignisse (inklusive Erhöhung der ALT-Werte)	216
4.3.1.3.11.7 Gastrointestinale Beschwerden.....	222
4.3.1.3.11.8 Infektionen und Infestationen	228
4.3.1.3.11.9 Bluthochdruck, Kopfschmerzen und periphere Neuropathie	231
4.3.1.3.11.10 Zusammenfassung der Ergebnisse zu unerwünschten Ereignissen von besonderem Interesse (AESI)	237
4.3.1.3.12 Subgruppenanalysen – RCT.....	239
4.3.1.3.12.1 Ergebnisse der Subgruppenanalysen der Studie TENERE.....	241
4.3.1.3.12.2 Ergebnisse der Subgruppenanalysen der Studie TEMSO	244
4.3.1.3.12.3 Ergebnisse der Subgruppenanalysen der Studie TOWER.....	249
4.3.1.3.12.4 Zusammenfassende Auswertung der Subgruppenanalysen.....	254
4.3.1.3.13 Zusammenfassung der Ergebnisse aus randomisierten kontrollierten Studien.....	257
4.3.1.3.13.1 Zusammenfassung der Ergebnisse zum medizinischen Nutzen aus randomisierten kontrollierten Studien.....	257
4.3.1.3.13.2 Zusammenfassung der Ergebnisse zum medizinischen Zusatznutzen aus randomisierten kontrollierten Studien.....	263
4.3.2 Weitere Unterlagen.....	269
4.3.2.1 Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien	269
4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche	269
4.3.2.1.1.1 Studien des pharmazeutischen Unternehmers	269
4.3.2.1.1.2 Studien aus der bibliografischen Literaturrecherche	274
4.3.2.1.1.3 Studien aus der Suche in Studienregistern	278
4.3.2.1.1.4 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel und der zweckmäßigen Vergleichstherapie.....	280
4.3.2.1.2 Charakteristika der Studien für indirekte Vergleiche.....	280
4.3.2.1.3 Verzerrungspotenzial auf Studienebene.....	288
4.3.2.1.4 Ergebnisse aus indirekten Vergleichen	289
4.3.2.1.4.1 Schübe – indirekte Vergleiche aus RCT.....	292
4.3.2.1.4.2 Behinderungsprogression – indirekte Vergleiche aus RCT	298
4.3.2.1.4.3 Mortalität – indirekte Vergleiche aus RCT	306
4.3.2.1.4.4 Unerwünschte Ereignisse (die zum Behandlungsabbruch führten) – indirekte Vergleiche aus RCT	311
4.3.2.1.4.5 Unerwünschte Ereignisse von besonderem Interesse – indirekte Vergleiche aus RCT.....	317

4.3.2.1.4.6	Zusammenfassung der Ergebnisse der indirekten Vergleiche.....	339
4.3.2.1.4.7	Subgruppenanalysen – indirekte Vergleiche aus RCT	341
4.3.2.2	Nicht randomisierte vergleichende Studien	342
4.3.2.2.1	Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien	342
4.3.2.2.2	Charakteristika der nicht randomisierten vergleichenden Studien.....	342
4.3.2.2.3	Ergebnisse aus nicht randomisierten vergleichenden Studien	343
4.3.2.2.3.1	<Endpunkt xxx> – nicht randomisierte vergleichende Studien.....	343
4.3.2.2.3.2	Subgruppenanalysen – nicht randomisierte vergleichende Studien	344
4.3.2.3	Weitere Untersuchungen.....	344
4.3.2.3.1	Ergebnis der Informationsbeschaffung – weitere Untersuchungen	344
4.3.2.3.2	Charakteristika der weiteren Untersuchungen	345
4.3.2.3.3	Ergebnisse aus weiteren Untersuchungen	345
4.3.2.3.3.1	<Endpunkt xxx> – weitere Untersuchungen	345
4.3.2.3.3.2	Subgruppenanalysen – weitere Untersuchungen	346
4.3.2.4	Zusammenfassung der Ergebnisse aus weiteren Unterlagen	346
4.4	Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens.....	347
4.4.1	Beurteilung der Aussagekraft der Nachweise	347
4.4.1.1	Studienqualität	347
4.4.1.2	Validität der Endpunkte	348
4.4.1.3	Evidenzstufe.....	349
4.4.1.4	Fazit	349
4.4.2	Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß.....	349
4.4.2.1	Beschreibung des Zusatznutzens auf Endpunktebene	350
4.4.2.1.1	Mortalität.....	353
4.4.2.1.2	Morbidität.....	354
4.4.2.1.3	Lebensqualität	354
4.4.2.1.4	Patientenzufriedenheit.....	354
4.4.2.1.5	Unerwünschte Ereignisse	355
4.4.2.2	Ausmaß des Zusatznutzens - Fazit.....	355
4.4.3	Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht	355
4.4.4	Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht – Arzneimittel zur Behandlung eines seltenen Leidens....	356
4.5	Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte	357
4.5.1	Begründung für die Vorlage indirekter Vergleiche.....	357
4.5.2	Begründung für die Vorlage nichtrandomisierter vergleichender Studien und weiterer Untersuchungen.....	358
4.5.3	Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen	358
4.5.4	Verwendung von Surrogatendpunkten	359
4.6	Liste der eingeschlossenen Studien.....	360
4.7	Referenzliste.....	362
	Anhang 4-A : Suchstrategien – bibliografische Literaturrecherche	369
	Anhang 4-B : Suchstrategien – Suche in Studienregistern.....	376

**Anhang 4-C : Liste der im Volltext gesichteten und ausgeschlossenen Studien mit
Ausschlussgrund 380**

Anhang 4-D : Liste der abgebrochenen Studien 382

Anhang 4-E : Liste der laufenden Studien..... 383

Anhang 4-F : Methodik der eingeschlossenen Studien – RCT..... 385

Anhang 4-G : Bewertungsbögen zur Einschätzung von Verzerrungsaspekten 440

Tabellenverzeichnis

	Seite
Tabelle 4-1: Einzelergebnisse aus denen sich der Zusatznutzen ableitet (aufgeschlüsselt nach Endpunkten) - direkter Vergleich (TENERE)	28
Tabelle 4-2: Einzelergebnisse aus denen sich der Zusatznutzen ableitet (aufgeschlüsselt nach Endpunkten) - indirekte Vergleiche.....	33
Tabelle 4-3: Einzelergebnisse aus denen sich der Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie ableitet (aufgeschlüsselt nach Endpunkten).....	39
Tabelle 4-4: Schlussfolgerung zum therapeutisch bedeutsamen Zusatznutzen	40
Tabelle 4-5: Patientenrelevante Endpunkte für die Bewertung eines krankheitsmodifizierenden Wirkstoffes in der MS-Therapie	44
Tabelle 4-6: Einschlusskriterien für Studien mit Teriflunomid als Intervention	46
Tabelle 4-7: Ausschlusskriterien für Studien mit Teriflunomid als Intervention	47
Tabelle 4-8: Einschlusskriterien für Interferon beta-1a, 44 µg-Studien für den indirekten Vergleich	49
Tabelle 4-9: Ausschlusskriterien für Interferon beta-1a, 44 µg-Studien für den indirekten Vergleich	50
Tabelle 4-10: Übersicht über die patientenrelevante Endpunkte, die für die Bewertung von Teriflunomid herangezogen werden.....	61
Tabelle 4-11: Liste der Studien des pharmazeutischen Unternehmers – RCT ^a mit dem zu bewertenden Arzneimittel	75
Tabelle 4-12: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel.....	77
Tabelle 4-13: Relevante Publikationen	81
Tabelle 4-14: Relevante Studien aus der Suche in Studienregistern – RCT mit dem zu bewertenden Arzneimittel	82
Tabelle 4-15: Studienpool – RCT mit dem zu bewertenden Arzneimittel.....	83
Tabelle 4-16: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel	84
Tabelle 4-17: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel.....	86
Tabelle 4-18: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel.....	87
Tabelle 4-19: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel.....	89
Tabelle 4-20: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel	91
Tabelle 4-21: Definitionen der Studienpopulationen in den Studienberichten von TEMSO, TOWER und TENERE	93

Tabelle 4-22: Für die Bewertung ausgewertete Patientenpopulationen.....	95
Tabelle 4-23: Übersicht der während der Studien aufgetretenen Schübe – in der gesamten Studienpopulation.....	97
Tabelle 4-24: Übersicht der während der Studien aufgetretenen Schübe in der RRMS-Subpopulation.....	98
Tabelle 4-25: Operationalisierung des Endpunktes „Jährliche Schubrate“	99
Tabelle 4-26: Bewertung des Verzerrungspotenzials für „Jährliche Schubrate“ in RCT mit dem zu bewertenden Arzneimittel.....	100
Tabelle 4-27: Ergebnisse für „Jährliche Schubrate“ aus RCT mit dem zu bewertenden Arzneimittel - für die gesamte Studienpopulation	101
Tabelle 4-28: Ergebnisse für jährliche Schubrate aus TEMSO und TOWER - für die Subgruppe der RRMS-Patienten	102
Tabelle 4-29: Operationalisierung des Endpunktes „Zeit bis zum Schub“	106
Tabelle 4-30: Bewertung des Verzerrungspotenzials für „Zeit bis zum Schub“ in RCT mit dem zu bewertenden Arzneimittel.....	107
Tabelle 4-31: Ergebnisse für „Zeit bis zum Schub“ aus RCT mit dem zu bewertenden Arzneimittel - für die gesamte Studienpopulation	108
Tabelle 4-32: Ergebnisse für „Patienten mit Schub“ aus TEMSO und TOWER - für die RRMS-Subpopulation	110
Tabelle 4-33: Operationalisierung des Endpunktes „Schubfreie Patienten“	113
Tabelle 4-34: Bewertung des Verzerrungspotenzials für „Schubfreie Patienten“ in RCT mit dem zu bewertenden Arzneimittel	113
Tabelle 4-35: „Schubfreie Patienten“ aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation.....	114
Tabelle 4-36: „Schubfreie Patienten“ aus TEMSO und TOWER - für die RRMS-Subpopulation.....	115
Tabelle 4-37: Schwere der Schübe und andere Schub-Effekte – Teil 1 Auswertungen der TEMSO Studie	116
Tabelle 4-38: Schwere der Schübe und andere Schub-Effekte – Teil 2 Auswertungen der TEMSO Studie	116
Tabelle 4-39: Schwere der Schübe und andere Schub-Effekte – Auswertungen der TEMSO - Studie- Teil 3 plus Auswertungen der TOWER-Studie ^a	117
Tabelle 4-40: Ergebnis der Metaanalyse von TEMSO und TOWER für die jährliche Schubrate – für die RRMS-Subpopulation.....	119
Tabelle 4-41: Ergebnis der Metaanalyse von TEMSO und TOWER für die Zeit bis zum Schub – für die gesamte Studienpopulation.....	120
Tabelle 4-42: Ergebnis der Metaanalyse von TEMSO und TOWER für „Patienten mit Schub“ – für die RRMS-Subpopulation.....	120
Tabelle 4-43: Ergebnis der Metaanalyse von TEMSO und TOWER für „schubfreie Patienten“ – für die RRMS-Subpopulation.....	121

Tabelle 4-44: Operationalisierung von „Zeit bis zur Behinderungsprogression“ (inklusive Anteil Patienten mit Behinderungsprogression zu einem bestimmten Zeitpunkt).....	123
Tabelle 4-45: Bewertung des Verzerrungspotenzials für „Zeit bis zur 12 Wochen anhaltenden Behinderungsprogression“ in RCT mit dem zu bewertenden Arzneimittel	124
Tabelle 4-46: Ergebnisse für „Zeit bis zur Behinderungsprogression“ aus RCT mit dem zu bewertenden Arzneimittel - für die gesamte Studienpopulation	126
Tabelle 4-47: Ergebnisse für „Zeit bis zur Behinderungsprogression“ aus TEMSO und TOWER - für die Subpopulation RRMS-Patienten	128
Tabelle 4-48: Ergebnisse für „Patienten mit Behinderungsprogression“ – für die gesamte Studienpopulation.....	129
Tabelle 4-49: Ergebnisse für „Patienten mit Behinderungsprogression“ aus TEMSO und TOWER - für die RRMS-Subpopulation	130
Tabelle 4-50: Operationalisierung von Mittlere Änderung des MSFC-Scores.....	132
Tabelle 4-51: Bewertung des Verzerrungspotenzials für „Mittlere Änderung des MSFC-Z-Scores“ in RCT mit dem zu bewertenden Arzneimittel	133
Tabelle 4-52: Ergebnisse für Mittlere Änderung des MSFC Z-Scores aus RCT mit dem zu bewertenden Arzneimittel - für die gesamte Studienpopulation	134
Tabelle 4-53: Operationalisierung von Mittlere Änderung des EDSS-Scores.....	136
Tabelle 4-54: Bewertung des Verzerrungspotenzials für „Mittlere Änderung des EDSS-Scores“ in RCT mit dem zu bewertenden Arzneimittel	137
Tabelle 4-55: Ergebnisse für „Mittlere Änderung des EDSS-Scores“ aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation	139
Tabelle 4-56: Metaanalyse-Ergebnisse für die „Zeit bis zur Behinderungsprogression“ aus TEMSO und TOWER für die Subpopulation RRMS-Patienten	142
Tabelle 4-57: Ergebnis der Metaanalyse von TEMSO und TOWER für „Patienten mit Behinderungsprogression“ für die Subpopulation der RRMS-Patienten	143
Tabelle 4-58: Ergebnis der Metaanalyse von TEMSO und TOWER für „Mittlere Änderung des EDSS“ – für die gesamte Studienpopulation	143
Tabelle 4-59: Operationalisierung von „Therapieversagen (Time to Failure)“	145
Tabelle 4-60: Bewertung des Verzerrungspotenzials für „Time to Failure“ in RCT mit dem zu bewertenden Arzneimittel.....	146
Tabelle 4-61: Ergebnisse für Time to Failure aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation.....	147
Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein. Tabelle 4-62: Operationalisierung von Fatigue (Fatigue Impact Scale = FIS)	151
Tabelle 4-63: Bewertung des Verzerrungspotenzials für Fatigue (Fatigue Impact Scale = FIS) in RCT mit dem zu bewertenden Arzneimittel	152
Tabelle 4-64: Ergebnisse für Fatigue aus RCT mit dem zu bewertenden Arzneimittel-FIS-Gesamtwert – für die gesamte Studienpopulation.....	153

Tabelle 4-65: Ergebnisse für Fatigue aus RCT mit dem zu bewertenden Arzneimittel-FIS-Subscore der kognitiven Dimension – für die gesamte Studienpopulation	154
Tabelle 4-66: Ergebnisse für Fatigue aus RCT mit dem zu bewertenden Arzneimittel-FIS-Subscore der physischen Dimension – für die gesamte Studienpopulation.....	156
Tabelle 4-67: Ergebnisse für Fatigue aus RCT mit dem zu bewertenden Arzneimittel-FIS-Subscore der psychosozialen Dimension – für die gesamte Studienpopulation	157
Tabelle 4-68: Operationalisierung von Krankheitslast (Burden of Disease - MRT)	161
Tabelle 4-69: Bewertung des Verzerrungspotenzials für Krankheitslast (Burden of Disease (MRT)) in RCT mit dem zu bewertenden Arzneimittel	161
Tabelle 4-70: Ergebnisse für Krankheitslast (Burden of Disease) aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation	163
Tabelle 4-71: Ergebnisse für ausgewählte weitere MRT-Parameter aus RCT mit dem zu bewertenden Arzneimittel - für die gesamte Studienpopulation.....	164
Tabelle 4-72: Ergebnisse für aktive Läsionen („unique active lesions“) aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation	167
Tabelle 4-73: Operationalisierung von Patientenzufriedenheit.....	169
Tabelle 4-74: Bewertung des Verzerrungspotenzials für Patientenzufriedenheit in RCT mit dem zu bewertenden Arzneimittel	170
Tabelle 4-75: Ergebnisse für Patientenzufriedenheit aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation.....	171
Tabelle 4-76: Operationalisierung von SF-36 (Lebensqualität).....	174
Tabelle 4-77: Bewertung des Verzerrungspotenzials für SF-36 (Lebensqualität) in RCT mit dem zu bewertenden Arzneimittel	175
Tabelle 4-78: Ergebnisse für Lebensqualität SF-36 aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation.....	176
Tabelle 4-79: Operationalisierung von Lebensqualität	178
Tabelle 4-80: Bewertung des Verzerrungspotenzials für EQ-5D (Lebensqualität) in RCT mit dem zu bewertenden Arzneimittel	179
Tabelle 4-81: Ergebnisse für Lebensqualität EQ-5D aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation.....	180
Tabelle 4-82: Operationalisierung von „Während der Studie verstorbene Patienten“ (Mortalität)	182
Tabelle 4-83: Bewertung des Verzerrungspotenzials für Mortalität in RCT mit dem zu bewertenden Arzneimittel	183
Tabelle 4-84: Ergebnisse für Mortalität aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation.....	184
Tabelle 4-85: Metaanalyse-Ergebnisse für Mortalität aus TEMSO und TOWER – für die gesamte Studienpopulation	185
Tabelle 4-86: Operationalisierung von Gesamtraten unerwünschter Ereignisse (AEs, SAEs, Behandlungsabbrüche).....	187

Tabelle 4-87: Definitionen der Safety-Population	188
Tabelle 4-88: Bewertung des Verzerrungspotenzials für Gesamtraten unerwünschter Ereignisse (AEs, SAEs, Behandlungsabbrüche) in RCT mit dem zu bewertenden Arzneimittel.....	190
Tabelle 4-89: Betrachteter Studienzeitraum, in dem die unerwünschten Ereignisse beobachtet werden konnten	191
Tabelle 4-90: Ergebnisse für die Gesamtrate unerwünschter Ereignisse aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation	191
Tabelle 4-91: Ergebnisse für die Gesamtrate schwerwiegender unerwünschter Ereignisse aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation	192
Tabelle 4-92: Ergebnisse für die Gesamtrate unerwünschter Ereignisse, die zum Behandlungsabbruch führten, aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation	193
Tabelle 4-93: Metaanalyse-Ergebnisse für Gesamtrate unerwünschter Ereignisse aus TEMSO und TOWER – für die gesamte Studienpopulation	195
Tabelle 4-94: Metaanalyse-Ergebnisse für Gesamtrate schwerwiegender unerwünschter Ereignisse aus TEMSO und TOWER – für die gesamte Studienpopulation	196
Tabelle 4-95: Metaanalyse-Ergebnisse für unerwünschte Ereignisse, die zum Behandlungsabbruch führten aus TEMSO und TOWER – für die gesamte Studienpopulation.....	197
Tabelle 4-96: Operationalisierung von unerwünschten Ereignissen von besonderem Interesse.....	198
Tabelle 4-97: Ausgewertete AESI aufgeschlüsselt nach SOC, HLGT, HLT und PT	201
Tabelle 4-98: Bewertung des Verzerrungspotenzials für „Unerwünschte Ereignisse von besonderem Interesse“ in RCT mit dem zu bewertenden Arzneimittel	202
Tabelle 4-99: Reaktionen an der Injektionsstelle – Ergebnisse für die gesamte Studienpopulation.....	203
Tabelle 4-100: Grippeähnliche Symptome – Ergebnisse für die gesamte Studienpopulation.....	205
Tabelle 4-101: Neutropenie – Ergebnisse für die gesamte Studienpopulation	209
Tabelle 4-102: Metaanalyse-Ergebnisse für Neutropenie aus TEMSO und TOWER – für die gesamte Studienpopulation.....	211
Tabelle 4-103: Alopezie – Ergebnisse für die gesamte Studienpopulation	211
Tabelle 4-104: Metaanalyse-Ergebnisse für verminderte Haardichte (HLT Alopecias) aus TEMSO und TOWER – Ergebnisse für die gesamte Studienpopulation.....	214
Tabelle 4-105: Überempfindlichkeit/Hautreaktionen – Ergebnisse für die gesamte Studienpopulation.....	214
Tabelle 4-106: Metaanalyse-Ergebnisse für Überempfindlichkeits- und Hautreaktionen aus TEMSO und TOWER – für die gesamte Studienpopulation.....	216
Tabelle 4-107: Hepatische unerwünschte Ereignisse – Ergebnisse für die gesamte Studienpopulation.....	217

Tabelle 4-108: Patienten mit hepatischen unerwünschten Ereignissen – Ergebnisse für die gesamte Studienpopulation	218
Tabelle 4-109: Erhöhung der Alanin-Aminotransferase (ALT)-Werte – Ergebnisse für die gesamte Studienpopulation	219
Tabelle 4-110: Gastrointestinale Beschwerden – Ergebnisse für die gesamte Studienpopulation.....	222
Tabelle 4-111: Metaanalyse-Ergebnisse für Gastrointestinale Beschwerden (SOC) aus TEMSO und TOWER – für die gesamte Studienpopulation	226
Tabelle 4-112: Metaanalyse-Ergebnisse für Durchfall (PT Diarrhoea) aus TEMSO und TOWER – für die gesamte Studienpopulation.....	227
Tabelle 4-113: Metaanalyse-Ergebnisse für Übelkeit (PT Nausea) aus TEMSO und TOWER –für die gesamte Studienpopulation.....	227
Tabelle 4-114: Infektionen und Infestationen – Ergebnisse für die gesamte Studienpopulation.....	228
Tabelle 4-115: Metaanalyse-Ergebnisse für Infektionen und Infestationen (SOC) aus TEMSO und TOWER –für die gesamte Studienpopulation	230
Tabelle 4-116: Bluthochdruck – Ergebnisse für die gesamte Studienpopulation	231
Tabelle 4-117: Kopfschmerzen und Periphere Neuropathie – Ergebnisse für die gesamte Studienpopulation.....	232
Tabelle 4-118: Parästhesie – Ergebnisse für die gesamte Studienpopulation	233
Tabelle 4-119: Metaanalyse-Ergebnisse für Bluthochdruck aus TEMSO und TOWER – für die gesamte Studienpopulation.....	235
Tabelle 4-120: Metaanalyse-Ergebnisse für periphere Neuropathie (SOC) aus TEMSO und TOWER – für die gesamte Studienpopulation.....	236
Tabelle 4-121: 14 mg Teriflunomid versus Placebo (Ergebnisse der TEMSO- und der TOWER-Studie).....	237
Tabelle 4-122: 14 mg Teriflunomid versus Rebif® 44 µg (Ergebnisse der TENERE-Studie)	238
Tabelle 4-123: Subgruppenanalysen für den primären Endpunkt Time to Failure in TENERE.....	242
Tabelle 4-124: Subgruppenanalysen für den sekundären Endpunkt jährliche Schubrate in TENERE.....	243
Tabelle 4-125: Subgruppenanalysen für den Endpunkt jährliche Schubrate in TEMSO	245
Tabelle 4-126: Subgruppenanalysen für den Endpunkt Zeit bis zur Behinderungsprogression in TEMSO	247
Tabelle 4-127: Subgruppenanalysen für den Endpunkt jährliche Schubrate in TOWER.....	250
Tabelle 4-128: Subgruppenanalysen für den Endpunkt Zeit bis zur Behinderungsprogression (12 week sustained disability progression) in TOWER.....	252
Tabelle 4-129: Vergleichende Gegenüberstellung der Subgruppenanalysen aus TEMSO und TOWER- Jährliche Schubrate	255

Tabelle 4-130: 14 mg Teriflunomid versus Placebo (Ergebnisse der TEMSO- und der TOWER-Studie).....	258
Tabelle 4-131: 14 mg Teriflunomid versus Rebif [®] 44 µg (Ergebnisse der TENERE-Studie ^a).....	264
Tabelle 4-132: Liste der Studien des pharmazeutischen Unternehmers – RCT ^a mit dem zu bewertenden Arzneimittel	271
Tabelle 4-133: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel.....	273
Tabelle 4-134: Relevante Studien aus der bibliografischen Literaturrecherche – RCT mit dem zu bewertenden Arzneimittel.....	276
Tabelle 4-135: Relevante Studien aus der bibliografischen Literaturrecherche – RCT mit der zweckmäßigen Vergleichstherapie.....	278
Tabelle 4-136: Relevante Studien aus der Suche in Studienregistern – RCT mit dem zu bewertenden Arzneimittel	279
Tabelle 4-137: Relevante Studien aus der Suche in Studienregistern – RCT mit der zweckmäßigen Vergleichstherapie (Rebif [®] (Interferon beta-1a, s. c.))	279
Tabelle 4-138: Studienpool – RCT mit dem zu bewertenden Arzneimittel und der zweckmäßigen Vergleichstherapie (placebokontrollierte Studien).....	280
Tabelle 4-139: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel	281
Tabelle 4-140: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel.....	283
Tabelle 4-141: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel	284
Tabelle 4-142: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel.....	288
Tabelle 4-143: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche	289
Tabelle 4-144: Ausgewertete Studienpopulationen der Teriflunomid-Studien in den indirekten Vergleichen nach Endpunkt	291
Tabelle 4-145: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden.....	292
Tabelle 4-146: Operationalisierung von Schüben	293
Tabelle 4-147: Bewertung des Verzerrungspotenzials für Schübe in RCT für indirekte Vergleiche	294
Tabelle 4-148: Ergebnisse für jährliche Schubrate aus RCT des indirekten und direkten Vergleichs – für die RRMS-Subpopulation	295
Tabelle 4-149: Definitionen für den indirekten Vergleich.....	296
Tabelle 4-150: Ergebnisse für die jährliche Schubrate aus RCT mit dem zu bewertenden Arzneimittel- 1. indirekter Vergleich – für die RRMS-Subpopulation.....	297

Tabelle 4-151: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden.....	299
Tabelle 4-152: Operationalisierung von Behinderungsprogression.....	299
Tabelle 4-153: Bewertung des Verzerrungspotenzials für Zeit bis zur Behinderungsprogression in RCT für indirekte Vergleiche.....	301
Tabelle 4-154: Ergebnisse für Zeit bis zur Behinderungsprogression aus RCT des indirekten und direkten Vergleichs – für die RRMS-Subpopulation.....	302
Tabelle 4-155: Definitionen für den indirekten Vergleich.....	303
Tabelle 4-156: Ergebnisse für die Zeit bis zur Behinderungsprogression aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die RRMS-Subpopulation	303
Tabelle 4-157: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden.....	306
Tabelle 4-158: Operationalisierung von Mortalität.....	307
Tabelle 4-159: Bewertung des Verzerrungspotenzials für Mortalität in RCT für indirekte Vergleiche	307
Tabelle 4-160: Ergebnisse für Mortalität aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation.....	308
Tabelle 4-161: Definitionen für den indirekten Vergleich.....	309
Tabelle 4-162: Ergebnisse für Mortalität aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich 1 – für die gesamte Studienpopulation.....	309
Tabelle 4-163: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden.....	311
Tabelle 4-164: Operationalisierung von „Unerwünschte Ereignisse, die zum Abbruch der Behandlung führten“	312
Tabelle 4-165: Bewertung des Verzerrungspotenzials für „Unerwünschte Ereignisse, die zum Abbruch der Behandlung führten“ in RCT für indirekte Vergleiche.....	313
Tabelle 4-166: Ergebnisse für „Unerwünschte Ereignisse, die zu einem Behandlungsabbruch führten“ aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation	314
Tabelle 4-167: Definitionen für den indirekten Vergleich.....	315
Tabelle 4-168: Ergebnisse für „Unerwünschte Ereignisse, die zu einem Behandlungsabbruch führten“aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die gesamte Studienpopulation	315
Tabelle 4-169: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden – für die gesamte Studienpopulation.....	317
Tabelle 4-170: Operationalisierung von „Unerwünschte Ereignisse von besonderem Interesse“	318
Tabelle 4-171: „Unerwünschte Ereignisse von besonderem Interesse“	321
Tabelle 4-172: Bewertung des Verzerrungspotenzials für „Unerwünschte Ereignisse von besonderem Interesse“ in RCT für indirekte Vergleiche	321

Tabelle 4-173: Definitionen für den indirekten Vergleich.....	322
Tabelle 4-174: Ergebnisse für grippeähnliche Symptome aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation	323
Tabelle 4-175: Ergebnisse für grippeähnliche Symptome aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die gesamte Studienpopulation.....	324
Tabelle 4-176: Ergebnisse für Reaktionen an der Injektionsstelle aus relevanten RCT – für die gesamte Studienpopulation	326
Tabelle 4-177: Ergebnisse für Kopfschmerzen aus RCT des indirekten und direkten Vergleichs– für die gesamte Studienpopulation.....	327
Tabelle 4-178: Ergebnisse für Kopfschmerzen aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die gesamte Studienpopulation	329
Tabelle 4-179: Ergebnisse für aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation	331
Tabelle 4-180: Ergebnisse für erhöhte ALT-Werte aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die gesamte Studienpopulation	332
Tabelle 4-181: Ergebnisse für Lymphopenie aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation.....	334
Tabelle 4-182: Ergebnisse für Lymphopenie aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die gesamte Studienpopulation	335
Tabelle 4-183: Ergebnisse für Leukopenie aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation.....	336
Tabelle 4-184: Ergebnisse für Leukopenie aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die gesamte Studienpopulation	338
Tabelle 4-185: Ergebnisse aus RCT mit dem zu bewertenden Arzneimittel- indirekte Vergleiche	340
Tabelle 4-186: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien	343
Tabelle 4-187: Operationalisierung von <Endpunkt xxx>.....	343
Tabelle 4-188: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien	344
Tabelle 4-189: Operationalisierung von <Endpunkt xxx>.....	345
Tabelle 4-190: Beurteilung der eingeschlossenen Studien	348
Tabelle 4-191: Einzelergebnisse aus denen sich der Zusatznutzen ableitet (aufgeschlüsselt nach Endpunkten)- direkter Vergleich (TENERE)	350
Tabelle 4-192: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens.....	356
Tabelle 4-193: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens – Arzneimittel zur Behandlung eines seltenen Leidens	357
Tabelle 4-194: Liste der eingeschlossenen Studien und der zugehörigen Datenquellen	360

Tabelle 4-195 (Anhang): Dokumentation der Suchstrategie in MEDLINE (1).....	370
Tabelle 4-196 (Anhang): Dokumentation der Suchstrategie in MEDLINE (2).....	371
Tabelle 4-197 (Anhang): Dokumentation der Suchstrategie in EMBASE	372
Tabelle 4-198 (Anhang): Dokumentation der Suchstrategie in Cochrane	373
Tabelle 4-199 (Anhang): Dokumentation der Suchstrategie in MEDLINE	373
Tabelle 4-200 (Anhang): Dokumentation der Suchstrategie in EMBASE	374
Tabelle 4-201 (Anhang): Dokumentation der Suchstrategie in Cochrane	374
Tabelle 4-202 (Anhang): Dokumentation des Volltext-Screenings (Recherche nach randomisierten, kontrollierten Studien, die Teriflunomid 14 mg, einmal pro Tag, mit Placebo vergleichen), ausgeschlossene Studien	380
Tabelle 4-203 (Anhang): Dokumentation des Volltext-Screenings (Recherche nach randomisierten, kontrollierten Studien, die Rebif 44 µg, dreimal die Woche, mit Placebo vergleichen), ausgeschlossene Studien	380
Tabelle 4-204 (Anhang): Studienpool, abgebrochene Studien, zu bewertendes Arzneimittel.....	382
Tabelle 4-205 (Anhang): Studienpool, laufende Studien, zu bewertendes Arzneimittel.....	383
Tabelle 4-206 (Anhang): Studiendesign und -methodik für Studie <Studienbezeichnung> .	386
Tabelle 4-207 (Anhang): Studiendesign und -methodik für die Studie TEMSO.....	388
Tabelle 4-208 (Anhang): Studiendesign und -methodik für Studie TOWER.....	402
Tabelle 4-209 (Anhang): Studiendesign und -methodik für Studie TENERE.....	418
Tabelle 4-210 (Anhang): Studiendesign und -methodik für Studie PRISMS	433
Tabelle 4-211 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie <Studienbezeichnung>	441
Tabelle 4-212 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie TEMSO	451
Tabelle 4-213 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie TOWER.....	483
Tabelle 4-214 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie TENERE.....	509
Tabelle 4-215 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie PRISMS	536

Abbildungsverzeichnis

	Seite
Abbildung 4-1: Netzwerk für den indirekten Vergleich	49
Abbildung 2: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten, kontrollierten Studien mit dem zu bewertenden Arzneimittel	79
Abbildung 4-3: Flussdiagramm für die systematische Literaturrecherche in MEDLINE, EMBASE und den Cochrane-Datenbanken	80
Abbildung 4-4: Kaplan-Meier-Kurven für Zeit bis zum ersten Schub in TEMSO (links) und TOWER (rechts)	109
Abbildung 4-5: Forest-Plot der Metaanalyse von TEMSO und TOWER für die jährliche Schubrate	119
Abbildung 4-6: Forest-Plot der Metaanalyse von TEMSO und TOWER für die Zeit bis zum Schub	120
Abbildung 4-7: Forest-Plot der Metaanalyse (basierend auf den relativen Risiken) von TEMSO und TOWER für die „Patienten mit Schub“	121
Abbildung 4-8: Forest-Plot der Metaanalyse (basierend auf den relativen Risiken) von TEMSO und TOWER für die „schubfreie Patienten“	121
Abbildung 4-9: Kaplan-Meier-Kurven für „Zeit bis zur Behinderungsprogression“ in TEMSO (links) und TOWER (rechts)	127
Abbildung 4-10: Forest-Plot der Metaanalyse von TEMSO und TOWER für die Zeit bis zur Behinderungsprogression	142
Abbildung 4-11: Forest-Plot der Metaanalyse von TEMSO und TOWER (basierend auf den relativen Risiken) für „Patienten mit Behinderungsprogression“	143
Abbildung 4-12: Forest-Plot der Metaanalyse von TEMSO und TOWER (basierend auf den Mittelwertsdifferenzen) für „Mittlere Änderung des EDSS“	143
Abbildung 4-13: Kaplan-Meier-Kurve für „Zeit bis zum Therapieversagen“ in TENERE .	149
Abbildung 4-14: Forest-Plot der Metaanalyse von TEMSO und TOWER für Mortalität.....	185
Abbildung 4-15: Forest-Plot der Metaanalyse von TEMSO und TOWER für Gesamtrate unerwünschter Ereignisse (basierend auf relativen Risiken)	196
Abbildung 4-16: Forest-Plot der Metaanalyse von TEMSO und TOWER für Gesamtrate schwerwiegender unerwünschter Ereignisse (basierend auf relativen Risiken)	196
Abbildung 4-17: Forest-Plot der Metaanalyse von TEMSO und TOWER für unerwünschte Ereignisse, die zum Behandlungsabbruch führten (basierend auf relativen Risiken)	197
Abbildung 4-18: Forest-Plot der Metaanalyse von TEMSO und TOWER für grippeähnliche Symptome (basierend auf relativen Risiken)	208
Abbildung 4-19: Forest-Plot der Metaanalyse von TEMSO und TOWER für Neutropenie (basierend auf relativen Risiken).....	211
Abbildung 4-20: Forest-Plot der Metaanalyse von TEMSO und TOWER für verminderte Haardichte (HLT Alopecias) (basierend auf relativen Risiken).....	214

Abbildung 4-21: Forest-Plot der Metaanalyse von TEMSO und TOWER für Überempfindlichkeits- und Hautreaktionen (basierend auf relativen Risiken).....	216
Abbildung 4-22: Forest-Plot der Metaanalyse von TEMSO und TOWER für Erhöhung der ALT-Werte (basierend auf relativen Risiken).....	221
Abbildung 4-23: Forest-Plot der Metaanalyse von TEMSO und TOWER für Gastrointestinale Beschwerden (SOC) (basierend auf relativen Risiken)	226
Abbildung 4-24: Forest-Plot der Metaanalyse von TEMSO und TOWER für Durchfall (PT Diarrhoea) (basierend auf relativen Risiken)	227
Abbildung 4-25: Forest-Plot der Metaanalyse von TEMSO und TOWER für Übelkeit (PT Nausea) (basierend auf relativen Risiken)	228
Abbildung 4-26: Forest-Plot der Metaanalyse von TEMSO und TOWER für Infektionen und Infestationen (SOC) (basierend auf relativen Risiken)	230
Abbildung 4-27: Forest-Plot der Metaanalyse von TEMSO und TOWER für Bluthochdruck (basierend auf relativen Risiken).....	235
Abbildung 4-28: Forest-Plot der Metaanalyse von TEMSO und TOWER für Kopfschmerzen (basierend auf relativen Risiken)	236
Abbildung 4-29: Forest-Plot der Metaanalyse von TEMSO und TOWER für periphere Neuropathie (basierend auf relativen Risiken).....	237
Abbildung 30: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten, kontrollierten Studien mit dem zu bewertenden Arzneimittel	275
Abbildung 4-31: Flussdiagramm der bibliografischen Literaturrecherche in MEDLINE, EMBASE und den Cochrane-Datenbanken	277
Abbildung 4-32: Netzwerk für den indirekten Vergleich	292
Abbildung 4-33: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Rate Ratio der jährlichen Schubrate.....	298
Abbildung 4-34: Netzwerk für den indirekten Vergleich	299
Abbildung 4-35: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Hazard Ratio der Zeit bis zur Behinderungsprogression	305
Abbildung 4-36: Netzwerk für den indirekten Vergleich	306
Abbildung 4-37: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risikodifferenz der Mortalität.....	310
Abbildung 4-38: Netzwerk für den indirekten Vergleich	311
Abbildung 4-39: Forest-Plot zur Metaanalyse 1 (von oben nach unten: TEMSO, TOWER, Ergebnis der Metaanalyse) – Risk Ratio	316
Abbildung 4-40: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risk Ratio.....	317
Abbildung 4-41: Netzwerk für den indirekten Vergleich	318
Abbildung 4-42: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risk Ratio.....	325

Abbildung 4-43: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risk Ratio.....	330
Abbildung 4-44: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risk Ratio.....	333
Abbildung 4-45: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse 2) – Risk Ratio.....	336
Abbildung 4-46: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risk Ratio.....	339
Abbildung 4-47: Netzwerk für den indirekten Vergleich	357

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AE	Unerwünschtes Ereignis (adverse event)
AESI	Unerwünschtes Ereignis von besonderem Interesse (adverse event of special interest)
ALT	Alanin-Aminotransferase
ARR	Jährliche Schubrate (annualized relapse rate)
CHMP	Committee for Medicinal Products for Human Use
CIS	Klinisch isoliertes Syndrom (clinically isolated syndrome)
CONSORT	Consolidated Standards of Reporting Trials
DIMDI	Deutsches Institut für Medizinische Dokumentation
EDSS	Expanded Disability Status Scale
EG	Europäische Gemeinschaft
EMA	European Medicines Agency
EQ-5D	EuroQoL-5D (Fragebogen zur Erfassung der Lebensqualität)
FIS	Fatigue Impact Scale
FS	Functional System
GCP	Good Clinical Practice
GEE	Generalized Estimating Equation
HLGT	High Level Group Term
HLT	High Level Term
HR	Hazard Ratio
ITT	Intention to treat
KI	Konfidenzintervall
LLT	Lowest Level Term
LS-Means	=Least Squares-Mittelwerte: sich aus dem MMRM-Modell ergebende adjustierte Schätzer
MD	Mittelwertsdifferenz
MedDRA	Medical Dictionary for Regulatory Activities
MMRM	Mixed-effect model with repeated measures
MRI	Magnetic resonance imaging
MRT	Magnetresonanztomographie
MS	Multiple Sklerose

MSFC	Multiple Sclerosis Functional Composite
MTC	Mixed Treatment Comparison
OR	Odds Ratio
PP	Per protocol
pU	Pharmazeutischer Unternehmer
PT	Preferred Term
qd	Einmal täglich, jeden Tag (=quaque die)
RCT	Randomized Controlled Trial
RD	Risikodifferenz
RMS	Schubförmige Multiple Sklerose (relapsing MS)
RR	Relatives Risiko
RRMS	Schubförmig-remittierende Multiple Sklerose (relapsing-remitting MS)
SD	Standardabweichung (standard deviation)
SDP	Sustained disability progression
SE	Standardfehler (standard error)
SF-36	Short-Form 36
SGB	Sozialgesetzbuch
SOC	System Organ Class
STE	Surrogate Threshold Effects
STEAE	Schwerwiegendes unerwünschtes Ereignis, das während der Behandlungsdauer aufgetreten ist (serious treatment-emergent adverse event)
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
TEAE	Unerwünschtes Ereignis, das während der Behandlungsdauer aufgetreten ist (treatment-emergent adverse event)
tiw	Dreimal wöchentlich (three times a week)
TREND	Transparent Reporting of Evaluations with Non-Randomized Design
TSQM	Treatment Satisfaction Questionnaire for Medication
WBC	White Blood Cell
WHO	World Health Organization

4 Modul 4 – allgemeine Informationen

Modul 4 enthält folgende Angaben:

- Zusammenfassung (Abschnitt 4.1)
- Angaben zur Methodik der im Dossier präsentierten Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens (Abschnitt 4.2)
- Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen (Abschnitt 4.3)
- eine abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens, einschließlich der Angabe von Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht (Abschnitt 4.4)
- ergänzende Informationen zur Begründung der vorgelegten Unterlagen (Abschnitt 4.5)

Für jedes zu bewertende Anwendungsgebiet ist eine separate Version des vorliegenden Dokuments zu erstellen. Die Kodierung der Anwendungsgebiete ist in Modul 2 hinterlegt. Sie ist je Anwendungsgebiet einheitlich für die Module 3, 4 und 5 zu verwenden.

Für Arzneimittel, die zur Behandlung eines seltenen Leidens nach der Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden zugelassen sind, müssen keine Nachweise zum medizinischen Nutzen und medizinischen Zusatznutzen vorgelegt werden (Abschnitt 4.3), solange der Umsatz des Arzneimittels mit der gesetzlichen Krankenversicherung zu Apothekenverkaufspreisen einschließlich Umsatzsteuer in den letzten 12 Monaten einen Betrag von 50 Millionen Euro nicht übersteigt. Angaben zu Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, müssen jedoch vorgelegt werden. Zu diesem Zweck enthält die Dokumentvorlage den Abschnitt 4.4.4, der ausschließlich in Dossiers für die oben genannten Arzneimittel zur Behandlung eines seltenen Leidens bearbeitet werden soll. In diesem Abschnitt sind Angaben zu Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich der begründenden Nachweise vorzulegen. Weitere Abschnitte von Modul 4 können dabei nach Bedarf bearbeitet werden.

Im Dokument verwendete Abkürzungen sind in das Abkürzungsverzeichnis aufzunehmen. Sofern Sie für Ihre Ausführungen Tabellen und Abbildungen verwenden, sind diese im Tabellen- bzw. Abbildungsverzeichnis aufzuführen.

4.1 Zusammenfassung der Inhalte von Modul 4

Stellen Sie eine strukturierte Zusammenfassung der Inhalte von Modul 4 zur Verfügung.

4.1.1 Hintergrund

Multiple Sklerose ist eine chronisch entzündliche demyelinisierende Erkrankung des zentralen Nervensystems. Der Erkrankungsbeginn liegt überwiegend im jungen Erwachsenenalter und hat für die Betroffenen privat wie beruflich weitreichende Konsequenzen. Durch die zunehmende Behinderung kann ihre Lebensqualität und ihre Arbeitsfähigkeit stark eingeschränkt sein. Mit den seit den 1990er Jahren zugelassenen Interferon beta-Präparaten und dem seit 2001 zugelassenen Glatirameracetat stehen krankheitsmodifizierende Wirkstoffe zur Behandlung der schubförmig-remittierenden MS zur Verfügung, die auch heute noch als Basistherapeutika der ersten Wahl gelten.

Das Leben der MS-Patienten kann jedoch nicht nur durch ihre Erkrankung, sondern auch durch die Therapie in seiner Qualität eingeschränkt sein. Die Therapie mit Interferon beta bzw. Glatirameracetat erfordert regelmäßige Injektionen und kann mit für den Patienten stark belastenden Nebenwirkungen wie z. B. grippeähnlichen Symptomen assoziiert sein.

Mit Teriflunomid (AUBAGIO®) steht den Patienten nun ein krankheitsmodifizierender Wirkstoff zur Verfügung, der in seiner Effektivität den etablierten Basistherapeutika vergleichbar ist und sich durch eine im allgemeinen gute Verträglichkeit und einfache orale Applikation auszeichnet.

4.1.2 Fragestellung

Die im vorliegenden Dossier zu beantwortende Frage lautet:

Hat die Behandlung mit 14 mg Teriflunomid, einmal täglich, oral, bei erwachsenen Patienten mit schubförmig-remittierender Multipler Sklerose (RRMS) hinsichtlich patientenrelevanter Endpunkte gegenüber der zweckmäßigen Vergleichstherapie Interferon beta-1a, 44 µg, s. c., dreimal pro Woche (Rebif® 44 µg) einen Zusatznutzen im Sinne des § 35a SGB V?

Darüber hinaus soll auch der medizinische Nutzen, der im Rahmen des Zulassungsverfahrens bereits bestätigt wurde, ergänzend dargestellt werden. Die Nachweise sollen mittels randomisierter kontrollierter Studien (Evidenzstufe 1b) erfolgen. Für die Darstellung des Nutzens stehen drei randomisierte, kontrollierte Studien zur Verfügung, wobei die Wirksamkeit und Verträglichkeit von Teriflunomid in zwei Studien (TEMSO und TOWER) gegen Placebo und in einer Studie (TENERE) gegen den aktiven Komparator Interferon beta-1a, 44 µg, s. c. (Rebif® 44 µg) gezeigt wird.

Wie für die Darstellung des Nutzens so ergibt sich auch für die Darstellung des Zusatznutzens die beste Evidenz aus direkt vergleichenden Studien, die als Intervention das zu bewertende Arzneimittel und als Komparator die zweckmäßige Vergleichstherapie untersuchen. Für die Bewertung des Zusatznutzens von Teriflunomid steht mit TENERE eine direkt vergleichende Studie zur Verfügung.

Ergänzend zu dem direkten Vergleich werden indirekte Vergleiche über Placebo als Brückenkomparator durchgeführt, was die Einbeziehung der Zulassungsstudien TEMSO und TOWER ermöglicht. In die indirekten Vergleiche fließen zusätzlich die Ergebnisse aus der PRISMS-Studie ein, eine randomisierte kontrollierte Studie im Anwendungsgebiet von Teriflunomid, die die zweckmäßige Vergleichstherapie Interferon beta-1a, 44 µg, s. c. (Rebif® 44 µg) gegen Placebo vergleicht.

4.1.3 Datenquellen

Für die vorliegende Bewertung des medizinischen Nutzens und Zusatznutzens wurden die Studien des pharmazeutischen Unternehmers (dargestellt in den Abschnitten 4.3.1 und 4.3.2), herangezogen, die nach den vorab definierten Ein- und Ausschlusskriterien zur Beantwortung der Fragestellung geeignet sind. Für die Bewertung des medizinischen Zusatznutzens wurden neben den Ergebnissen der direkten Komparatorstudie zusätzlich indirekte Vergleiche durchgeführt. Dies machte eine bibliographische Literaturrecherche erforderlich, über die die relevanten Studien identifiziert worden sind. Das Vorgehen, die Suchstrategie und die gefundenen Treffer sind in den entsprechenden Dossierabschnitten dokumentiert (Abschnitt 4.2.3 „Informationsbeschaffung“ und in den Ergebnisteilen 4.3.1 und 4.3.2 der direkten und indirekten Vergleiche inklusive der entsprechenden Anhänge).

Die Suche in Studienregistern ergab keine weiteren relevanten Treffer zur Ergänzung des Studienpools.

4.1.4 Ein-/Ausschlusskriterien für Studien

Die vorab definierten Ein- und Ausschlusskriterien für Studien (PICOS-Kriterien) richten sich nach der medizinischen Fragestellung und berücksichtigen die folgenden Aspekte:

- Patientenpopulation
- Intervention
- Komparator
- Endpunkte (Outcomes)
- Studientypen/-dauer.

Gemäß der Fragestellung sind für die Bewertung des Nutzens und des Zusatznutzens von Teriflunomid nur solche Studien relevant, die die entsprechend der Zulassung infrage kommende Patientenpopulation einschließen. Diese Patientenpopulation besteht aus erwachsenen Patienten mit gesicherter MS-Diagnose, schubförmiger-remittierender Verlaufsform der MS und einem EDSS von $\leq 5,5$. Die Intervention muss Teriflunomid (bzw. die zweckmäßige Vergleichstherapie Interferon beta-1a 44 µg, s. c., dreimal pro Woche (Rebif® 44 µg)) sein. Je nach Detailfragestellung (Darstellung des Nutzens, Zusatznutzens oder indirekter Vergleich) kann der Komparator entweder die zweckmäßige

Vergleichstherapie, Interferon beta-1a (44 µg), s. c., dreimal pro Woche (Rebif® 44 µg) oder Placebo sein.

Es werden nur Studien eingeschlossen, die sich auf patientenrelevante Endpunkte stützen, die einer der folgenden Kategorien zugeordnet werden können:

- Mortalität
- Morbidität
- Lebensqualität
- Patientenzufriedenheit
- Unerwünschte Ereignisse.

Studien, die ausschließlich Surrogatparameter zur Bewertung heranziehen, wurden nicht in den Studienpool aufgenommen.

Zur adäquaten Beurteilung der Wirksamkeit und Verträglichkeit wird für Studien zu krankheitsmodifizierenden Wirkstoffen in der MS-Therapie eine Studiendauer von mindestens 2 Jahren gefordert, wobei für den Nachweis von Effekten auf die Schubaktivität auch ein Zeitraum von einem Jahr als adäquat angesehen wird, auch wenn längere Beobachtungszeiträume zu bevorzugen sind [1, 2]. Deswegen wird als minimale Studiendauer 1 Jahr, bzw. geplante Behandlungsdauer 48 Wochen, festgelegt. Die Aussagekraft von Studien mit einer noch kürzeren Studiendauer wird als nicht ausreichend hoch eingestuft.

4.1.5 Methoden zur Bewertung der Aussagekraft der Nachweise und zur Synthese von Ergebnissen

Zur Bewertung des Nutzens und Zusatznutzens von Teriflunomid werden Unterlagen der Evidenzstufe 1 (randomisierte kontrollierte Studien = Evidenzstufe 1b [3]) herangezogen, die die vorab definierten Einschlusskriterien erfüllen.

Zur Aufwertung der Beleglage für den Zusatznutzen werden neben dem direkten Vergleich in der TENERE-Studie indirekte Vergleiche über Placebo als Brückenkompator herangezogen. Die placebokontrollierten Teriflunomid- bzw. Rebif®-Studien werden hierfür gegebenenfalls durch Metaanalysen mit zufälligen Effekten zusammengefasst.

Da es sich bei allen eingeschlossenen Studien um randomisierte, kontrollierte Studien handelt, wird das Design und die Methodik der betreffenden Studien entsprechend den Items 2b bis 14 (inklusive zugehöriges Flow-Chart) des CONSORT-Statements dargestellt. Anhand der dargestellten Daten zum Design und der Methodik lässt sich die Qualität der Studien ableiten.

Das Verzerrungspotential der Studien (z. B. in Bezug auf Erzeugung der Randomisierungssequenz, Verdeckung der Gruppenzuteilung, Verblindung, ergebnisgesteuerte Berichterstattung, sonstige Aspekte) wurde einerseits auf Studienebene als auch

auf Endpunktebene bewertet. Auf Studienebene wurden alle eingeschlossenen Studien als niedrig verzerrt bewertet, auf Endpunktebene wurde das Verzerrungspotential für einzelne Endpunkte aufgrund fehlender Verblindung als hoch eingestuft; dies betrifft maßgeblich die Erfassung von patientenberichteten Endpunkten in der TENERE-Studie, in der die Patienten nicht verblindet waren.

4.1.6 Vorteil einer oralen gegenüber einer parenteralen Applikationsform

Teriflunomid wird einmal täglich als Filmtablette eingenommen. Die orale Darreichungsform alleine stellt schon einen patientenrelevanten Vorteil gegenüber parenteralen Alternativen dar. Dieser Vorteil ist jedoch in klinischen Studien sehr schwer nachzuweisen und in einem doppelblinden Studiendesign grundsätzlich nicht möglich, da die Patienten in jedem Fall wüssten, ob sie eine Tablette nehmen oder eine Injektion vornehmen, womit sie nicht verblindet sein können. Patientenberichtete Endpunkte wären in jedem Fall potentiell verzerrt.

In bester wissenschaftlicher Näherung lassen sich die Vorteile der oralen Applikationsform in den folgenden Endpunkten abbilden:

- Patientenzufriedenheit
- Reaktionen an der Injektionsstelle

4.1.7 Ergebnisse zum medizinischen Nutzen

Der medizinische Nutzen von Teriflunomid gegenüber Placebo wurde in zwei großen, qualitativ hochwertigen, randomisierten placebokontrollierten Studien geprüft (TEMPO und TOWER). Der primäre Endpunkt beider Studien waren die Schübe (operationalisiert als jährliche Schubrate). Der wesentliche sekundäre Endpunkt war die Behinderungsprogression (operationalisiert als Zeit bis zur Behinderungsprogression). In diesen beiden wichtigen Effektivitätseindpunkten konnte für Teriflunomid ein statistisch signifikanter Vorteil gezeigt werden.

Der medizinische Nutzen, der sich aus der aktivkontrollierten Studie TENERE ergibt, leitet sich aus dem therapeutisch relevanten Zusatznutzen gegenüber Interferon beta-1a, s. c, 44 µg (Rebif® 44 µg) ab. Dies ist ein etablierter Therapiestandard in der MS-Therapie und auch die für die Nutzenbewertung von Teriflunomid festgelegte zweckmäßige Vergleichstherapie. Der medizinisch relevante Zusatznutzen durch 14 mg Teriflunomid wird unter 4.1.8 im Detail ausgeführt.

4.1.7.1 Mortalität

In der TEMPO-Studie sind keine teilnehmenden Patienten verstorben. In der TOWER-Studie gab es über die Behandlungsgruppen verteilt 4 Todesfälle (2 in der 14 mg Teriflunomid-Gruppe und jeweils einer in der 7 mg Teriflunomid- und der Placebo-Gruppe). Ein direkter Zusammenhang mit der Studienmedikation ist für keinen dieser Fälle anzunehmen.

4.1.7.2 Morbidität

Die jährliche Schubrate wurde in beiden placebokontrollierten Studien signifikant um über 30 % reduziert (31,5 % (95 % KI: [15,3 %; 44,6 %]) in der TEMSO-Studie und 36,3 % (95 % KI: [20,7 %; 48,8 %]) in der TOWER-Studie). Für die Zielpopulation von Teriflunomid (erwachsene Patienten mit schubförmig-remittierender Multipler Sklerose) betrug die entsprechende Reduktion der Schubrate 35,6 % (Metaanalyse der Ergebnisse aus TEMSO und TOWER). Dies ist im vorliegenden Krankheitsbild (schubförmig-remittierende MS) als klinisch relevante Minderung der Schubrate anzusehen.

Auch in Bezug auf den zweiten patientenrelevanten klinischen Endpunkt, der Behinderungsprogression, konnte durch 14 mg Teriflunomid eine signifikante Verbesserung gegenüber Placebo gezeigt werden. In der TEMSO-Studie wurde die Behinderungsprogression gegenüber Placebo signifikant verlangsamt (Hazard Ratio von 0,702; 95 % KI: [0,506; 0,973]). In der TOWER-Studie zeigte sich ebenfalls ein Vorteil von Teriflunomid (Hazard Ratio von 0,685; 95 % KI: [0,467; 1,004]). Für die Zielpopulation von Teriflunomid (erwachsene Patienten mit schubförmig-remittierender Multipler Sklerose) betrug das entsprechende Hazard Ratio 0,696 (95 % KI: [0,539; 0,898]) mit einem zugehörigen p-Wert von 0,0053 (Metaanalyse der Ergebnisse aus TEMSO und TOWER bei guter Homogenität mit $I^2=0\%$ und $p_Q=0,7501$).

Hinsichtlich der Fatigue, einem sekundären Effektivitätspunkt, konnte durch die Therapie mit 14 mg Teriflunomid keine Verbesserung gegenüber Placebo festgestellt werden.

14 mg Teriflunomid hatte einen positiven Einfluss auf diverse MRT-Parameter, darunter auch auf die „Krankheitslast“ (Burden of disease), definiert als das Gesamtvolumen an abnormalem Gewebe im ZNS. Das Gesamtvolumen nahm unter 14 mg Teriflunomid signifikant weniger zu als unter Placebo (Mittelwertsdifferenz -0,089; 95 % KI: [-0,137; -0,041]). Über bildgebende Verfahren kann die Effektivität eines Wirkstoffes sehr sensitiv gezeigt werden. So konnte nachgewiesen werden, dass die entzündliche Aktivität im ZNS durch 14 mg Teriflunomid signifikant verringert wird.

4.1.7.3 Lebensqualität

Lebensqualität wurde als tertiärer Endpunkt mit den etablierten und validierten Messinstrumenten SF-36 (TEMSO und TOWER) und EQ-5D (TEMSO) bewertet. Es konnte kein signifikanter Vorteil von 14 mg Teriflunomid gegenüber Placebo gezeigt werden.

4.1.7.4 Patientenzufriedenheit

Patientenzufriedenheit wurde weder in TEMSO noch in TOWER erhoben.

4.1.7.5 Unerwünschte Ereignisse

Teriflunomid zeigte in beiden Studien eine gute Verträglichkeit. Weder die Gesamtrate an unerwünschten Nebenwirkungen (relatives Risiko von 1,038; 95 % KI: [0,998; 1,079] (Ergebnis der Metaanalyse von TEMSO und TOWER bei guter Homogenität mit $I^2 = 0\%$ und $p_Q = 0,9957$)), noch die Gesamtrate schwerwiegender unerwünschter Nebenwirkungen

(relatives Risiko von 1,110; 95 % KI: [0,853; 1,444] (Ergebnis der Metaanalyse von TEMSO und TOWER bei guter Homogenität mit $I^2 = 0$ % und $p_Q = 0,3550$)) unterschied sich signifikant von der Placebo-Gruppe. Lediglich die Anzahl an Studienabbruchern aufgrund von Nebenwirkungen war in der TOWER-Studie im Vergleich zu Placebo erhöht (relatives Risiko von 2,508; 95 % KI: [1,593; 3,948]). In der TEMSO-Studie konnte mit einem relativen Risiko von 1,352 und einem zugehörigen 95 %-Konfidenzintervall von [0,856; 2,137] auch in Bezug auf die unerwünschten Ereignisse, die zu einem Therapieabbruch führen, kein signifikanter Unterschied zwischen den Behandlungsarmen (14 mg Teriflunomid und Placebo) festgestellt werden.

Spezifische unerwünschte Ereignisse, die unter 14 mg Teriflunomid signifikant häufiger auftraten als unter Placebo, waren gastrointestinale Beschwerden, Neutropenie, periphere Neuropathie, eine verminderte Haardichte (HLT „Alopecias“) und eine Erhöhung der Alanin-Amino-Transferase-Werte (ALT-Werte). Bei den aufgetretenen Ereignissen, die dem High Level Term (HLT) „Alopecias“ zugeordnet werden, handelt es sich um einen überwiegend leichten Haarausfall innerhalb der ersten Behandlungsmonate. Nur in Einzelfällen führte dieser zu einem Therapieabbruch. Neutropenien und Leberwerte sind bei einer qualitätsgesicherten Anwendung entsprechend der Fachinformation zu überwachen (s. Modul 3.4) und stellen somit keine wirklichen Sicherheitsrisiken für die Behandlung mit 14 mg Teriflunomid dar.

4.1.7.6 Fazit zum medizinischen Nutzen von 14 mg Teriflunomid

14 mg Teriflunomid hat einen belegten Nutzen bei der Therapie erwachsener Patienten mit schubförmiger MS. Teriflunomid senkt die Schubrate in klinisch relevantem Ausmaß und hat einen positiven Effekt auf die Behinderungsprogression. Teriflunomid wird im Allgemeinen gut vertragen, so dass auch das Nutzen-Risiko-Verhältnis positiv ist.

4.1.8 Medizinischer Zusatznutzen von 14 mg Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie Interferon beta-1a, 44 µg, s. c., (Rebif® 44 µg)

Der medizinische Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie Interferon beta-1a, 44 µg, s. c., (Rebif® 44 µg) wurde in einer direkten Vergleichsstudie (TENERE) und in indirekten Vergleichen mit Placebo als Brückenkompator gezeigt.

4.1.8.1 TENERE

Tabelle 4-1: Einzelergebnisse aus denen sich der Zusatznutzen ableitet (aufgeschlüsselt nach Endpunkten) - direkter Vergleich (TENERE)

Direkter Vergleich (TENERE, 14 mg Teriflunomid vs. Rebif® 44 µg)		
Endpunkt	Effektschätzer 95 % KI []	p-Wert
Mortalität		
Gesamt mortalität	entfällt (keine Patienten mit dem Ereignis)	
Morbidität		

Direkter Vergleich (TENERE, 14 mg Teriflunomid vs. Rebif® 44 µg)		
Endpunkt	Effektschätzer 95 % KI []	p-Wert
Schübe (jährliche Schubrate)	RR 1,197 [0,623; 2,299]	p = 0,5896
Behinderungsprogression (Zeit bis zur Behinderungsprogression)	HR 0,982 [0,399; 2,420]	p = 0,9690
Therapieversagen (Time to Failure)	HR 0,861 [0,564; 1,314]	p = 0,5953
Fatigue	Mittelwertsdifferenz -5,003 [-12,31; 2,307]	p = 0,1789
Lebensqualität		
Lebensqualität (SF-36 oder EQ-5D)	entfällt, da in TENERE die entsprechende Erhebung fehlt	
Patientenzufriedenheit		
Allgemeine Zufriedenheit (TSQM)	Mittelwertsdifferenz 7,84 [1,458; 14,23]	p = 0,0162
Zufriedenheit in Bezug auf die Wirksamkeit (TSQM)	Mittelwertsdifferenz 3,835 [-3,082; 10,75]	p = 0,2759
Zufriedenheit in Bezug auf die Verträglichkeit (TSQM)	Mittelwertsdifferenz 21,77 [16,48; 27,07]	p < 0,0001
Zufriedenheit in Bezug auf die Anwenderfreundlichkeit (TSQM)	Mittelwertsdifferenz 27,96 [23,35; 32,56]	p < 0,0001
Unerwünschte Ereignisse		
Gesamtrate unerwünschter Ereignisse	RR 0,966 [0,904; 1,031]	p = 0,2946
Gesamtrate schwerwiegender unerwünschter Ereignisse	RR 0,787 [0,274; 2,264]	p = 0,6568
Gesamtrate unerwünschter Ereignisse, die zu einem Behandlungsabbruch führten	RR 0,501 [0,262; 0,959]	p = 0,0369
Reaktionen an der Injektionsstelle	RR nicht berechenbar, da das Ereignis unter 14 mg Teriflunomid nicht auftritt (orale Applikationsform)	
Grippeähnliche Symptome	RR 0,051 [0,017; 0,158]	p < 0,0001
Verminderung der Haardichte (HLT «Alopecias»)	RR 20,2 [2,773; 147,1]	p = 0,0030
Mit der Leberfunktion assoziierte unerwünschte Ereignisse	RR 0,25 [0,13; 0,45]	p < 0,0001

Direkter Vergleich (TENERE, 14 mg Teriflunomid vs. Rebif® 44 µg)		
Endpunkt	Effektschätzer 95 % KI []	p-Wert
Erhöhung der ALT-Werte	RR 0,326 [0,173; 0,613]	p = 0,0005
Durchfall	RR 2,640 [1,237; 5,632]	p = 0,0121
Übelkeit und Erbrechen	RR 2,295 [0,743; 7,089]	p = 0,1487
Neutropenie	RR 0,612 [0,178; 2,107]	p = 0,4364
Infektionen und Infestationen	RR 1,055 [0,795; 1,400]	p = 0,7108
Bluthochdruck	RR 1,148 [0,317; 4,156]	p = 0,8338
Periphere Neuropathie	RR 4,591 [0,546; 38,63]	p = 0,1608
Kopfschmerzen	RR 0,570 [0,338; 0,961]	p = 0,0349

4.1.8.1.1 Mortalität

In der TENERE-Studie sind keine Todesfälle aufgetreten.

4.1.8.1.2 Morbidität

Für den primären Endpunkt Therapieversagen („Time to Failure“) konnte keine signifikante Überlegenheit für 14 mg Teriflunomid nachgewiesen werden. Ein indirekter Vergleich ist nicht möglich, da dieser Endpunkt von den eingeschlossenen Studien nur in TENERE erhoben worden ist.

In der TENERE-Studie ließ sich hinsichtlich der wichtigen Effektivitätseindpunkte Schübe und Behinderungsprogression keine signifikante Überlegenheit von Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie Rebif® 44 µg zeigen.

Auch in Bezug auf die Fatigue, einem sekundären Effektivitätseindpunkt, konnte durch die Therapie mit 14 mg Teriflunomid keine Verbesserung gegenüber Rebif® 44 µg festgestellt werden.

4.1.8.1.3 Lebensqualität

Lebensqualität ist in der TENERE-Studie nicht erhoben worden.

4.1.8.1.4 Patientenzufriedenheit

Mit dem für die Patientenzufriedenheit validierten Messinstrument TSQM [4, 5] (treatment satisfaction questionnaire for medication) konnte eine signifikant größere Zufriedenheit der

Patienten mit 14 mg Teriflunomid als mit Rebif[®] 44 µg gezeigt werden (mittlere Differenz des Scores für allgemeine Zufriedenheit von 7,843; 95 % KI: [1,458; 14,23]). Für die weiteren Dimensionen ergab sich dabei folgendes Bild:

- Kein signifikant besseres Ergebnis im Bereich Wirksamkeit (Mittelwertsdifferenz 3,835; 95 % KI: [-3,082; 10,75])
- Signifikant besseres Abschneiden von 14 mg Teriflunomid im Bereich Verträglichkeit (Mittelwertsdifferenz 21,77; 95 % KI: [16,48; 27,07])
- Signifikant besseres Abschneiden von 14 mg Teriflunomid im Bereich Anwenderfreundlichkeit (Mittelwertsdifferenz 27,96; 95 % KI: [23,35; 32,56])

Das Ergebnis in den einzelnen Dimensionen des TSQM ist konsistent mit den Studienergebnissen, die sich für Teriflunomid bei der Auswertung anderer Endpunkte ergeben haben:

- Die Patienten nehmen keine signifikante Verbesserung im Bereich der Wirksamkeit wahr, was mit einer vergleichbaren Effektivität von 14 mg Teriflunomid und Rebif[®] 44 µg bezogen auf die Morbiditätsendpunkte Schübe und Behinderungsprogression korreliert.
- Die Patienten empfinden 14 mg Teriflunomid als besser verträglich als die zweckmäßige Vergleichstherapie. Dies ist in guter Übereinstimmung mit einer geringeren Rate an unerwünschten Ereignissen, die zu einem Therapieabbruch führten, und der bedeutsamen Vermeidung unerwünschter Ereignisse von besonderem Interesse (s. 4.1.8.1.5).
- Die Patienten bewerten 14 mg Teriflunomid als deutlich besser in der Anwenderfreundlichkeit, wodurch die Relevanz der oralen Applikationsform für die Patienten unterstrichen wird.

Das zeigt, dass die Patienten mit dem zu bewertenden Arzneimittel 14 mg Teriflunomid insgesamt zufriedener waren als mit der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg.

4.1.8.1.5 Unerwünschte Ereignisse

Sowohl Teriflunomid als auch Interferon beta-1a sind gut verträgliche Wirkstoffe. Die Gesamtrate unerwünschter Ereignisse (relatives Risiko von 0,966; 95 % KI: [0,904; 1,031]) und die Gesamtrate schwerwiegender unerwünschter Ereignisse (relatives Risiko von 0,787; 95 % KI: [0,274; 2,264]) waren zwischen den Behandlungsgruppen nicht signifikant unterschiedlich. Jedoch gab es in der Rebif[®] 44 µg-Gruppe signifikant mehr Patienten, die die Studienmedikation abbrechen, als in der 14 mg Teriflunomid-Gruppe, das relative Risiko betrug 0,501 (95 % KI: [0,262; 0,959]), d. h. das Risiko, aufgrund von Nebenwirkungen die Therapie abzubrechen, war unter Rebif[®] 44 µg doppelt so hoch wie unter 14 mg Teriflunomid.

Neutropenie und eine Erhöhung der ALT-Werte sind für 14 mg Teriflunomid, jedoch auch für Interferon beta-1a, 44 µg, s. c., (Rebif[®] 44 µg) häufige unerwünschte Ereignisse, die kritisch zu beobachten sind. Im direkten Vergleich (TENERE) zeigte sich bezüglich der Erhöhung der ALT-Werte ein deutlicher, statistisch signifikanter Vorteil für Teriflunomid (relatives Risiko von 0,326 (95 % KI: [0,173; 0,613])); das Ergebnis bezogen auf mit der Leberfunktion assoziierte unerwünschte Ereignisse ist ähnlich (relatives Risiko von 0,25 (95 % KI: [0,13; 0,45])).

Ereignisse, die dem High Level Term (HLT) „Alopecias“ (Verminderung der Haardichte) zuzuordnen sind, traten unter 14 mg Teriflunomid häufiger auf (relatives Risiko von 20,2; 95 % KI: [2,773; 147,1]), dagegen leiden die Rebif[®] 44 µg - Patienten öfter an Kopfschmerzen (relatives Risiko von 0,570; 95 % KI: [0,338; 0,961]), grippeähnlichen Symptomen (relatives Risiko 0,051; 95 % KI: [0,017; 0,158]) und Reaktionen an der Injektionsstelle (21,8 % der Patienten). Reaktionen an der Injektionsstelle, die unter und Rebif[®] 44 µg sehr häufig sind, treten bei der Therapie mit 14 mg Teriflunomid naturgemäß nicht auf, da es sich um eine Filmtablette handelt.

MS als chronische Erkrankung erfordert eine langfristige Therapie. Das reduzierte Risiko für Therapieabbrüche aufgrund von Nebenwirkungen ist ein patientenrelevanter Vorteil von 14 mg Teriflunomid gegenüber der etablierten Therapie. In Bezug auf unerwünschte Ereignisse von besonderem Interesse ergibt sich insgesamt eine bedeutsame Vermeidung unerwünschter Ereignisse durch die Therapie mit 14 mg Teriflunomid. Damit hat Teriflunomid einen beträchtlichen patientenrelevanten Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg.

4.1.8.2 Indirekte Vergleiche

Die direkt vergleichende TENERE-Studie war auf den Nachweis der Überlegenheit in Bezug auf die Zeit bis zu dem kombinierten Endpunkt Therapieversagen angelegt. Ein vorgeschalteter Nachweis von Nicht-Unterlegenheit war nicht geplant. Da die placebo-kontrollierten Studien TEMSO und TOWER deutlich größer angelegt waren als TENERE, wurden indirekte Vergleiche angestellt, um die Konsistenz mit den Ergebnissen aus TENERE zu prüfen und gegebenenfalls die Präzision der in TENERE geschätzten Abstände zu verbessern.

Dazu wurden jeweils die Ergebnisse der zwei placebokontrollierten Teriflunomid-Studien TEMSO und TOWER in einer Metaanalyse mit zufälligen Effekten zusammengefasst und anschließend mit den Ergebnissen der placebokontrollierten Interferon beta-1a, s. c., (Rebif[®])-Studie PRISMS in einem indirekten Vergleich nach Bucher [6] kombiniert (siehe dazu Abbildung 4-1). Anschließend erfolgte die metaanalytische Zusammenfassung des Ergebnisses aus TENERE mit dem Ergebnis des indirekten Vergleichs zur Prüfung der Konsistenz und gegebenenfalls der Verbesserung der Präzision geschätzter Abstände. Im Folgenden werden die Ergebnisse der zwei Metaanalysen und der indirekten Vergleiche tabellarisch dargestellt und in den Unterpunkten 4.1.8.2.1 bis 4.1.8.2.5 zusammenfassend beschrieben.

Tabelle 4-2: Einzelergebnisse aus denen sich der Zusatznutzen ableitet (aufgeschlüsselt nach Endpunkten) - indirekte Vergleiche

Indirekte Vergleiche (14 mg Teriflunomid vs. Rebif® 44 µg)				
	Metaanalyse TEMZO/ TOWER	PRISMS	Indirekter Vergleich	Metaanalyse TENERE/ Indirekter Vergleich
Endpunkt	Effektschätzer 95 % KI [] p-Wert Heterogenitäts- maße	Effektschätzer 95 % KI [] p-Wert Heterogenitäts- maße	Effektschätzer 95 % KI [] p-Wert	Effektschätzer 95 % KI [] p-Wert Heterogenitäts- maße
Mortalität				
Gesamt- mortalität	Risikodifferenz 0,0007 [-0,0039; 0,0054] p = 0,7539 I ² =0%, p _Q =0,6032	Risikodifferenz -0,0053 [-0,0201; 0,0094] p = 0,4775	Risikodifferenz 0,0061 [-0,0094; 0,0216] p = 0,4403	Risikodifferenz 0,0035 [-0,0098; 0,0168] p = 0,6094 I ² =0%, p _Q = 0,6195
Morbidität				
Schübe (jährliche Schubrate)	Rate Ratio 0,644 [0,552; 0,751] p < 0,0001 ^a I ² =0%, p _Q =0,7444	Rate Ratio 0,67 [0,56; 0,79] p < 0,0001 ^a	Rate Ratio 0,961 [0,763; 1,211] p = 0,7369	Rate Ratio 0,985 [0,792; 1,224] p = 0,8911 I ² = 0 %; p _Q = 0,5349
Behinderungs- progression (Zeit bis zur Behinderungs- progression)	Hazard Ratio 0,696 ^b [0,539; 0,898] p = 0,0053 I ² =0%, p _Q =0,7501	Hazard Ratio 0,62 ^b [0,43; 0,91] p < 0,05	Hazard Ratio 1,122 [0,713; 1,766] p = 0,6195 ^c	Hazard Ratio 1,092 [0,728; 1,638] p = 0,6701 I ² = 0 %; p _Q = 0,7965
Therapie- versagen (Time to Failure)	Entfällt, da außer in TENERE die entsprechenden Erhebungen fehlen			
Fatigue	Entfällt, da die Fatigue in PRISMS nicht mit dem validierten Messinstrument FIS erhoben worden ist			
Lebensqualität				
Lebensqualität (SF-36 oder EQ-5D)	Entfällt, da in PRISMS die entsprechenden Erhebungen fehlen			
Patientenzufriedenheit				
TSQM	Entfällt, da außer in TENERE die entsprechenden Erhebungen fehlen			
Unerwünschte Ereignisse				
Gesamtrate unerwünschter	Entfällt, da das entsprechende Ereignis in PRISMS nicht berichtet ist			

Indirekte Vergleiche (14 mg Teriflunomid vs. Rebif® 44 µg)				
	Metaanalyse TEMZO/ TOWER	PRISMS	Indirekter Vergleich	Metaanalyse TENERE/ Indirekter Vergleich
Endpunkt	Effektschätzer 95 % KI [] p-Wert Heterogenitäts- maße	Effektschätzer 95 % KI [] p-Wert Heterogenitäts- maße	Effektschätzer 95 % KI [] p-Wert	Effektschätzer 95 % KI [] p-Wert Heterogenitäts- maße
Ereignisse				
Gesamtrate schwerwiegender unerwünschter Ereignisse	Entfällt, da das entsprechende Ereignis in PRISMS nicht berichtet ist			
Gesamtrate unerwünschter Ereignisse, die zu einem Behandlungsabbruch führten	Relatives Risiko 1,843 [1,006; 3,376] p = 0,0477 I ² =72%; p _Q =0,0604 Aber deutliche gleichgerichtete Effekte.	Relatives Risiko 4,573 [1,002; 20,88] p = 0,0498	Relatives Risiko 0,403 [0,079; 2,067] p = 0,2759	Relatives Risiko 0,486 [0,266; 0,889] p = 0,0192 I ² =0%, p _Q =0,8086
Reaktionen an der Injektionsstelle	RR nicht berechenbar, da das Ereignis unter 14 mg Teriflunomid nicht auftritt (orale Applikationsform)			
Grippeähnliche Symptome	Relatives Risiko 0,908 [0,467; 1,767] p = 0,7763 I ² = 0%, p _Q =0,9630	Relatives Risiko 1,129 [0,798; 1,599] p = 0,4931	Relatives Risiko 0,804 [0,380; 1,704] p = 0,5693	Relatives Risiko 0,210 [0,014; 3,122] p = 0,2568 I ² =94%, p _Q <0,0001 Keine deutlichen gleichgerichteten Effekte.
Alopezie	Entfällt, da das entsprechende Ereignis in PRISMS nicht berichtet ist			
ALT-Wert-Erhöhung	Relatives Risiko 1,875 [1,376; 2,555] p < 0,0001 I ² =0%, p _Q =0,4560	Relatives Risiko 6,098 [1,384; 26,87] p = 0,0169	Relatives Risiko 0,308 [0,068; 1,399] p = 0,1271	Relatives Risiko 0,323 [0,180; 0,579] p = 0,0001 I ² =0%, p _Q =0,9449
Gastrointestinale Beschwerden	Entfällt, da das entsprechende Ereignis in PRISMS nicht berichtet ist			
Neutropenie	Entfällt, da das entsprechende Ereignis in PRISMS nicht berichtet ist			

Indirekte Vergleiche (14 mg Teriflunomid vs. Rebif® 44 µg)				
	Metaanalyse TEMZO/ TOWER	PRISMS	Indirekter Vergleich	Metaanalyse TENERE/ Indirekter Vergleich
Endpunkt	Effektschätzer 95 % KI [] p-Wert Heterogenitäts- maße	Effektschätzer 95 % KI [] p-Wert Heterogenitäts- maße	Effektschätzer 95 % KI [] p-Wert	Effektschätzer 95 % KI [] p-Wert Heterogenitäts- maße
Lymphopenie	Relatives Risiko 1,791 [0,377; 8,502] p = 0,4632 I ² =0%, p _Q =0,6981	Relatives Risiko 3,485 [1,539; 7,888] p = 0,0027	Relatives Risiko 0,514 [0,089; 2,984] p = 0,4584	Relatives Risiko 0,456 [0,098; 2,125] p = 0,3171 I ² =0%, p _Q =0,7805
Leukopenie	Relatives Risiko 4,580 [0,992; 21,16] p = 0,0513 I ² =0%, p _Q =0,8705	Relatives Risiko 5,082 [1,496; 17,26] p = 0,0092	Relatives Risiko 0,901 [0,127; 6,391] p = 0,9172	Relatives Risiko 0,910 [0,229; 3,612] p = 0,8931 I ² =0%, p _Q =0,9895
Infektionen und Infestationen	Entfällt, da das entsprechende Ereignis in PRISMS nicht berichtet ist			
Bluthochdruck	Entfällt, da das entsprechende Ereignis in PRISMS nicht berichtet ist			
Periphere Neuropathie	Entfällt, da das entsprechende Ereignis in PRISMS nicht berichtet ist			
Kopfschmerzen	Relatives Risiko 1,026 [0,822; 1,281] p = 0,8209 I ² =0%, p _Q =0,8284	Relatives Risiko 1,029 [0,820; 1,291] p = 0,8073	Relatives Risiko 0,997 [0,726; 1,371] p = 0,9871	Relatives Risiko 0,785 [0,456; 1,350] p = 0,3813 I ² =69%, p _Q =0,0729 Keine deutlichen gleichgerichteten Effekte.

^aH₀: Rate Ratio = 1^bH₀: Hazard Ratio = 1

4.1.8.2.1 Mortalität

Die Gesamtzahl der Todesfälle in den Studien ist erwartungsgemäß niedrig. Es konnte kein Unterschied zwischen den Interventionen festgestellt werden.

4.1.8.2.2 Morbidität

Die indirekten Vergleiche unterstützen die Aussage, die sich aus der direkt vergleichenden Studie TENERE ableitet, nämlich, dass es sich bei Teriflunomid und Interferon beta-1a, 44 µg, s. c., (Rebif® 44 µg) um Wirkstoffe handelt, die bei erwachsenen Patienten mit RRMS die jährliche Schubrate und die Behinderungsprogression in vergleichbarem Ausmaß senken.

Die Metaanalyse von TEMSO und TOWER zeigt mit 35,6 % (95 % KI: [24,9 %; 44,8 %]) eine ähnliche Reduktion der jährlichen Schubrate bei RRMS-Patienten wie die PRISMS-Studie (33 % (95 % KI: [21 %; 44 %])). Im indirekten Vergleich ergibt sich daraus eine Rate Ratio von 0,961 (95 % KI: [0,763;1,211]). Durch den indirekten Vergleich konnte die Präzision der Schätzung aus TENERE deutlich verbessert werden.

Für die Zeit bis zur Behinderungsprogression ergibt sich für die RRMS-Patienten ein ähnliches Bild. Der Wert für die Hazard Ratio von 0,696 (95 % KI: [0,539; 0,898]), der sich aus der Metaanalyse von TEMSO und TOWER ergibt, ist der Hazard Ratio in der PRISMS-Studie (0,62; 95 % KI: [0,43; 0,91]) ähnlich. Im indirekten Vergleich ergibt sich daraus ein Wert von 1,122 (95 % KI: [0,713; 1,766]) und damit kein nachweisbarer Unterschied zwischen den beiden Interventionen. Auch in diesem Endpunkt kann die Präzision der Schätzung aus TENERE durch den indirekten Vergleich deutlich verbessert werden.

Ein indirekter Vergleich bezüglich des Endpunktes Fatigue ist nicht möglich, da die Fatigue in der Rebif[®] 44 µg-Studie PRISMS nicht mit dem validierten Messinstrument FIS [7, 8] erhoben worden ist.

4.1.8.2.3 Lebensqualität

Gesundheitsbezogene Lebensqualität wurde in den placebokontrollierten Teriflunomid-Studien mit den etablierten Messinstrumenten SF-36 (TEMSO und TOWER) und EQ-5D (TEMSO) bewertet. Eine entsprechende Erhebung fand sich in der Rebif[®] 44 µg-Studie PRISMS nicht, so dass die Durchführung eines indirekten Vergleichs für diesen Endpunkt nicht möglich ist.

4.1.8.2.4 Patientenzufriedenheit

Patientenzufriedenheit wurde von den zur Nutzenbewertung eingeschlossenen Studien lediglich in der TENERE-Studie erhoben, ein indirekter Vergleich ist somit nicht durchführbar.

4.1.8.2.5 Unerwünschte Ereignisse

Einige unerwünschte Ereignisse (Gesamtrate unerwünschter Ereignisse, Gesamtrate schwerwiegender unerwünschter Ereignisse, Reaktionen an der Injektionsstelle, verminderte Haardichte (Alopezie), Gastrointestinale Beschwerden, Neutropenie, Überempfindlichkeit/Hautreaktionen, Infektionen und Infestationen, Bluthochdruck und periphere Neuropathie) konnten in indirekten Vergleichen nicht berücksichtigt werden, weil sie entweder unter Teriflunomid nicht auftreten konnten (Reaktionen an der Injektionsstelle) oder aber in der PRISMS-Studienpublikation nicht dargestellt sind. Die Ergebnisse der unerwünschten Ereignisse, die sich auswerten lassen, sind für die Mehrheit der patientenrelevanten Endpunkte konsistent mit den Ergebnissen der TENERE-Studie. Durch den indirekten Vergleich wird der Vorteil für 14 mg Teriflunomid bezüglich der Gesamtrate unerwünschter Ereignisse, die zum Behandlungsabbruch führten, und der ALT-Wert-Erhöhung belegt.

4.1.8.3 Fazit zum medizinischen Zusatznutzen von 14 mg Teriflunomid

Für das zu bewertende Arzneimittel Teriflunomid 14 mg lässt sich aus der vorliegenden Evidenz eine mit der zweckmäßigen Vergleichstherapie Interferon beta-1a (Rebif® 44 µg) vergleichbare Effektivität ableiten, so dass sich in diesem Bereich kein Zusatznutzen zeigen lässt.

Der medizinisch relevante Zusatznutzen ergibt sich aus folgenden patientenrelevanten Endpunkten:

- der erheblich größeren Patientenzufriedenheit,
- der statistisch signifikant niedrigeren Therapieabbruchrate aufgrund von unerwünschten Ereignissen und
- der bedeutsamen Vermeidung nicht schwerwiegender unerwünschter Ereignisse, u. a. Reaktionen an der Injektionsstelle und grippeähnliche Symptome.

4.1.9 Ausmaß des Zusatznutzens und Beleglage

Der neue Wirkstoff hat als orale Behandlungsoption einen beträchtlichen Zusatznutzen in der krankheitsmodifizierenden Therapie der schubförmig-remittierenden Multiplen Sklerose. Dieser beträchtliche Zusatznutzen wird wie oben ausgeführt in den klinischen Studien abgebildet in statistisch signifikanten Ergebnissen für

- eine größere Patientenzufriedenheit,
- weniger unerwünschte Ereignisse, die zu einem Therapieabbruch führten, und
- der Vermeidung spezifischer unerwünschter Ereignisse.

Die Ergebnisse für die einzelnen Endpunkte und der resultierende Zusatznutzen, der beansprucht wird, sind in den folgenden Abschnitten 4.1.9.1 bis 4.1.9.1.5 ausführlich zusammengestellt.

4.1.9.1 Beschreibung des Zusatznutzens auf Endpunktebene

4.1.9.1.1 Mortalität

Im Bereich Mortalität konnte für 14 mg Teriflunomid weder ein patientenrelevanter Vorteil noch ein größeres Risiko abgeleitet werden.

4.1.9.1.2 Morbidität

Wie unter 4.1.8 ausgeführt ergibt sich aus der vorhanden Evidenz eine vergleichbare Effektivität des zu bewertenden Arzneimittels 14 mg Teriflunomid (AUBAGIO®) und Interferon beta-1a (Rebif® 44 µg). In einer direkt vergleichenden Studie (TENERE) konnte kein Wirksamkeitsunterschied zwischen den Interventionen nachgewiesen werden. Die

vergleichbare Effektivität leitet sich ebenfalls aus indirekten Vergleichen mit Placebo als Brückenkomparator ab.

In Bezug auf die Effektivität wird für das zu bewertende Arzneimittel kein Zusatznutzen beansprucht.

4.1.9.1.3 Lebensqualität

Für den Zusatznutzen von Teriflunomid 14 mg im Bereich der gesundheitsbezogenen Lebensqualität gibt es keinen Beleg.

4.1.9.1.4 Patientenzufriedenheit

In Bezug auf die Patientenzufriedenheit ergibt sich ein beträchtlicher Zusatznutzen.

Für MS-Patienten, die einer Dauertherapie mit krankheitsmodifizierenden Wirkstoffen zum Teil über Jahrzehnte ausgesetzt sind, ist eine bessere Verträglichkeit und Handhabung ihres angewendeten Arzneimittels von zentraler Bedeutung. Auch aus Sicht der Therapeuten ist es wichtig, dass Compliance und Therapieadhärenz sichergestellt sind, da die Wirkstoffe nur so effektiv greifen und die Therapieziele erreicht werden können. Die notwendige Compliance und Therapieadhärenz sind auf Dauer nur mit Wirkstoffen zu erreichen, die von dem Patienten selbst als anwenderfreundlich und gut verträglich empfunden werden. Im Anwendungsgebiet A (krankheits-modifizierende Therapie der schubförmig-remittierenden Multiplen Sklerose) ist die größere Zufriedenheit der Patienten resultierend aus einer besseren Anwenderfreundlichkeit und der Vermeidung von unerwünschten Ereignissen ausgesprochen bedeutsam. Damit liegt eine „bedeutsame Vermeidung anderer [d. h. nicht schwerwiegender] Nebenwirkungen“ [9] vor. Nach §5 Absatz 7 Satz 2 AM-NutzenV [9] resultiert daraus ein beträchtlicher Zusatznutzen.

Dieser wird als Hinweis eingestuft, weil entsprechende Daten lediglich aus der TENERE-Studie vorliegen

4.1.9.1.5 Unerwünschte Ereignisse

Die Behandlung mit Teriflunomid führte zu signifikant weniger unerwünschten Ereignissen, die zu einem Therapieabbruch führten, als die Behandlung mit der zweckmäßigen Vergleichstherapie Interferon beta-1a, s.c, dreimal pro Woche (Rebif[®] 44 µg). Dies zeigt einen beträchtlichen Zusatznutzen des neuen Wirkstoff in der Therapie von erwachsenen Patienten mit RRMS. Der Unterschied in Bezug auf diesen Endpunkt war statistisch signifikant. Im indirekten Vergleich ergab sich ein Effektschätzer (relatives Risiko) zugunsten von Teriflunomid von ähnlicher Größe, der jedoch nicht signifikant war. Dementsprechend kann der Zusatznutzen in Bezug auf die unerwünschten Ereignisse, die zu einem Therapieabbruch führten, aus der vorhandenen Evidenz lediglich als gering eingestuft werden.

Im speziellen Anwendungsgebiet von Teriflunomid (AUBAGIO[®]) ist die Vermeidung von nicht schwerwiegenden unerwünschten Ereignissen von großer Bedeutung für den Therapieerfolg, da es sich sowohl bei Teriflunomid als auch bei der zweckmäßigen Vergleichstherapie um dauerhaft bzw. lebenslang anzuwendende Wirkstoffe handelt. Aus

einer bedeutsamen Vermeidung nicht-schwerwiegender unerwünschter Ereignisse resultiert nach §5 Absatz 7 Satz 2 AM-NutzenV [9] ein beträchtlicher Zusatznutzen.

Die unerwünschten Ereignisse, die unter 14 mg Teriflunomid häufiger auftraten als unter Rebif[®] 44 µg –Therapie, waren von milder oder moderater Natur und eine auftretende Verminderung der Haardichte bildet sich in der Regel auch unter Behandlung mit Teriflunomid ohne Therapie wieder zurück. Dementsprechend führt neben der oralen Darreichungsform auch die Verträglichkeit von Teriflunomid zu einer für die Patienten bedeutend angenehmeren Anwendung als die Behandlung mit der zweckmäßigen Vergleichstherapie, die nicht nur mit den Injektionen selbst, sondern auch mit Reaktionen an der Injektionsstelle, Kopfschmerzen und grippeähnlichen Symptomen verbunden ist.

Der patientenrelevante Zusatznutzen bezüglich der unerwünschten Ereignisse ist durch die zusätzlichen indirekten Vergleiche belegt.

Als Resultat wird in Bezug auf die bedeutsame Vermeidung unerwünschter Ereignisse ein Beleg für einen beträchtlichen Zusatznutzen beansprucht.

4.1.9.2 Ausmaß des Zusatznutzens - Fazit

Aus den oben zusammengefassten Ergebnissen ergibt sich ein Beleg für einen beträchtlichen Zusatznutzen für 14 mg Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg.

Tabelle 4-3: Einzelergebnisse aus denen sich der Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie ableitet (aufgeschlüsselt nach Endpunkten)

Endpunkt	Ausmaß des Zusatznutzens	Beleglage
Mortalität		
Mortalität	Zusatznutzen nicht belegt	-
Morbidität		
Schübe	Zusatznutzen nicht belegt	-
Behinderungsprogression	Zusatznutzen nicht belegt	-
Therapieversagen	Zusatznutzen nicht belegt	-
Fatigue	Zusatznutzen nicht belegt	-
Lebensqualität		
EQ-5D	Zusatznutzen nicht belegt	-
SF-36	Zusatznutzen nicht belegt	-
Patientenzufriedenheit		
Patientenzufriedenheit	Beträchtlicher Zusatznutzen	Hinweis

Endpunkt	Ausmaß des Zusatznutzens	Beleglage
(TSQM)		
Unerwünschte Ereignisse		
Gesamtrate unerwünschter Ereignisse	Zusatznutzen nicht belegt	-
Gesamtrate schwerwiegender unerwünschter Ereignisse	Zusatznutzen nicht belegt	-
Gesamtrate unerwünschter Ereignisse, die zu einem Behandlungsabbruch führten	Geringer Zusatznutzen	Beleg
Unerwünschte Ereignisse von besonderem Interesse ^a	Beträchtlicher Zusatznutzen	Beleg

^aeine aufgeschlüsselte Darstellung der unerwünschten Ereignisse von besonderem Interesse findet sich in den entsprechenden Abschnitten der Module 4.3.1 und 4.3.2

4.1.9.3 Schlussfolgerungen zum Zusatznutzen und zum therapeutisch bedeutsamen Zusatznutzen

Aus der vorhandenen Evidenz lässt sich insgesamt ein **Beleg für einen beträchtlichen Zusatznutzen** ableiten.

Tabelle 4-4: Schlussfolgerung zum therapeutisch bedeutsamen Zusatznutzen

Zu bewertendes Arzneimittel	Ausmaß des Zusatznutzens	Beleglage
Teriflunomid 14 mg (AUBAGIO®)	beträchtlich	Beleg

4.2 Methodik

Abschnitt 4.2 soll die Methodik der im Dossier präsentierten Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens beschreiben. Der Abschnitt enthält Hilfestellungen für die Darstellung der Methodik sowie einige Vorgaben, die aus den internationalen Standards der evidenzbasierten Medizin abgeleitet sind. Eine Abweichung von diesen methodischen Vorgaben ist möglich, bedarf aber einer Begründung.

4.2.1 Fragestellung

Nach den internationalen Standards der evidenzbasierten Medizin soll eine Bewertung unter einer definierten Fragestellung vorgenommen werden, die mindestens folgende Komponenten enthält:

- Patientenpopulation
- Intervention
- Vergleichstherapie
- Endpunkte
- Studientypen

Geben Sie die Fragestellung der vorliegenden Aufarbeitung von Unterlagen zur Untersuchung des medizinischen Nutzens und des medizinischen Zusatznutzens des zu bewertenden Arzneimittels an. Begründen Sie Abweichungen von den oben beschriebenen Vorgaben.

Folgende Fragestellung liegt der vorliegenden Analyse zugrunde:

Hat die Behandlung mit 14 mg Teriflunomid, einmal täglich, oral, bei erwachsenen Patienten mit schubförmig-remittierender Multipler Sklerose (RRMS) hinsichtlich patientenrelevanter Endpunkte gegenüber der zweckmäßigen Vergleichstherapie Interferon beta-1a, 44 µg, s. c., dreimal pro Woche (Rebif® 44 µg) einen Zusatznutzen im Sinne des § 35a SGB V?

Zielpopulation (Patientenpopulation) bezüglich dieser Fragestellung sind erwachsene Patienten mit schubförmig-remittierender Multipler Sklerose (RRMS) und einem EDSS≤5,5. Erwachsene Patienten sind MS-Patienten, die das 18. Lebensjahr erreicht haben oder älter sind. Die Patientenpopulation schließt sowohl weibliche als auch männliche Patienten ein.

Intervention ist der selektive, reversible Inhibitor der Dihydroorotat-Dehydrogenase Teriflunomid. Teriflunomid wird in der Dosierung von 14 mg einmal täglich als Filmtablette oral verabreicht.

Die zweckmäßige Vergleichstherapie, die der Gemeinsame Bundesausschuss gemeinsam mit dem pharmazeutischen Unternehmer im Beratungsgespräch vom 04.07.2012 und in anschließender weiterer Korrespondenz vom 30.10.2012 (Vorgangsnummer 2012-B-023) festgelegt hat, ist Beta-Interferon (Rebif® 44 µg) [10, 11]. Dies entspricht dem Therapiestandard im Bereich der krankheitsmodifizierenden Therapie der schubförmig-remittierenden Multiplen Sklerose. Eine ausführliche Darlegung ist in Modul 3.1 erfolgt.

Die Bewertung erfolgt anhand patientenrelevanter Endpunkte, die in Kapitel 4.2.2 dargestellt sind.

Zur Bewertung werden randomisierte kontrollierte Studien herangezogen.

4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung

Die Untersuchung der in Abschnitt 4.2.1 benannten Fragestellung soll auf Basis von klinischen Studien vorgenommen werden. Für die systematische Auswahl von Studien für diese Untersuchung sollen Ein- und Ausschlusskriterien für die Studien definiert werden.

Benennen Sie die Ein- und Ausschlusskriterien für Studien zum medizinischen Nutzen und Zusatznutzen. Machen Sie dabei mindestens Aussagen zur Patientenpopulation, zur Intervention, zur Vergleichstherapie, zu den Endpunkten, zum Studientyp und zur Studiendauer und begründen Sie diese. Stellen Sie die Ein- und Ausschlusskriterien zusammenfassend in einer tabellarischen Übersicht dar.

4.2.2.1 Grundlagen zur Festlegung von Ein- und Ausschlusskriterien

Die Ein- und Ausschlusskriterien für Studien, die für die Ableitung von Aussagen zum Nutzen und Zusatznutzen von Teriflunomid herangezogen werden, orientieren sich an der unter 4.2.1 definierten Fragestellung. Im vorliegenden Dossier sollen der medizinische Nutzen und der Zusatznutzen sowohl anhand direkter als auch anhand indirekter Evidenz nachgewiesen werden. Für die verschiedenen Herangehensweisen werden die Ein- und Ausschlusskriterien getrennt ausgewiesen. In jedem Fall müssen in der Definition der Ein- und Ausschlusskriterien folgende Parameter (gemäß PICOS=population, intervention, comparison, outcome, study design [12]) berücksichtigt werden:

- Patientenpopulation
- Intervention
- Vergleichstherapie (Komparator/Comparison)
- Endpunkte (Zielgrößen/Outcomes)
- Studientypen

Allgemeine Angaben und Einschätzungen zu diesen Parametern in Studien zu krankheitsmodifizierenden Wirkstoffen in der MS-Therapie werden im Folgenden dargestellt. Die Einschätzungen erfolgten in Bezug auf bzw. sind übereinstimmend mit den Empfehlungen des CHMP (Committee for Medicinal Products for Human Use) der EMA. Aktuell existiert eine Leitlinie von 2007 (Zeitpunkt des Beginns der Gültigkeit) [1] und ein Entwurf einer aktualisierten Fassung [2], die beide Berücksichtigung finden, da der Entwurf erst nach Einreichen des vorliegenden Dokuments in der endgültigen Fassung vorliegen wird, jedoch den aktuelleren Wissensstand darstellt.

4.2.2.1.1 Patientenpopulation

4.2.2.1.1.1 Diagnosekriterien

In Studien zur krankheitsmodifizierenden Therapie der MS sollten nur Patienten eingeschlossen sein, die auch tatsächlich an dieser Erkrankung leiden. Deshalb sind nur solche Studien relevant, die Patienten mit einer gesicherten MS-Diagnose (McDonald-Kriterien bzw. Poser-Kriterien) einschließen [13-16].

4.2.2.1.1.2 Krankheitsschwere

Die Schwere der permanenten Behinderung sollte eine Erfassung von Schüben nicht ungenau bis unmöglich machen (unabdingbare Voraussetzung für die Erfassung schubbezogener Endpunkte). Darüber hinaus sollte die betrachtete Patientenpopulation hinreichend homogen sein. Deswegen muss die Patientenpopulation nach Grad der Behinderung (EDSS-Score) vorab charakterisiert werden. Eine gängige Einteilung erfolgt nach dem Kriterium der Gehfähigkeit, so dass Studien mit ausnahmslos Patienten, die bereits auf Gehhilfen angewiesen sind (EDSS>5,5), ausgeschlossen werden (allgemeine Erläuterungen zum EDSS finden sich in Modul 3.2.2.1.2).

4.2.2.1.1.3 Verlaufsform

Es können nur Studien berücksichtigt werden, deren Ergebnisse für die Bewertung krankheitsmodifizierender Therapie schubförmig-remittierender Multipler Sklerose aussagekräftig sind. Das bedeutet, dass der Großteil der Patientenkollektive durch die Zulassung von Teriflunomid abgedeckt werden muss (Anwendungsgebiet A: Behandlung von erwachsenen Patienten mit schubförmig-remittierender MS (RRMS)). Studien, deren Patientenkollektive lediglich Patienten mit SPMS, PPMS oder CIS enthalten, sind ausgeschlossen.

4.2.2.1.1.4 Alter

Die Ein- und Ausschlusskriterien zur Identifizierung relevanter Studien müssen geeignet sein, um Antworten auf die oben ausgeführte Fragestellung zu bekommen (s. Kapitel 4.2.1). Dementsprechend sind Studien, deren Patientenpopulation aus Kindern und Jugendlichen (<18 Jahre) besteht, ausgeschlossen (eine Zulassung von Teriflunomid (AUBAGIO®) besteht lediglich für erwachsene Patienten).

4.2.2.1.2 Intervention

a) Für die Bewertung des Nutzens und für die Bewertung des Zusatznutzens über einen direkten Vergleich werden nur Studien eingeschlossen, die Teriflunomid in der Dosierung 14 mg bei einmal täglicher Applikation als Intervention benennen. Studien mit anderen Wirkstoffen, anderen Dosierungen oder Kombinationen sind zur Beantwortung der Fragestellung nicht geeignet.

b) Für den indirekten Vergleich dient Placebo als Brückenkompator. Dementsprechend werden Studien eingeschlossen, die als Interventionen entweder Teriflunomid 14 mg (oral, 1 x täglich) oder Interferon beta-1a 44 µg (subkutan, 3 x wöchentlich) mit Placebo vergleichen.

4.2.2.1.3 Vergleichstherapie (Komparator/Comparison)

a) Zur Darstellung des Nutzens von Teriflunomid wird die Wirksamkeit und Verträglichkeit der Therapie im Vergleich zu Placebo oder zur zweckmäßigen Vergleichstherapie evaluiert.

Dementsprechend ist die Vergleichstherapie (Komparator) in direkt vergleichenden eingeschlossenen Studien zur Darstellung des Nutzens Placebo oder Interferon beta-1a, 44 µg, subkutan, 3 x wöchentlich (Rebif®).

b) Zur Darstellung des Zusatznutzens werden direkt vergleichende Studien eingeschlossen, in denen Interferon beta-1a, 44 µg (subkutan, 3 x wöchentlich) der festgelegte Komparator ist.

Im indirekten Vergleich zur Darstellung des Zusatznutzens ist Placebo der Brückenkompator (s. Kapitel 4.2.2.2) und damit auch der Komparator in den eingeschlossenen Studien.

Fazit: je nach Fragestellung kann der Komparator Placebo oder die zweckmäßige Vergleichstherapie Interferon beta-1a, 44 µg, subkutan, 3 x wöchentlich, (Rebif®) sein.

4.2.2.1.4 Endpunkte (Zielgrößen/Outcomes)

In der folgenden Tabelle sind die für die Darstellung des Nutzens und Zusatznutzens von Teriflunomid im Anwendungsgebiet A relevanten Endpunkte dargestellt. Begründungen für die Patientenrelevanz und Validität der einzelnen Endpunkte erfolgen in Kapitel 4.2.5.2.

Tabelle 4-5: Patientenrelevante Endpunkte für die Bewertung eines krankheitsmodifizierenden Wirkstoffes in der MS-Therapie

Endpunktkategorie und Ausprägung	Erhebung/Messinstrument
Mortalität	
Gesamtmortalität	Absolute Anzahl der Todesfälle
Morbidität	
Schübe	(mittlere) Jährliche Schubrate
	Zeit bis zum ersten Schub
	Anteil schubfreier Patienten oder Patienten mit ≥ 1 Schub
Behinderungsprogression	Zeit bis zur Behinderungsprogression
	Anteil Patienten ohne Behinderungsprogression bis zu einem definierten Zeitpunkt
	Zunahme der Behinderung (EDSS)
	Zunahme der Behinderung (MSFC)
Therapieversagen („Time to Failure“)	Zeit bis zum ersten Schub oder Abbruch der Therapie aus jedwedem Grund
Fatigue	Fatigue Impact Scale (FIS)
Unerwünschte Ereignisse	

unerwünschte Ereignisse (AEs)	Gesamtrate von Patienten mit Ereignis
schwerwiegende unerwünschte Ereignisse (SAEs)	Gesamtrate von Patienten mit Ereignis
unerwünschte Ereignisse, die zum Behandlungsabbruch führten	Gesamtrate von Patienten mit Ereignis
Unerwünschte Ereignisse von besonderem Interesse	Reaktionen an der Injektionsstelle
	Grippeähnliche Symptome
	Mit der Leberfunktion assoziierte Ereignisse (Gesamtrate, ALT-Erhöhung)
	Alopezien
	Gastrointestinale Beschwerden (Durchfall, Übelkeit und Erbrechen)
	Infektionen
	Neutropenie
	Hypertonie
	Kopfschmerzen
Periphere Neuropathie (inklusive Parästhesie)	
Gesundheitsbezogene Lebensqualität	
Gesundheitsbezogene Lebensqualität	Short Form 36 (SF-36)
	EuroQoL 5D (EQ-5D)
Patientenzufriedenheit	
Patientenzufriedenheit	Treatment Satisfaction Questionnaire for Medication (TSQM)

4.2.2.1.5 Studientypen

Zur Bewertung des Zusatznutzens von Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie werden nur RCTs eingeschlossen. „Evidenzeinstufungen im Hinblick auf die (Nutzen-)Bewertung medizinischer Interventionen basieren grundsätzlich auf einer Hierarchie der Evidenz/Studientypen“ [17]. Entsprechend würden Studien eines niedrigeren Evidenzgrades nur dann in die Bewertung eingeschlossen, wenn keine RCTs zur Verfügung stünden.

Die Auswahl der Studien beschränkt sich auf solche, die in deutscher oder englischer Sprache publiziert worden sind. Englisch ist die allgemein anerkannte Sprache der Wissenschaft, so dass davon auszugehen ist, dass relevante medizinische Erkenntnisse und Studien in dieser Sprache publiziert sind. Zusätzlich werden deutsche Quellen eingeschlossen, weil sie für eine aktuelle Bewertung für die GKV in Deutschland möglicherweise relevant sind.

4.2.2.1.6 Studiendauer

Studien mit einer geplanten Behandlungsdauer von mindestens 48 Wochen werden eingeschlossen. MS ist eine chronische Erkrankung, die klinisch wahrnehmbare

Verschlechterung geht oft langsam voran. Deswegen ist eine ausreichend lange Studiendauer notwendig. Das CHMP empfiehlt in seiner Leitlinie zur Durchführung von klinischen Studien in diesem Krankheitsbild eine Studiendauer von 2 Jahren [1], da gerade eine valide Erfassung der Behinderungsprogression in kürzeren Studien mit Unsicherheit verbunden ist. Für den Nachweis von Effekten auf die Schubaktivität wird auch ein Zeitraum von einem Jahr als prinzipiell ausreichend angesehen, auch wenn längere Beobachtungszeiträume zu bevorzugen sind [1, 2]. Deswegen wird als minimale Studiendauer 1 Jahr, bzw. geplante Behandlungsdauer von mindestens 48 Wochen, festgelegt. Die Aussagekraft von Studien mit einer noch kürzeren Studiendauer wird als nicht ausreichend hoch eingestuft.

4.2.2.2 Ein- und Ausschlusskriterien für randomisierte kontrollierte Studien mit dem zu bewertenden Arzneimittel und den indirekten Vergleich

Eine Suche in Studienregistern wird entsprechend den Vorgaben des G-BA durchgeführt. Eine bibliografische Literaturrecherche ist ebenfalls durchzuführen, s. 4.2.3.2.

4.2.2.2.1 Ein- und Ausschlusskriterien für Studien mit Teriflunomid als Intervention

Tabelle 4-6: Einschlusskriterien für Studien mit Teriflunomid als Intervention

Einschlusskriterien	
Patientenpopulation	<ul style="list-style-type: none"> • Erwachsene Patienten (≥ 18 Jahre) • Gesicherte MS-Diagnose (McDonald- oder Poser-Kriterien) • EDSS $\leq 5,5$ • Patienten mit schubförmig-remittierender Verlaufsform der Multiplen Sklerose (RRMS)
Intervention	<ul style="list-style-type: none"> • Teriflunomid 14 mg, einmal täglich, oral
Vergleichstherapie	<ul style="list-style-type: none"> • Interferon beta-1a, 44 μg, dreimal wöchentlich, subkutan (Rebif[®]) oder • Placebo
Endpunkte	<ul style="list-style-type: none"> • Schubbezogene Endpunkte, z. B. <ul style="list-style-type: none"> – Zeit bis zum ersten Schub – Jährliche Schubrate – Anteil schubfreier Patienten oder Patienten mit ≥ 1 Schub • Behinderungsprogression, z. B. <ul style="list-style-type: none"> – Zeit bis zur bestätigten Behinderungsprogression – Anzahl Patienten mit oder ohne Behinderungsprogression bis zu einem bestimmten Zeitpunkt – Änderung, der mittleren EDSS/MSFC-Werte • Gesundheitsbezogene Lebensqualität, z. B. gemessen mit

	<ul style="list-style-type: none"> – SF-36 – EQ-5D • Fatigue, z. B. gemessen mit <ul style="list-style-type: none"> – Fatigue Impact Scale (FIS) • Patientenzufriedenheit, z. B. gemessen mit <ul style="list-style-type: none"> – Treatment Satisfaction Questionnaire for Medication (TSQM) • Unerwünschte Ereignisse, z. B. <ul style="list-style-type: none"> – Gesamtrate, Gesamtrate schwerwiegender unerwünschter Ereignisse, Behandlungsabbrüche aufgrund von unerwünschten Ereignissen, Todesfälle, unerwünschte Ereignisse von besonderem Interesse
Studien- und Publikationstypen	<ul style="list-style-type: none"> • Randomisierte kontrollierte Studien • Volltextpublikation oder Studienbericht • Veröffentlichung in deutscher oder englischer Sprache
Studiendauer	<ul style="list-style-type: none"> • mindestens 1 Jahr (geplante Behandlungsdauer mind. 48 Wochen)

Tabelle 4-7: Ausschlusskriterien für Studien mit Teriflunomid als Intervention

Ausschlusskriterien		
A1	Patientenpopulation	<ul style="list-style-type: none"> • Kinder und Jugendliche (<18 Jahre) • Keine gesicherte MS-Diagnose • EDSS >5,5 • Patienten mit anderen Verlaufsformen der MS (z. B. CIS, SPMS, PPMS)
A2	Intervention	<ul style="list-style-type: none"> • Andere Wirkstoffe • Andere Dosierungen • Ausschließlich Kombination mit anderen Wirkstoffen
A3	Vergleichstherapie	<ul style="list-style-type: none"> • Andere Präparate als die unter den Einschlusskriterien genannten (z. B. Azathioprin, Glatirameracetat, Natalizumab, Fingolimod, Mitoxantron, Cyclophosphamid) • Ausschließlich andere Dosierungen als die in den Einschlusskriterien genannten
A4	Endpunkte	<ul style="list-style-type: none"> • Studien mit ausschließlich Surrogat-Parametern (z. B. nur MRT-bezogene Parameter)
A5	Studien- und	<ul style="list-style-type: none"> • Nur Abstract vorliegend

	Publikations- typen	<ul style="list-style-type: none"> • Review-Artikel, Surveys, Notes, Letter, Editorial, Errata, Konferenzabstracts ,Poster, Studienregistereinträge • Publikation nur in einer anderen Sprache als Deutsch oder Englisch
A6	Studiendauer	<ul style="list-style-type: none"> • <1 Jahr (geplante Behandlungsdauer mind. 48 Wochen)
A7	Art der Infor- mation	<ul style="list-style-type: none"> • Zusätzliche Publikation einer eingeschlossenen Studie ohne relevante Zusatzinformation

4.2.2.2.2 Ein- und Ausschlusskriterien für Studien mit Interferon beta-1a, 44 µg, subkutan, 3 x wöchentlich, (Rebif[®]) als Intervention (für indirekte Vergleiche)

Für Teriflunomid gibt es zum gegenwärtigen Zeitpunkt drei abgeschlossene randomisierte kontrollierte Studien, die die oben aufgeführten Einschlusskriterien erfüllen. Bei diesen Studien handelt es sich um die Studien TEMSO, TOWER und TENERE. TEMSO und TOWER sind placebokontrollierte Studien, wogegen es sich bei TENERE um einen direkten Vergleich mit der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg (Interferon beta-1a, subkutan, 3 x wöchentlich) handelt.

Mit der TENERE-Studie liegt ein direkter Vergleich mit der zweckmäßigen Vergleichstherapie vor. Das Studienziel, einen signifikanten Vorteil für Teriflunomid bezüglich des primären Endpunktes „Zeit bis zum Therapieversagen“ nachzuweisen, wurde nicht erreicht. Ein vorgeschalteter Nachweis der Nichtunterlegenheit in Bezug auf den primären Endpunkt war nicht geplant. Der sekundäre Endpunkt, die Reduktion der jährlichen Schubrate, war in den beiden Behandlungsgruppen vergleichbar (siehe 4.3.1.3.1.2). Deswegen sollen die Ergebnisse der TENERE-Studie durch einen indirekten Vergleich unterstützt werden, in den auch die größer angelegten placebokontrollierten Studien TEMSO und TOWER einbezogen werden können (s. Modul 4.3.2).

Ausgehend von der vorhandenen Evidenz mit zwei großen und qualitativ hochwertigen placebokontrollierten Studien ist eine Erhöhung der Aussagekraft der Ergebnisse, die sich auf die bereits ausgewerteten Informationen beziehen, durch einen indirekten Vergleich mit Placebo als Brückenkomparator möglich.

Im Folgenden werden nur die Ein- und Ausschlusskriterien für den Vergleich Rebif[®] 44 µg (Interferon beta-1a, subkutan, 3 x wöchentlich) versus Placebo aufgeführt, da die Ein- und Ausschlusskriterien für Studien, die Teriflunomid gegen Placebo vergleichen, schon in der entsprechenden Darstellung für den direkten Vergleich aufgeführt worden sind.

Abbildung 4-1: Netzwerk für den indirekten Vergleich

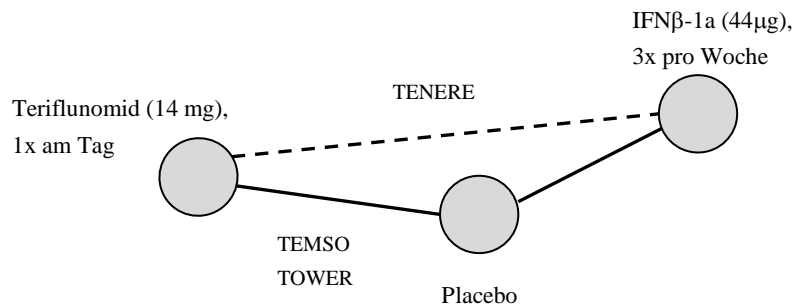


Tabelle 4-8: Einschlusskriterien für Interferon beta-1a, 44 µg-Studien für den indirekten Vergleich

Einschlusskriterien	
Patientenpopulation	<ul style="list-style-type: none"> • Erwachsene Patienten (≥ 18 Jahre) • Gesicherte MS-Diagnose (McDonald- oder Poser-Kriterien) • EDSS $\leq 5,5$ • Patienten mit schubförmig-remittierender Verlaufsform der Multiplen Sklerose (RRMS)
Intervention	<ul style="list-style-type: none"> • Interferon beta-1a, 44 µg, dreimal wöchentlich (Rebif®), subkutan
Vergleichstherapie	<ul style="list-style-type: none"> • Placebo
Endpunkte	<ul style="list-style-type: none"> • Schubbezogene Endpunkte, z. B. <ul style="list-style-type: none"> – Zeit bis zum ersten Schub – Jährliche Schubrate – Anteil schubfreier Patienten oder Patienten mit ≥ 1 Schub • Behinderungsprogression, z. B. <ul style="list-style-type: none"> – Zeit bis zur bestätigten Behinderungsprogression – Anzahl Patienten mit oder ohne Behinderungsprogression bis zu einem bestimmten Zeitpunkt – Änderung, der mittleren EDSS/MSFC-Werte • Gesundheitsbezogene Lebensqualität, z. B. gemessen mit <ul style="list-style-type: none"> – SF-36 – EQ-5D • Fatigue, z. B. gemessen mit <ul style="list-style-type: none"> – Fatigue Impact Scale (FIS)

	<ul style="list-style-type: none"> • Patientenzufriedenheit, z. B. gemessen mit <ul style="list-style-type: none"> – Treatment Satisfaction Questionnaire for Medication (TSQM) • Unerwünschte Ereignisse, z. B. <ul style="list-style-type: none"> – Gesamtrate, Gesamtrate schwerwiegender unerwünschter Ereignisse, Behandlungsabbrüche aufgrund von unerwünschten Ereignissen, Todesfälle, unerwünschte Ereignisse von besonderem Interesse
Studien- und Publikationstypen	<ul style="list-style-type: none"> • Randomisierte kontrollierte Studien • Volltextpublikation, Studienbericht • Veröffentlichung in deutscher oder englischer Sprache
Studiendauer	<ul style="list-style-type: none"> • mindestens 1 Jahr (geplante Behandlungsdauer mind. 48 Wochen)

Tabelle 4-9: Ausschlusskriterien für Interferon beta-1a, 44 µg-Studien für den indirekten Vergleich

Ausschlusskriterien		
A1	Patientenpopulation	<ul style="list-style-type: none"> • Kinder und Jugendliche (<18 Jahre) • Keine gesicherte MS-Diagnose • EDSS >5,5 • Patienten mit anderen Verlaufsformen der MS (z. B. CIS, SPMS, PPMS)
A2	Intervention	<ul style="list-style-type: none"> • Andere Wirkstoffe oder Präparate als die unter den Einschlusskriterien genannten • Andere Dosierungen • Ausschließlich Kombination mit anderen Wirkstoffen
A3	Vergleichstherapie	<ul style="list-style-type: none"> • Jegliche von Placebo abweichende Vergleichstherapie
A4	Endpunkte	<ul style="list-style-type: none"> • Studien mit ausschließlich Surrogat-Parametern (z. B. nur MRT-bezogene Parameter)
A5	Studien- und Publikationstypen	<ul style="list-style-type: none"> • Keine randomisierte kontrollierte Studie • Nur Abstract vorliegend • Review-Artikel, Surveys, Notes, Letter, Editorial, Errata, Konferenzabstracts, Poster, Studienregistereinträge • Publikation nur in einer anderen Sprache als Deutsch oder Englisch

A6	Studiendauer	• <1 Jahr (geplante Behandlungsdauer mind. 48 Wochen)
A7	Art der Information	• Zusätzliche Publikation einer eingeschlossenen Studien ohne relevante Zusatzinformation

4.2.3 Informationsbeschaffung

In den nachfolgenden Abschnitten ist zu beschreiben, nach welcher Methodik Studien identifiziert wurden, die für die Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens in dem in diesem Dokument bewerteten Anwendungsgebiet herangezogen werden. Dies bezieht sich sowohl auf publizierte als auch auf unpublizierte Studien. Die Methodik muss dazu geeignet sein, die relevanten Studien (gemäß den in Abschnitt 4.2.2 genannten Kriterien) systematisch zu identifizieren (systematische Literaturrecherche).

4.2.3.1 Studien des pharmazeutischen Unternehmers

Für die Identifikation der Studien des pharmazeutischen Unternehmers ist keine gesonderte Beschreibung der Methodik der Informationsbeschaffung erforderlich. Die vollständige Auflistung aller Studien, die an die Zulassungsbehörde übermittelt wurden (Zulassungsstudien), sowie aller Studien, für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, erfolgt in den Abschnitten 4.3.1 und 4.3.2, jeweils im Unterabschnitt „Studien des pharmazeutischen Unternehmers“. Die Darstellung soll auf Studien mit Patienten in dem Anwendungsgebiet, für das das vorliegende Dokument erstellt wird, beschränkt werden.

4.2.3.2 Bibliografische Literaturrecherche

Die Durchführung einer bibliografischen Literaturrecherche ist in folgenden Fällen erforderlich, um sicherzustellen, dass ein vollständiger Studienpool in die Bewertung einfließt:

- Für das zu bewertende Arzneimittel wurde eine Zulassung in dem Anwendungsgebiet, für das das vorliegende Dokument erstellt wurde, mindestens 12 Monate vor dem Zeitpunkt der Dossiereinreichung erteilt. Dabei sind auch Zulassungen außerhalb Deutschlands bzw. Europas relevant.
- Es werden indirekte Vergleiche zur Bewertung des medizinischen Zusatznutzens herangezogen.

Sofern eine bibliografische Literaturrecherche erforderlich ist, soll diese mindestens in den Datenbanken MEDLINE und EMBASE sowie in den Cochrane-Datenbanken durchgeführt werden. Optional kann zusätzlich eine Suche in weiteren themenspezifischen Datenbanken (z. B. CINAHL, PsycINFO etc.) durchgeführt werden.

Die Suche soll in jeder Datenbank einzeln und mit einer für die jeweilige Datenbank adaptierten Suchstrategie durchgeführt werden. Die Suchstrategien sollen jeweils in Blöcken, insbesondere getrennt nach Indikation, Intervention und ggf. Studientypen, aufgebaut werden. Wird eine Einschränkung der Strategien auf bestimmte Studientypen vorgenommen (z. B.

randomisierte kontrollierte Studien), sollen aktuelle validierte Filter hierfür verwendet werden. Eine gemeinsame Suche nach Studien zu mehreren Fragestellungen (z. B. direkt vergleichende Studien sowie Studien für einen indirekten Vergleich) ist möglich. Alle Suchstrategien sind in Anhang 4-A zu dokumentieren.

Beschreiben Sie nachfolgend, in welchen Datenbanken eine bibliografische Literaturrecherche durchgeführt wurde. Geben Sie dabei an, ob Sie eine gemeinsame Suche für mehrere Fragestellungen durchgeführt haben (z. B. gemeinsame Suche nach Studien mit dem zu bewertenden Arzneimittel und Studien zur zweckmäßigen Vergleichstherapie für indirekte Vergleiche). Begründen Sie Abweichungen von den oben beschriebenen Vorgaben. Geben Sie auch an, ob bei der Recherche generelle Einschränkungen vorgenommen wurden (z. B. Sprach- oder Jahreseinschränkungen), und begründen Sie diese.

Die Zulassung von Teriflunomid zur Behandlung der schubförmigen MS wurde in den USA am 12.09.2013 also über 12 Monate vor dem Zeitpunkt der Dossiereinreichung erteilt, so dass für die Suche nach randomisierten kontrollierten Studien eine bibliographische Literaturrecherche vorgeschrieben ist.

Zur Erhöhung der Aussagesicherheit wurde zusätzlich zu der direkt vergleichenden Studie TENERE indirekte Vergleiche der Interventionen Teriflunomid und Interferon beta-1a subkutan 3 x wöchentlich (Rebif® 44 µg) mit dem Brückenkomparator Placebo durchgeführt. Auch aufgrund dieses indirekten Vergleiches wurde eine systematische bibliographische Literaturrecherche durchgeführt.

Es wurde in MEDLINE und EMBASE sowie in den Cochrane-Datenbanken gesucht. Für die Suche nach Studien des zu bewertenden Arzneimittels im Vergleich mit dem Brückenkomparator Placebo und nach entsprechenden Studien der zweckmäßigen Vergleichstherapie wurden zwei separate Recherchen durchgeführt. Es wurde keine Jahreseinschränkung vorgenommen, die Suche wurde auf Studien in englischer und deutscher Sprache beschränkt, dies wurde in Kapitel 4.2.2 begründet.

4.2.3.3 Suche in Studienregistern

Eine Suche in öffentlich zugänglichen Studienregistern ist grundsätzlich durchzuführen, um sicherzustellen, dass laufende Studien sowie abgeschlossene Studien von Dritten vollständig identifiziert werden.

Die Suche soll mindestens in den Studienregistern clinicaltrials.gov, clinicalstudyresults.org sowie über das International Clinical Trials Registry Platform Search Portal (ICTRP Search Portal, Suchportal der WHO) durchgeführt werden. Optional kann zusätzlich eine Suche in weiteren themenspezifischen Studienregistern (z. B. krankheitsspezifische Studienregister oder Studienregister einzelner pharmazeutischer Unternehmen) durchgeführt werden.

Die Suche soll in jedem Studienregister einzeln und mit einer für das jeweilige Studienregister adaptierten Suchstrategie durchgeführt werden. Die Suche soll abgeschlossene, abgebrochene und laufende Studien erfassen. Eine gemeinsame Suche nach

Studien zu mehreren Fragestellungen (z. B. direkt vergleichende Studien sowie Studien für einen indirekten Vergleich) ist möglich. Alle Suchstrategien sind in Anhang 4-B zu dokumentieren.

Beschreiben Sie nachfolgend, in welchen Studienregistern die Suche durchgeführt wurde. Begründen Sie dabei Abweichungen von den oben beschriebenen Vorgaben. Geben Sie auch an, ob bei der Recherche generelle Einschränkungen vorgenommen wurden (z. B. Jahreseinschränkungen), und begründen Sie diese.

Die Suche wurde in den Studienregistern clinicaltrials.gov und über das International Clinical Trials Registry Platform Search Portal (ICTRP Search Portal, Suchportal der WHO) durchgeführt. Es besteht kein Zugriff auf clinicalstudyresults.org, so dass eine entsprechende Suche nicht durchgeführt werden konnte.

Zusätzlich wurde im EU Clinical Trials Register und im Internetportal PharmNet.Bund.de nach klinischen Prüfungen gesucht.

4.2.3.4 Selektion relevanter Studien

Beschreiben Sie das Vorgehen bei der Selektion relevanter Studien aus dem Ergebnis der in den Abschnitten 4.2.3.2 und 4.2.3.3 beschriebenen Rechenschritte. Begründen Sie das Vorgehen, falls die Selektion nicht von zwei Personen unabhängig voneinander durchgeführt wurde.

Die Literaturrecherche wurde von zwei Personen unabhängig voneinander durchgeführt. Die gefundenen Treffer wurden nach den in Kapitel 4.2.2 dargelegten Kriterien bewertet. Dabei wurde in einem mehrstufigen Verfahren vorgegangen. In einem ersten Schritt wurden die Titel gesichtet und ein Großteil der Treffer bereits ausgeschlossen. In einem weiteren Schritt wurden Studien des verbleibenden Studienpools anhand der Abstracts bewertet und Kandidaten für die Analyse der Volltexte identifiziert. Nach Lesen und Bewerten der Volltexte konnte die endgültige Liste der eingeschlossenen Studien aufgestellt werden. Bei Unklarheiten oder unterschiedlichen Treffern in den parallel durchgeführten Recherchen wurde eine dritte Person zu Rate gezogen.

4.2.4 Bewertung der Aussagekraft der Nachweise

Zur Bewertung der Aussagekraft der im Dossier vorgelegten Nachweise sollen Verzerrungsaspekte der Ergebnisse für jede eingeschlossene Studie beschrieben werden, und zwar separat für jeden patientenrelevanten Endpunkt. Dazu sollen insbesondere folgende endpunktübergreifende (A) und endpunktspezifische (B) Aspekte systematisch extrahiert werden (zur weiteren Erläuterung der einzelnen Aspekte siehe Bewertungsbogen in Anhang 4-G):

A: Verzerrungsaspekte der Ergebnisse auf Studienebene

- Erzeugung der Randomisierungssequenz (*bei randomisierten Studien*)
- Verdeckung der Gruppenzuteilung (*bei randomisierten Studien*)

- zeitliche Parallelität der Gruppen (*bei nicht randomisierten vergleichenden Studien*)
- Vergleichbarkeit der Gruppen bzw. Berücksichtigung prognostisch relevanter Faktoren (*bei nicht randomisierten vergleichenden Studien*)
- Verblindung des Patienten sowie des Behandlers
- ergebnisgesteuerte Berichterstattung
- sonstige Aspekte

B: Verzerrungsaspekte der Ergebnisse auf Endpunktebene

- Verblindung der Endpunkterheber
- Umsetzung des ITT-Prinzips
- ergebnisgesteuerte Berichterstattung
- sonstige Aspekte

Für randomisierte Studien soll darüber hinaus das Verzerrungspotenzial bewertet und als „niedrig“ oder „hoch“ eingestuft werden. Ein niedriges Verzerrungspotenzial liegt dann vor, wenn mit großer Wahrscheinlichkeit ausgeschlossen werden kann, dass die Ergebnisse relevant verzerrt sind. Unter einer relevanten Verzerrung ist zu verstehen, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

Eine zusammenfassende Bewertung der Verzerrungsaspekte soll nicht für nicht randomisierte Studien erfolgen.

Für die Bewertung eines Endpunkts soll für randomisierte Studien zunächst das Verzerrungspotenzial endpunktübergreifend anhand der unter A aufgeführten Aspekte als „niedrig“ oder „hoch“ eingestuft werden. Falls diese Einstufung als „hoch“ erfolgt, soll das Verzerrungspotenzial für den Endpunkt in der Regel auch als „hoch“ bewertet werden, Abweichungen hiervon sind zu begründen. Ansonsten sollen die unter B genannten endpunktspezifischen Aspekte Berücksichtigung finden.

Eine Einstufung des Verzerrungspotenzials des Ergebnisses für einen Endpunkt als „hoch“ soll nicht zum Ausschluss der Daten führen. Die Klassifizierung soll vielmehr der Diskussion heterogener Studienergebnisse und der Einschätzung der Aussagekraft der Nachweise dienen. Für nicht randomisierte Studien können für solche Diskussionen einzelne Verzerrungsaspekte herangezogen werden.

Beschreiben Sie die für die Bewertung des Verzerrungspotenzials eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.

Von der oben angegebenen Methodik zur Bewertung des Verzerrungspotentials wird nicht abgewichen. Zunächst wird auf Studienebene das endpunktübergreifende Verzerrungspotential als hoch oder niedrig eingestuft. Im Anschluss erfolgt eine Bewertung für jeden einzelnen Endpunkt. Das Ergebnis der Analysen ist in Anhang 4-G dokumentiert.

4.2.5 Informationssynthese und -analyse

4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien

Das Design und die Methodik der eingeschlossenen Studien soll in den Abschnitten 4.3.1 und 4.3.2, jeweils in den Unterabschnitten „Charakteristika der in die Bewertung eingeschlossenen Studien“ und den dazugehörigen Anhängen, dargestellt werden. Die Darstellung der Studien soll für randomisierte kontrollierte Studien mindestens die Anforderungen des CONSORT-Statements erfüllen (Items 2b bis 14, Informationen aus dem CONSORT-Flow-Chart)¹. Die Darstellung nicht randomisierter Interventionsstudien und epidemiologischer Beobachtungsstudien soll mindestens den Anforderungen des TREND-² bzw. STROBE-Statements³ folgen. Design und Methodik weiterer Untersuchungen sollen gemäß den verfügbaren Standards dargestellt werden.

Beschreiben Sie, nach welchen Standards und mit welchen Informationen (Items) Sie das Design und die Methodik der eingeschlossenen Studien in Modul 4 dargestellt haben. Begründen Sie Abweichungen von den oben beschriebenen Vorgaben.

Das Design und die Methodik der eingeschlossenen Studien wird in den Abschnitten 4.3.1 und 4.3.2 und den zugehörigen Anhängen dokumentiert. Die Bewertung folgte den CONSORT-Vorgaben (Items 2b bis 14) sowie dem CONSORT-Flow-Chart. Bei allen eingeschlossenen Studien handelt es sich um randomisierte kontrollierte Studien, somit entspricht die Beurteilung und Darstellung des Designs und der Methodik der Studien den Vorgaben. Nicht randomisierte Interventionsstudien und epidemiologische Beobachtungsstudien werden nicht zur Bewertung des Nutzens und des Zusatznutzens von Teriflunomid im Anwendungsgebiet A herangezogen.

4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien

Die Ergebnisse der einzelnen Studien sollen in den Abschnitten 4.3.1 und 4.3.2 in den entsprechenden Unterabschnitten zunächst für jede eingeschlossene Studie separat dargestellt werden. Die Darstellung soll die Charakteristika der Studienpopulationen sowie die Ergebnisse zu allen in den eingeschlossenen Studien berichteten patientenrelevanten Endpunkten (Verbesserung des Gesundheitszustands, Verkürzung der Krankheitsdauer, Verlängerung des Überlebens, Verringerung von Nebenwirkungen, Verbesserung der Lebensqualität) umfassen. Anforderungen an die Darstellung werden in den Unterabschnitten beschrieben.

¹ Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c332.

² Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Publ Health 2004; 94(3): 361-366.

³ Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007; 147(8): 573-577.

Benennen Sie die Patientencharakteristika und Endpunkte, zu denen Ergebnisse dargestellt werden. Begründen Sie, wenn Sie von den oben benannten Vorgaben abgewichen sind. Beschreiben Sie für jeden Endpunkt, warum Sie ihn als patientenrelevant einstufen, und machen Sie Angaben zur Validität des Endpunkts (z. B. zur Validierung der eingesetzten Fragebögen). Beachten Sie bei der Berücksichtigung von Surrogatendpunkten Abschnitt 4.5.4.

4.2.5.2.1 Patientencharakteristika

Entsprechend der Zulassung von Teriflunomid zur Behandlung erwachsener Patienten mit schubförmig-remittierender Multipler Sklerose (RRMS) werden Ergebnisse zu Patienten dargestellt, die folgende Charakteristika aufweisen:

- Erwachsene Patienten (≥ 18 Jahre)
- gesicherte MS-Diagnose (McDonald- oder Poser-Kriterien)
- EDSS $\leq 5,5$
- Patienten mit RRMS

Für die Charakterisierung der Studienpopulation in Tabelle 4-18 werden dementsprechend das Alter, der Baseline-EDSS-Wert und die Verlaufsform der Patienten in den einzelnen Studien angegeben. Für eine weitere Charakterisierung werden außerdem folgende Charakteristika dargestellt:

- Geschlecht
- Anzahl der Schübe im vorangegangenen Jahr
- Zeit seit der ersten MS-Diagnose.

4.2.5.2.2 Endpunkte

Nach der Verfahrensordnung des Gemeinsamen Bundesausschusses ist der Nutzen eines Arzneimittels definiert als „der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustands, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“ [11]. Dementsprechend werden die relevanten Studienendpunkte in die folgenden Kategorien eingeteilt [17]:

- Mortalität
- Morbidität (Beschwerden und Komplikationen)
- Gesundheitsbezogene Lebensqualität
- Unerwünschte Ereignisse

Zusätzlich:

- Patientenzufriedenheit

Nach Einschätzung des IQWiG und des pU ist die Zufriedenheit der Patienten mit der Behandlung für die Beurteilung relevant, kann jedoch nicht allein als Basis für die Bewertung des Nutzens oder Zusatznutzens herangezogen werden [17]. Die Zufriedenheit der Patienten mit der Therapie ist gerade im Bereich der MS ein wesentliches Therapieziel, weil eine größere Zufriedenheit mit der Therapie aufgrund einer besseren Wirksamkeit, eines günstigeren Nebenwirkungsprofils oder einer einfacheren Handhabung einen wesentlichen Einfluss auf Compliance und Therapieadhärenz hat [18-20]. Diese Aspekte wiegen im Bereich der krankheitsmodifizierenden MS-Therapie besonders schwer, weil eine dauerhafte Medikamenteneinnahme Voraussetzung für eine effiziente Hinauszögerung gravierender gesundheitlicher Einschränkungen ist. Deshalb sollte die Patientenzufriedenheit im betrachteten Anwendungsgebiet A (Behandlung erwachsener Patienten mit schubförmig-remittierender Multipler Sklerose), wenn möglich, valide erfasst werden.

4.2.5.2.2.1 Mortalität

Mortalität ist ein patientenrelevanter Endpunkt, der in jeder Studie erhoben wird. Mortalität ist jedoch kein für die MS-Therapie typischer Endpunkt, da es eine chronische Erkrankung ist, die erwartungsgemäß im Verlauf einer Studie von 1-3 Jahren Dauer nicht zum Tod führt. Darüber hinaus handelt es sich weder bei Teriflunomid noch bei der zweckmäßigen Vergleichstherapie um Wirkstoffe, bei denen mit einer erhöhten Anzahl an Todesfällen zu rechnen ist.

4.2.5.2.2.2 Morbidität

Die schubförmig-remittierende Multiple Sklerose ist in ihrem Krankheitsverlauf durch das Auftreten von Schüben und eine zunehmende Behinderung charakterisiert. Dementsprechend sind die Schübe und die Behinderungsprogression die beiden wichtigsten klinischen Endpunkte zur Bewertung krankheitsmodifizierender Wirkstoffe in der MS-Therapie. Entsprechend sollten auch nach Empfehlung der CHMP-Leitlinie zur Durchführung von klinischen Studien in diesem Anwendungsgebiet [1, 2] entweder die Schubrate oder die Behinderungsprogression das primäre Zielkriterium sein, wobei der jeweils andere ein wichtiger sekundärer Endpunkt sein sollte.

Schübe

Die Schübe sind einer der beiden klinischen Endpunkte, die bei der Bewertung eines krankheitsmodifizierenden MS-Wirkstoffes berücksichtigt werden sollten. Die Schübe werden üblicherweise als „Jährliche Schubrate“ operationalisiert, aber auch eine Darstellung als „Zeit bis zum ersten Schub“ oder „Anteil schubfreier Patienten“ oder „Patienten mit ≥ 1 Schub“ ist als verlässlich einzustufen [10]. Für eine valide Erfassung der Schübe ist es essentiell, dass z. B. deren Erscheinungsform und Mindestdauer vorab genau definiert sind [2]. Dies ist bei der Operationalisierung der Schübe in den Studien entsprechend zu beachten.

Behinderungsprogression

Die Behinderungsprogression kann beispielsweise erhoben werden als „Zeit bis zur Behinderungsprogression“, „Anteil Patienten ohne Behinderungsprogression“ oder „Zunahme der Behinderungsprogression“. Vergleichbar mit den schubbezogenen Endpunkten ist es unstrittig, dass es sich bei der Behinderungsprogression um einen patientenrelevanten Endpunkt handelt. Besonderes Augenmerk muss jedoch auf die Anwendbarkeit der verwendeten Messinstrumente gelegt werden. Zur Messung der Behinderungsprogression sind in der MS-Forschung verschiedene Skalen etabliert. Der EDSS (Expanded Disability Status Scale) ist eine in hohem Maße etablierte Skala, die nach Empfehlung der EMA allein aus Gründen der Vergleichbarkeit in jedem Fall eingesetzt werden sollte. Aufgrund der fehlenden Linearität (keine metrische Skala) ist der EDSS jedoch nicht geeignet, um eine mittlere Änderung der Behinderung valide abzubilden. Als Kriterien für die Bestimmung der Zeit bis zur Behinderungsprogression oder des Anteils an Patienten ohne Behinderungsprogression ist er jedoch akzeptiert [2, 10].

Für die mittlere Änderung der Behinderung ist der MSFC (Multiple Sclerosis Functional Composite) ein geeignetes Instrument, das explizit für die Bewertung der Behinderung validiert worden ist [21].

Therapieversagen

Das Therapieversagen ist operationalisiert als „Time-to-event-Variable“ (Zeit bis zum Therapieversagen; „Time to Failure“). „Time to Failure“ ist ein kombinierter Endpunkt, der in der TENERE-Studie definiert wurde als die Zeit bis zum Auftreten eines Schubes oder bis zum Abbruch der Studienmedikation aus jedwedem anderen Grund [22]. Dieser Endpunkt bildet neben der klinischen Wirksamkeit die Bedeutung einer dauerhaften Therapie bei der Behandlung einer chronischen Erkrankung ab und ist damit von Bedeutung für den klinischen Alltag.

Kombinierte Endpunkte haben einige grundsätzliche Vorteile: durch die Erhöhung der Anzahl an Ereignissen, wird die statistische Aussagekraft erhöht und es wird ermöglicht, zwei Aspekte gleichwertig in einem Endpunkt zu erfassen. Gleichzeitig birgt dies aber auch die Gefahr, irreführende Aussagen zu treffen, wenn die Endpunkte nicht die gleiche Wichtung haben und nicht gleichgerichtet sind [23].

Im vorliegenden Fall des Endpunktes „Time to Failure“ ist der positive Aspekt sehr deutlich, nämlich, dass der wirkliche Vorteil einer Therapie – sowohl eine positive Beeinflussung der Morbidität (ausgedrückt durch die Zeit bis zum ersten Schub) als auch der Verträglichkeit und Lebensqualität (ausgedrückt durch die Zeit bis zum Therapieabbruch aus jedwedem Grund) - mit **einem** Endpunkt (früher eintretendes Ereignis der beiden oben definierten Einzelendpunkte) erfasst werden können. Diese beiden Aspekte eignen sich insofern zu einer Zusammenfassung zu einem kombinierten Endpunkt, als dass die Schübe und die unerwünschten Ereignisse (Hauptgrund für einen Therapieabbruch) gleichgerichtete Faktoren sind, die bei Eintritt zu einem Therapieabbruch bzw. einer Therapieumstellung führen

können. Des Weiteren sind keine Endpunkte ausgeschlossen, die eingeschlossen werden müssten. Die beiden einzelnen Endpunkte können als „von vergleichbarer Schwere“ eingestuft werden und der kombinierte Endpunkt wurde vorab definiert. Damit sind wichtige Kriterien für einen kombinierten Endpunkt erfüllt [17].

Trotzdem ist es möglich, dass die Ergebnisse der einzelnen Abbruchkriterien in unterschiedlichem Maß zum Gesamtergebnis beitragen, was zu einem verzerrten Bild führen würde. Deswegen werden die Ergebnisse der Einzelendpunkte auch separat berichtet (s. Modul 4.3.1.3.3).

Fatigue

Die starke Erschöpfung (Fatigue) ist ein sehr häufiges Symptom bei MS und gehört zu den Faktoren, die die Lebensqualität der Betroffenen sehr stark einschränken, weil sowohl die berufliche als auch die soziale Leistungsfähigkeit erheblich vermindert ist [24]. Die Fatigue ist ein für MS-Patienten sehr belastendes Symptom ihrer Erkrankung und damit auch ein patientenrelevanter Endpunkt bei der Bewertung eines krankheitsmodifizierenden Wirkstoffes. Voraussetzung für die Erfassung der Fatigue ist die Anwendung eines validierten Messinstrumentes, was mit dem FIS (Fatigue Impact Scale) gegeben ist [7, 8].

4.2.5.2.2.3 Gesundheitsbezogene Lebensqualität

1948 definierte die WHO Gesundheit als Zustand vollständigen physischen, psychischen und sozialen Wohlbefindens und nicht lediglich als die Abwesenheit von Krankheit oder Gebrechen [25]. Auf dieser seitdem nicht geänderten Definition basiert auch das Konzept der „gesundheitsbezogenen Lebensqualität“ (health-related quality of life, HRQoL). Gesundheitsbezogene Lebensqualität berücksichtigt die Relevanz der Quantität, aber zusätzlich auch der Qualität der verlebten Jahre [26]. Gesundheitsbezogene Lebensqualität ist patientenrelevant und damit auch für die Darstellung eines Nutzens und auch Zusatznutzens einer Intervention bedeutsam, der „insbesondere hinsichtlich der Verbesserung des Gesundheitszustands, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“ [11] gezeigt werden soll. Die gesundheitsbezogene Lebensqualität muss mit Hilfe validierter Messinstrumente erhoben werden, dazu zählen der SF-36 [27-29] und der EQ-5D [30].

4.2.5.2.2.4 Patientenzufriedenheit

Wie oben ausgeführt ist nach Einschätzung des IQWiG und des pU auch die Zufriedenheit der Patienten mit der Behandlung für die Beurteilung relevant, kann jedoch nicht **allein** als Basis für die Bewertung des Nutzens oder Zusatznutzens herangezogen werden [17]. Bei einer Behandlung, die nachgewiesenermaßen einen Nutzen für den Patienten in Bezug auf die Wirksamkeit hat, ist die Zufriedenheit der Patienten, die auch maßgeblich für folgende Therapieadhärenz und Compliance der Patienten ist, gerade bei der Bewertung einer Dauermedikation von großer Bedeutung.

Voraussetzung für die Auswertung dieses Endpunktes ist auch hier die Anwendung validierter Messinstrumente. Ein validiertes Messinstrument ist z. B. der Treatment Satisfaction Questionnaire for Medication, TSQM [4, 5].

4.2.5.2.2.5 Nebenwirkungen (Unerwünschte Ereignisse)

Unerwünschte Ereignisse werden zur Bewertung des Sicherheitsprofils eines Wirkstoffs als Gesamtrate unerwünschter Ereignisse, Gesamtrate schwerwiegender unerwünschter Ereignisse, unerwünschter Ereignisse, die zum Behandlungsabbruch führten sowie als unerwünschte Ereignisse von besonderem Interesse aufgeführt und bewertet.

Für jeden Wirkstoff gibt es zusätzlich zu den oben aufgeführten allgemeinen Kriterien zur Abschätzung des Schadenspotentials spezielle Risiken, die aufgrund der bestehenden Erfahrungen (z. B. aus den klinischen Studien) oder in Bezugnahme auf den Wirkmechanismus ein besonderes Risiko darstellen könnten. Für eine objektive Bewertung des Nutzen-Risiko-Verhältnisses eines Wirkstoffes (auch im Bezug auf einen Zusatznutzen), ist eine Auswertung dieser unerwünschten Ereignisse von besonderem Interesse unabdingbar.

Im Rahmen der Nutzenbewertung von Teriflunomid werden dementsprechend ausgewählte unerwünschte Ereignisse von besonderem Interesse ausgewertet. Die Auswahl der unerwünschten Ereignisse folgt dabei der folgenden Rationale:

- Ereignisse, die in den placebokontrollierten Studien [31, 32] signifikant häufiger aufgetreten sind als unter Placebo, können nach derzeitiger Studienlage als spezielles Sicherheitsrisiko für Teriflunomid eingestuft werden und müssen dementsprechend berichtet werden. Dabei handelt es sich um:
 - Verminderte Haardichte (HLT „Alopecias“)
 - Mit der Leberfunktion assoziierte Ereignisse (Gesamtrate, Erhöhung der Alanin-Amino-Transferase (ALT)-Werte)
 - Gastrointestinale Beschwerden (Durchfall, Übelkeit und Erbrechen)
 - Neutropenie
 - Periphere Neuropathie
- Ereignisse, die in der Fachinformation [33] zusätzlich zu den oben genannten Ereignissen als Sicherheitshinweis oder als „sehr häufig“ angegeben sind. Dies sind:
 - Hypertension
 - Infektionen (als SOC „Infections and Infestations“ ausgewertet)
 - Parästhesie (dieses unerwünschte Ereignis wird unter „Periphere Neuropathie“ in Abschnitt 4.3.1 berücksichtigt)

- Ereignisse, die in der Fachinformation von Rebif 44 µg als „sehr häufig“ aufgeführt sind [34] und mit den für Teriflunomid berücksichtigten Ereignissen nicht abgedeckt sind:
 - Reaktionen an der Injektionsstelle
 - Grippeähnliche Symptome
 - Kopfschmerzen

Tabelle 4-10: Übersicht über die patientenrelevante Endpunkte, die für die Bewertung von Teriflunomid herangezogen werden

Endpunktkategorie und Ausprägung	Erhebung	Kommentar
Mortalität		
Gesamtmortalität	Absolute Anzahl der Todesfälle	
Morbidität		
Schübe	Jährliche Schubrate	Die Häufigkeit und Schwere der Schübe ist einer der beiden wichtigen klinischen Endpunkte zur Bewertung krankheitsmodifizierender Wirkstoffe in der MS-Therapie [1, 2].
	Zeit bis zum ersten Schub	
	Anteil schubfreier Patienten oder Patienten mit ≥ 1 Schub	
Behinderungsprogression	Zeit bis zur Behinderungsprogression	Die Behinderungsprogression ist der zweite wichtige klinische Endpunkt zur Bewertung krankheitsmodifizierender Wirkstoffe in der MS-Therapie [1, 2].
	Anteil Patienten mit oder ohne Behinderungsprogression bis zu einem definierten Zeitpunkt	
	Zunahme der Behinderungsprogression (EDSS)	
	Zunahme der Behinderungsprogression (MSFC)	
Therapieversagen	Time to Failure	„Time to Failure“ kombiniert die Endpunkte „Zeit bis zum nächsten Schub“ und „Zeit bis zum Therapieabbruch aus jedwedem Grund“. Die Endpunkte, aus denen sich „Time to Failure“ zusammensetzt, sind patientenrelevante Endpunkte, die als von ähnlicher Schwere

		eingestuft und so zu einem kombinierten Endpunkt zusammengefasst werden können.
Fatigue	Fatigue Impact Scale (FIS)	Der Fatigue Impact Scale ist ein validiertes Messinstrument zur Bewertung der Fatigue [7, 8].
Unerwünschte Ereignisse		
unerwünschte Ereignisse (AEs)	Gesamtrate von Patienten mit Ereignis	
schwerwiegende unerwünschte Ereignisse (SAEs)	Gesamtrate von Patienten mit Ereignis	
Behandlungsabbruch wegen unerwünschter Ereignisse	Gesamtrate von Patienten mit Ereignis	
Unerwünschte Ereignisse von besonderem Interesse	Reaktionen an der Injektionsstelle	Die unerwünschten Ereignisse, die bezogen auf Teriflunomid, Interferon beta-1a oder das Immunsystem von besonderer Bedeutung sind, werden mit Blick auf die Häufigkeit der entsprechenden MedDRA [®] Terme ausgewertet. Für eine valide, nicht verfälschende Aussage wird für jedes unerwünschte Ereignis auch die Häufigkeit der Ereignisse in der zugehörigen primären System Organ Class (SOC) angegeben.
	Grippeähnliche Symptome	
	Mit der Leberfunktion assoziierte Ereignisse (Gesamtrate, ALT-Erhöhung)	
	Verminderte Haardichte (HLT „Alopecias“)	
	Gastrointestinale Beschwerden (Durchfall, Übelkeit und Erbrechen)	
	Infektionen und Infestationen	
	Neutropenie	
	Hypertonie	
	Kopfschmerzen	
Periphere Neuropathie (inklusive Parästhesie)		
Gesundheitsbezogene Lebensqualität		
Gesundheitsbezogene Lebensqualität	Short Form 36 (SF-36)	Der SF-36 [29] ist ein gut etablierter, weit eingesetzter Fragebogen zur Erfassung der Lebensqualität von Patienten. Der Fragebogen ist nicht nur in weiten Bereichen als Goldstandard zur Erfassung des Gesundheitsstatus anerkannt und gut etabliert [28], sondern auch wissenschaftlich

		validiert [27].
Gesundheitsbezogene Lebensqualität	EuroQoL 5D (EQ-5D)	Ebenfalls zur Erhebung der Lebensqualität geeignet ist der EQ-5D, der für die Anwendung bei Multiple Sklerose validiert ist [30].
Patientenzufriedenheit		
Patientenzufriedenheit	Treatment Satisfaction Questionnaire for Medication (TSQM)	Der TSQM ist ein validiertes Messinstrument zur Erfassung der Patientenzufriedenheit [4, 5].

4.2.5.3 Metaanalysen

Sofern die vorliegenden Studien dazu geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst werden. Dabei sollen verschiedene Studientypen nicht in einer Metaanalyse zusammengefasst werden. Für Metaanalysen soll die im Folgenden beschriebene Methodik eingesetzt werden.

Für die statistische Auswertung sollen primär die Ergebnisse aus Intention-to-treat-Analysen, so wie sie in den vorliegenden Dokumenten beschrieben sind, verwendet werden. Die Metaanalysen sollen in der Regel auf Basis von Modellen mit zufälligen Effekten⁴ erfolgen. In begründeten Ausnahmefällen sollen zusätzlich Modelle mit festen Effekten eingesetzt werden. Falls die für eine Metaanalyse notwendigen Schätzer für Lage und Streuung in den Studienunterlagen nicht vorliegen, sollen diese nach Möglichkeit aus den vorhandenen Informationen eigenständig berechnet beziehungsweise näherungsweise bestimmt werden.

Für kontinuierliche Variablen soll die Mittelwertdifferenz, gegebenenfalls standardisiert mittels Hedges' g, als Effektmaß eingesetzt werden. Bei binären Variablen sollen Metaanalysen primär anhand des Odds Ratio durchgeführt werden. In begründeten Ausnahmefällen können auch andere Effektmaße zum Einsatz kommen. Bei kategorialen Variablen soll ein geeignetes Effektmaß in Abhängigkeit vom konkreten Endpunkt und den verfügbaren Daten verwendet⁵ werden.

Die Effektschätzer und Konfidenzintervalle aus den Studien sollen mittels Forest Plots zusammenfassend dargestellt werden. Anschließend soll die Einschätzung einer möglichen Heterogenität der Studienergebnisse anhand des Maßes I^2 und des statistischen Tests auf Vorliegen von Heterogenität⁶ erfolgen. Ist die Heterogenität der Studienergebnisse nicht bedeutsam, soll der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt werden. Bei bedeutsamer Heterogenität sollen die Ergebnisse nur in begründeten

⁴ DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-188.

⁵ Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (Ed). *Cochrane handbook for systematic reviews of interventions*. Chichester: Wiley; 2008. S. 243-296.

⁶ Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557-560.

Ausnahmefällen gepoolt werden. Außerdem soll untersucht werden, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu zählen methodische Faktoren (siehe Abschnitt 4.2.5.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.2.5.5).

Beschreiben Sie die für Metaanalysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.

Für den direkten Vergleich von Teriflunomid versus Interferon beta-1a (Rebif®) kommen Metaanalysen nicht in Betracht, da mit TENERE nur eine Studie vorliegt, in der beide Medikamente direkt miteinander verglichen wurden.

Für den indirekten Vergleich mit Placebo als Brückenkomparator dagegen wurden in zwei Situationen Metaanalysen durchgeführt:

- Metaanalysen der Studien TEMSO und TOWER für den Vergleich Teriflunomid versus Placebo
- Metaanalysen des direkten Vergleichs (TENERE) und des indirekten Vergleichs, der auf den o. g. Metaanalysen aufbaut.

Für die Diskussion, Interpretation und weitere Verwendung in indirekten Vergleichen wurden ausschließlich die Ergebnisse von Metaanalysen mit zufälligen Effekten (nach DerSimonian und Laird) herangezogen.

Metaanalysen für Wirksamkeits-Endpunkte wurden generell auf Grundlage der Intention-to-treat Population (ITT) durchgeführt. Für Verträglichkeitsendpunkte wurden aus den Teriflunomid-Studienberichten die Ergebnisse auf Grundlage der Safety-Population herangezogen, die alle randomisierten und mit Studienmedikation anbehandelten Patienten umfasst und sie der de facto verabreichten Studienmedikation zuweist.

Bei binären Endpunkten wurde neben dem relativen Risiko zusätzlich das Odds Ratio als Abstandsmaß für die Analysen verwendet.

Bei Vorliegen von Einfach-Null-Studien (kein Ereignis in einer der beiden Behandlungsgruppen) wurde, wie im IQWiG-Dokument Allgemeine Methoden, Version 4 [17] vorgeschlagen, für die Berechnung und Metaanalyse von Odds Ratios auf die Peto-Odds-Ratio-Methode ausgewichen, für relative Risiken dagegen wurde vor den Berechnungen zu allen Zellenhäufigkeiten jeweils 0,5 addiert (s. Patienten mit Lymphopenie in Modul 4.3.2.1.4.5).

Bei Vorliegen von Doppel-Null-Studien (kein Ereignis in beiden Behandlungsgruppen) wurden die Analysen, wie im IQWiG-Dokument Allgemeine Methoden, Version 4 [17] vorgeschlagen, ausschließlich auf Grundlage von Risiko-Differenzen durchgeführt (s. Patienten, die während der Studie verstorben sind, Modul 4.3.2.1.4.3).

Beim Endpunkt "Reaktionen an der Injektionsstelle" wurde auf die Berechnung von Abstandsmaßen und indirekte Vergleiche ganz verzichtet, da in TEMSO und TOWER Placebo per os appliziert wurde, in PRISMS dagegen parenteral – Placebo kommt daher mit Blick auf den Endpunkt "Reaktionen an der Injektionsstelle" als Brückenkomparator nicht in Betracht. Da Teriflunomid als orale Therapie die Problematik von Reaktionen an der Injektionsstelle nicht hat, wird dieser Vorteil vor dem Hintergrund der Raten von Patienten mit Reaktionen an der Injektionsstelle in den Rebif-Armen der Studien TENERE und PRISMS diskutiert.

Für die Beurteilung der Heterogenität der in Metaanalysen zusammengefassten Studien wurde primär die Überschreitungswahrscheinlichkeit p_Q auf Grundlage der Cochran Q-Statistik herangezogen. Das I^2 -Maß wird zusätzlich präsentiert.

Das I^2 -Maß dient zur groben Einschätzung der Heterogenität, folgende Aussagen zur Heterogenität leiten sich aus dem entsprechenden Wert ab:

- 0 – 40 % = wahrscheinlich unbedeutende Heterogenität
- 30 – 60 % = mittelmäßige Heterogenität
- 50 – 90 % = substanzielle Heterogenität und
- 75 – 100 % = erhebliche Heterogenität [17].

Mit Blick auf p_Q werden folgende Fälle unterschieden:

- $p_Q \geq 0,2$

In diesem Fall wurde eine gute Homogenität der in einer Metaanalyse zusammenzuführenden Studien als gegeben angesehen.

- $p_Q < 0,2$ und *deutliche gleichgerichtete Effekte* der Studien

In diesem Fall wurde eine akzeptable Homogenität der in einer Metaanalyse zusammenzuführenden Studien als gegeben angesehen. Mögliche Gründe für die beobachtete Heterogenität werden diskutiert.

- *gleichgerichtet* wurde dabei in dem Sinne interpretiert, dass die Punktschätzer aller beteiligten Studien auf der gleichen Seite (bezogen auf RR, OR, HR = 1 bzw. RD, MD = 0) liegen.
- *deutlich* wurde subjektiv anhand des Forest Plots bewertet und nicht an Hand des Vorliegens signifikanter Unterschiede in den einzelnen Studien. Dies ist dadurch gerechtfertigt, dass in den Studien die Fallzahlplanungen nur für den primären Endpunkt ausgelegt waren. Die statistische Power für ein jeweils

signifikantes Ergebnis in den übrigen Endpunkten ist damit unklar und ein signifikantes Ergebnis nicht zu erwarten.

- $p_Q < 0,2$ und keine deutlichen gleichgerichteten Effekte der Studien

In diesem Fall wird von einer deutlichen Heterogenität der Studien ausgegangen. Die Ergebnisse von Metaanalysen sind in dieser Situation nur unter Vorbehalt zu interpretieren – im Vordergrund der Diskussion stehen mögliche Gründe für die beobachtete deutliche Heterogenität.

4.2.5.4 Sensitivitätsanalysen

Zur Einschätzung der Robustheit der Ergebnisse sollen Sensitivitätsanalysen hinsichtlich methodischer Faktoren durchgeführt werden, wenn sich die Studien bezüglich dieser Faktoren unterscheiden. Die methodischen Faktoren bilden sich aus den im Rahmen der Informationsbeschaffung und -bewertung getroffenen Entscheidungen, zum Beispiel die Festlegung von Cut-off-Werten für Erhebungszeitpunkte oder die Wahl des Effektmaßes. Insbesondere die Einstufung des Verzerrungspotenzials der Ergebnisse in die Kategorien „hoch“ und „niedrig“ soll für Sensitivitätsanalysen verwendet werden.

Das Ergebnis der Sensitivitätsanalysen kann die Einschätzung der Aussagekraft der Nachweise beeinflussen.

Begründen Sie die durchgeführten Sensitivitätsanalysen oder den Verzicht auf Sensitivitätsanalysen. Beschreiben Sie die für Sensitivitätsanalysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.

4.2.5.4.1 Als sekundäre Analysen der patientenrelevanten Endpunkte

In den Studien sind für einzelne Endpunkte Sensitivitätsanalysen durchgeführt worden. Diese Sensitivitätsanalysen umfassten, z. B.:

- Auswertung der Per-Protokoll-Population
- Analysen mit zusätzlicher Berücksichtigung von Daten, die während der Follow-up-Periode erhoben worden sind
- Analyse unter Verwendung aller Schübe (ungeachtet der Tatsache, ob es sich um bestätigte Schübe handelt)

Die Ergebnisse der Sensitivitätsanalyse sind Bestandteil der entsprechenden Studienberichte und nicht im Rahmen des vorliegenden Dossiers aufgeführt. Sämtliche durchgeführte Sensitivitätsanalysen unterstützten das jeweilige Ergebnis der Primäranalyse, so dass für Details auf die jeweiligen Studienberichte verwiesen wird [31, 32, 35, 36].

4.2.5.4.2 Im Rahmen von Metaanalysen und indirekten Vergleichen

Im Rahmen von Metaanalysen und indirekten Vergleichen boten sich für einen Wirksamkeits-Endpunkt Sensitivitätsanalysen an:

- Anteil schubfreier Patienten.

Bei der Auswertung des Anteils schubfreier Patienten fällt auf, dass in den placebo-kontrollierten Teriflunomid-Studien der Anteil schubfreier Patienten zwischen 49,3% und 67,0% liegt (im RRMS-Teilkollektiv betrug die entsprechende Spanne 48,6 % - 67,0 %), in PRISMS dagegen zwischen 16% und 32%. Dieser Unterschied ist nicht ohne Konsequenzen für die Punktschätzer von relativen Risiken, insbesondere für die Länge der zugehörigen Konfidenzintervalle.

Metaanalysen und indirekte Vergleiche wurden daher nicht nur für den Anteil schubfreier Patienten sondern auch für das *Gegenereignis*, d. h. den Anteil von Patienten mit mindestens einem Schub durchgeführt. Für das Odds Ratio als Abstandsmaß ändern sich durch den Übergang zum Gegenereignis die Ergebnisse nur insofern, als Punktschätzer und Konfidenzintervallgrenzen durch ihren Kehrwert zu ersetzen sind bzw. die Forest Plots auf der logarithmischen Skala bzgl. der Ordinate zu $OR=1$ achsensymmetrisch liegen. Für das relative Risiko als Abstandsmaß dagegen ergeben sich in den Metaanalysen und im indirekten Vergleich für den Endpunkt "schubfrei" und das Gegenereignis "mindestens ein Schub" durchaus unterschiedliche Ergebnisse, die in Modul 5 abgelegt sind.

4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren

Die Ergebnisse sollen hinsichtlich potenzieller Effektmodifikatoren, das heißt klinischer Faktoren, die die Effekte beeinflussen, untersucht werden. Dies können beispielsweise direkte Patientencharakteristika (Subgruppenmerkmale) sowie Spezifika der Behandlungen (z. B. die Dosis) sein. Im Gegensatz zu den in Abschnitt 4.2.5.4 beschriebenen methodischen Faktoren für Sensitivitätsanalysen besteht hier das Ziel, mögliche Effektunterschiede zwischen Patientengruppen und Behandlungsspezifika aufzudecken. Eine potenzielle Effektmodifikation soll anhand von Homogenitäts- bzw. Interaktionstests oder von Interaktionstermen aus Regressionsanalysen (mit Angabe von entsprechenden Standardfehlern) untersucht werden. Subgruppenanalysen auf der Basis individueller Patientendaten haben in der Regel eine größere Ergebnissicherheit als solche auf Basis von Meta-Regressionen oder Metaanalysen unter Kategorisierung der Studien bezüglich der möglichen Effektmodifikatoren, sie sind deshalb zu bevorzugen. Es sollten, soweit sinnvoll, folgende Faktoren bezüglich einer möglichen Effektmodifikation berücksichtigt werden:

- Geschlecht
- Alter
- Krankheitsschwere bzw. -stadium

Sollten sich aus den verfügbaren Informationen Anhaltspunkte für weitere mögliche Effektmodifikatoren ergeben, können diese ebenfalls begründet einbezogen werden. Die

Ergebnisse von in Studien a priori geplanten und im Studienprotokoll festgelegten Subgruppenanalysen für patientenrelevante Endpunkte sind immer darzustellen.

Bei Identifizierung möglicher Effektmodifikatoren kann gegebenenfalls eine Präzisierung der aus den für die Gesamtgruppe beobachteten Effekten abgeleiteten Aussagen erfolgen. Ergebnisse von Subgruppenanalysen können die Identifizierung von Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen unterstützen.

Benennen Sie die durchgeführten Subgruppenanalysen. Begründen Sie die durchgeführten Subgruppenanalysen bzw. die Untersuchung von Effektmodifikatoren oder den Verzicht auf solche Analysen. Beschreiben Sie die für diese Analysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen. Begründen Sie die Wahl von Trennpunkten, wenn quantitative Merkmale kategorisiert werden. Verwenden Sie dabei nach Möglichkeit die in dem jeweiligen Gebiet gebräuchlichen Einteilungen und begründen Sie etwaige Abweichungen. Kennzeichnen Sie in einzelnen Studien a priori geplante Subgruppenanalysen.

In den drei eingeschlossenen Studien TEMSO, TOWER und TENERE waren Subgruppenanalysen a priori geplant, die unter anderem auch die Faktoren Geschlecht, Alter und Krankheitsschwere- bzw. -stadium als mögliche Effektmodifikatoren berücksichtigen. Dementsprechend waren nach den Vorgaben keine zusätzlichen Subgruppenanalysen durchzuführen. Alle Subgruppenanalysen, die für die Studien TEMSO und TOWER aufgeführt sind, waren vorab geplant. Für die TENERE-Studie wurde neben den a priori geplanten Subgruppen, zusätzlich eine Subgruppenanalyse nach Vorbehandlung mit krankheitsmodifizierenden MS-Wirkstoffen durchgeführt.

TEMSO

Folgende Subgruppenanalysen wurden für den primären Effektivitätspunkt, die jährliche Schubrate, durchgeführt:

- Geschlecht
- Ethnische Zugehörigkeit
- Alter (<38, ≥38)
- Region (Osteuropa, Westeuropa, Amerika)
- Baseline EDSS-Stratum
- Anzahl Schübe in den vorangegangenen 2 Jahren (0, 1, 2, 3, und ≥4)
- MS subtypes (RRMS, SPMS, PRMS)
- Vorherige Behandlung mit krankheitsmodifizierenden MS-Wirkstoffen (ja, nein)

- Anzahl an Gadolinium (Gd)-anreichernden (aktiven) Läsionen (Baseline) ($0, \geq 1$)
- Krankheitslast im MRT (Baseline) ($<13, \geq 13$)

Um Inkonsistenzen des Behandlungseffektes in verschiedenen Subgruppen nachzuweisen, wurde die GEE-Methode (GEE = generalized estimation equation) angewendet; jeder Subgruppenfaktor wurde einzeln in einem Behandlung-Subgruppe-Interaktionstest analysiert.

Das Modell beinhaltete Behandlung, EDSS-Strata, Region, Subgruppe und Behandlung-Subgruppe-Interaktion als fixe Effekte und log-transformierte standardisierte Studiendauer als Offset-Variable.

Für das wesentliche sekundäre Zielkriterium, die Zeit bis zur Behinderungsprogression, wurden die gleichen Subgruppen analysiert wie für den primären Endpunkt durchgeführt.

Alle aufgeführten Subgruppenanalysen waren a priori im Studienprotokoll definiert.

Bei einem p-Wert von unter 0,05 wurde auf eine qualitative Interaktion geprüft.

TOWER

Folgende Subgruppenanalysen wurden für den primären Effektivitätseindpunkt, die jährliche Schubrate, durchgeführt:

- Geschlecht
- Ethnische Zugehörigkeit
- Alter ($<38, \geq 38$)
- Region (Osteuropa, Westeuropa und Afrika, Amerika, Asien und Australien)
- Baseline EDSS-Stratum
- Anzahl Schübe im vorangegangenen Jahr ($\leq 1, 2, \text{ und } \geq 3$)
- Anzahl Schübe in den vorangegangenen 2 Jahren ($0, 1, 2, 3, \text{ und } \geq 4$)
- MS Subtypen (RRMS, SPMS, PRMS)
- Vorherige Behandlung mit krankheitsmodifizierenden MS-Wirkstoffen (ja, nein)
- Gebiet (USA, nicht USA)

Um Inkonsistenzen des Behandlungseffektes in verschiedenen Subgruppen nachzuweisen, wurde die GEE-Methode (GEE = generalized estimation equation) angewendet; jeder Subgruppenfaktor wurde einzeln in einem Behandlung-Subgruppe-Interaktionstest analysiert.

Das Modell beinhaltet Behandlung, EDSS-Strata, Region, Subgruppe und Behandlung-Subgruppe-Interaktion als fixe Effekte und log-transformierte standardisierte Studiendauer als Offset-Variable.

Für das wesentliche sekundäre Zielkriterium, die Zeit bis zur Behinderungsprogression, wurden die gleichen Subgruppen analysiert wie für den primären Endpunkt durchgeführt.

Alle aufgeführten Subgruppenanalysen waren a priori im Studienprotokoll definiert.

Bei einem p-Wert von unter 0,05 wurde auf eine qualitative Interaktion geprüft.

TENERE

Folgende a priori festgelegte Subgruppenanalysen wurden für den primären Effektivitätseffekt, das Therapieversagen (Time to Failure), durchgeführt:

- Geschlecht
- Alter (<38, ≥38)
- Region
- Baseline EDSS-Stratum
- Anzahl Schübe in den vorangegangenen 2 Jahren (0, 1, 2, 3, und ≥4)

Um Inkonsistenzen des Behandlungseffektes in verschiedenen Subgruppen nachzuweisen, wurde ein Proportional-Hazards-Modell angewendet, unter Verwendung eines Behandlung-Subgruppe-Interaktionstests. Jeder Subgruppenfaktor wurde einzeln analysiert.

Das Modell beinhaltet Behandlung, Subgruppe und Behandlung-Subgruppe-Interaktion als fixe Effekte.

Für den wesentlichen sekundären Effektivitätseffekt, die jährliche Schubrate, wurden die gleichen Subgruppen analysiert, wie für den primären Endpunkt vorgeschrieben.

Zusätzlich zu den a priori festgelegten Subgruppenanalysen wurde getestet, ob es sich bei der Vorbehandlung mit einem krankheitsmodifizierenden Wirkstoff um einen Effektmodifikator handelt.

Bei einem p-Wert von unter 0,05 wurde auf eine qualitative Interaktion geprüft.

4.2.5.6 Indirekte Vergleiche

Zurzeit sind internationale Methoden in der Entwicklung, um indirekte Vergleiche zu ermöglichen. Nicht adjustierte indirekte Vergleiche (d. h. Vergleiche einzelner Behandlungsgruppen aus verschiedenen Studien ohne Bezug zu einem gemeinsamen Komparator) stellen dabei keine valide Analysemethoden dar, der Einsatz einfacher adjustierter

indirekter Vergleiche ist möglich⁷. Komplexe Verfahren für den simultanen Vergleich von mehr als zwei Therapien unter Berücksichtigung sowohl direkter als auch indirekter Vergleiche werden in der Literatur unterschiedlich bezeichnet, z. B. als „Mixed-Treatment-Comparison(MTC)-Metaanalysen“⁸, „Multiple-Treatment-Metaanalysen“⁹ oder auch „Netzwerk-Metaanalysen“¹⁰, sie gehen aber im Prinzip von denselben wesentlichen Annahmen aus.

Grundannahme für solche komplexen Analysen ist die Annahme der Konsistenz innerhalb des zu analysierenden Netzwerkes. Als Inkonsistenz wird dabei die Diskrepanz zwischen dem Ergebnis eines direkten und eines oder mehreren indirekten Vergleichen verstanden, die nicht mehr nur durch Zufallsfehler oder Heterogenität erklärbar ist¹¹. Insgesamt ist es notwendig, die zugrunde liegende Methodik genau und reproduzierbar zu beschreiben und die Annahme der Konsistenz zu untersuchen¹².

⁷ Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D'Amico R et al. Indirect comparisons of competing interventions. *Health Technol Assess* 2005; 9(26): 1-148.

⁸ Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. *Stat Med* 2004; 23(20): 3105-3124.

⁹ Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. *BMJ* 2005; 331(7521): 897-900.

¹⁰ Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. *Stat Methods Med Res* 2008; 17(3): 279-301.

¹¹ B. Schöttker, D. Lühmann, D. Boulkhemair, and H. Raspe. Indirekte Vergleiche von Therapieverfahren. *Schriftenreihe Health Technology Assessment Band 88, DIMDI, Köln, 2009.*

¹² Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DJ. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. *BMJ* 2009; 338: b1147.

Beschreiben Sie detailliert und vollständig die zugrunde liegende Methodik des indirekten Vergleichs. Dabei sind mindestens folgende Angaben notwendig:

- *Genaue Spezifikation des statistischen Modells inklusive aller Modellannahmen. Bei Verwendung eines Bayesianischen Modells sind dabei auch die angenommenen A-priori-Verteilungen (falls informative Verteilungen verwendet werden, mit Begründung), die Anzahl der Markov-Ketten und deren Startwerte und Länge zu spezifizieren.*
- *Art der Prüfung der Homogenität der Ergebnisse direkter paarweiser Vergleiche.*
- *Art der Prüfung der Konsistenz zwischen den Ergebnissen direkter und indirekter Vergleiche.*
- *Bilden Sie den Code des Computerprogramms in lesbarer Form ab und geben Sie an, welche Software Sie zur Berechnung eingesetzt haben (ggf. inklusive Spezifizierung von Modulen, Prozeduren, Packages etc.; siehe auch Modul 5 zur Ablage des Programmcodes).*
- *Art und Umfang von Sensitivitätsanalysen.*

Indirekte Vergleiche wurden ausgehend von den in 4.2.5.3 beschriebenen Metaanalysen mit zufälligen Effekten (nach DerSimonian und Laird) durchgeführt:

- Metaanalysen der Studien TEMSO und TOWER für den Vergleich Teriflunomid versus Placebo

Die Beurteilung der Heterogenität der in Metaanalysen zusammengefassten Studien erfolgt dabei, wie in 4.2.5.3 beschreiben, primär auf Grundlage der Überschreitungswahrscheinlichkeit p_Q der Cochrane Q-Statistik. Das I^2 -Maß wird zusätzlich präsentiert. Das I^2 -Maß dient zur groben Einschätzung der Heterogenität, folgende Aussagen zur Heterogenität leiten sich aus dem entsprechenden Wert ab:

- 0 – 40 % = wahrscheinlich unbedeutende Heterogenität
- 30 – 60 % = mittelmäßige Heterogenität
- 50 – 90 % = substanzielle Heterogenität und
- 75 – 100 % = erhebliche Heterogenität [17].

Für den indirekten Vergleich wurden endpunktweise die Ergebnisse der Metaanalysen über Placebo als Brückenkompator mit den Ergebnissen der PRISMS-Studie nach der Methode von Bucher [6] verknüpft. Angegeben werden hierzu der Punktschätzer für das jeweilige Abstandsmaß, das zugehörige 95%-Konfidenzintervall und der p-Wert unter der Nullhypothese gleicher Effekte unter Teriflunomid 14 mg und Interferon beta-1a, s.c., dreimal pro Woche 44 µg (Rebif 44 µg).

Die Ergebnisse des direkten Vergleichs aus TENERE und des indirekten Vergleichs wurden für jeden Endpunkt in einer weiteren Metaanalyse mit zufälligen Effekten (nach DerSimonian und Laird) zusammengeführt. Die Bewertung der Konsistenz erfolgte dabei auf Grundlage der Überschreitungswahrscheinlichkeit p_Q der Cochran Q-Statistik. Das I^2 -Maß wird zusätzlich präsentiert (Einteilung s. oben).

Mit Blick auf p_Q werden folgende Fälle unterschieden:

- $p_Q \geq 0,2$
In diesem Fall wurde die Konsistenz des indirekten Vergleichs mit dem Ergebnis aus TENERE als gegeben angesehen.
- $p_Q < 0,2$ und *deutliche gleichgerichtete Effekte* in TENERE und dem indirekten Vergleich
In diesem Fall wurde eine akzeptable Konsistenz zwischen TENERE und dem indirekten Vergleich als gegeben angesehen. Mögliche Gründe für gewisse Defizite in der Konsistenz werden diskutiert.
 - *gleichgerichtet* wurde dabei in dem Sinne interpretiert, dass die Punktschätzer aus TENERE und dem indirekten Vergleich auf der gleichen Seite (bezogen auf RR, OR, HR = 1 bzw. RD, MD = 0) liegen.
 - *deutlich* wurde subjektiv an Hand des Forest Plots bewertet und nicht an Hand des Vorliegens signifikanter Unterschiede in TENERE und dem indirekten Vergleich. Dies erscheint dadurch gerechtfertigt, dass in TENERE die Fallzahlplanungen nur für den primären Endpunkt ausgelegt war und dem post-hoc durchgeführten indirekten Vergleich keine Fallzahlplanung zu Grunde liegt. Die statistische Power für ein signifikantes Ergebnis in den übrigen Endpunkten von TENERE und in dem indirekten Vergleich generell ist damit unklar und ein jeweils signifikantes Ergebnis nicht zu erwarten.
- $p_Q < 0,2$ und keine *deutlichen gleichgerichteten Effekte* in TENERE und dem indirekten Vergleich
In diesem Fall wird von einer deutlichen Inkonsistenz zwischen TENERE und dem indirekten Vergleich ausgegangen. Die Ergebnisse der Metaanalyse sind in dieser Situation nur unter Vorbehalt zu interpretieren – im Vordergrund der Diskussion stehen mögliche Gründe für die beobachtete deutliche Inkonsistenz.

Beim Endpunkt "Reaktionen an der Injektionsstelle" wurde auf die Berechnung von Abstandsmaßen und indirekte Vergleiche ganz verzichtet, da in TEMSO und TOWER Placebo oral appliziert wurde, in PRISMS dagegen parenteral – Placebo kommt daher mit Blick auf den Endpunkt "Reaktionen an der Injektionsstelle" als Brückenkomparator nicht in Betracht. Da Teriflunomid als orale Therapie die Problematik von Reaktionen an der Injektionsstelle nicht hat, wird dieser Vorteil vor dem Hintergrund der Raten von Patienten mit Reaktionen an der Injektionsstelle in den Rebif[®]-Armen der Studien TENERE und PRISMS diskutiert.

Die Realisation der Metaanalysen erfolgte in MIX Ver. 1.7 für Windows, einem validierten und geschützten Add-in für Excel 2003 [37, 38]. Die Realisation der indirekten Vergleiche nach Bucher [6] erfolgten in Excel 2003 unter Microsoft Windows XP Professional. Die Ergebnisse aus MIX (weitere Informationen unter <http://www.meta-analysis-made-easy.com/>) und die Berechnungen der indirekten Vergleiche sind in Modul 5 als ungeschützte Excel-Files einsehbar.

Im Rahmen der indirekten Vergleiche wurden die in 4.2.5.4 beschriebenen Sensitivitätsanalysen für den Anteil schubfreier Patienten durchgeführt.

4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen

In den nachfolgenden Abschnitten sind die Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen zu beschreiben. Abschnitt 4.3.1 enthält dabei die Ergebnisse aus randomisierten kontrollierten Studien, die mit dem zu bewertenden Arzneimittel durchgeführt wurden (Evidenzstufen Ia/Ib).

Abschnitt 4.3.1.3.13 enthält weitere Unterlagen anderer Evidenzstufen, sofern diese aus Sicht des pharmazeutischen Unternehmers zum Nachweis des Zusatznutzens erforderlich sind. Diese Unterlagen teilen sich wie folgt auf:

- Randomisierte, kontrollierte Studien für einen indirekten Vergleich mit der zweckmäßigen Vergleichstherapie, sofern keine direkten Vergleichsstudien mit der zweckmäßigen Vergleichstherapie vorliegen oder diese keine ausreichenden Aussagen über den Zusatznutzen zulassen (Abschnitt 4.3.2.1)
- Nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2)
- Weitere Untersuchungen (Abschnitt 4.3.2.3)

4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel

4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel

4.3.1.1.1 Studien des pharmazeutischen Unternehmers

Benennen Sie in der nachfolgenden Tabelle alle Studien, die an die Zulassungsbehörde übermittelt wurden (Zulassungsstudien), sowie alle Studien, für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war.

Von den randomisierten kontrollierten Studien, die der Zulassungsbehörde im Zulassungsdossier übermittelt wurden, sollen die Studien, deren Studienberichte im Abschnitt 5.3.5 des Zulassungsdossiers enthalten sind, aufgeführt werden. Darüber hinaus sollen alle randomisierten kontrollierten Studien, für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, aufgeführt werden.

Benennen Sie nur randomisierte, kontrollierte Studien, die ganz oder teilweise innerhalb des in diesem Dokument beschriebenen Anwendungsgebiets durchgeführt wurden. Fügen Sie dabei für jede Studie eine neue Zeile ein.

Folgende Informationen sind in der Tabelle darzulegen: Studienbezeichnung, Angabe „Zulassungsstudie ja/nein“, Studienstatus (abgeschlossen, abgebrochen, laufend), Studiendauer und Therapiearme. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile.

Tabelle 4-11: Liste der Studien des pharmazeutischen Unternehmers – RCT^a mit dem zu bewertenden Arzneimittel

Studienbezeichnung	Zulassungsstudie (ja/nein)	Status (abgeschlossen / abgebrochen / laufend)	Studiendauer ^b	Therapiearme
HMR1726D/2001 (Phase II Studie als Monotherapie)	ja	abgeschlossen	36 Wochen	Teriflunomid 7 mg Teriflunomid 14 mg Placebo
HMR1726D/2002 ^e (LTS6048, Verlängerung der Phase II Monotherapie-Studie)	ja	laufend	528 Wochen (oder bis Teriflunomid in dem Land, in dem der Patient lebt, kommerziell erhältlich ist)	Teriflunomid 7 mg Teriflunomid 14 mg
HMR1726D-2003 (PDY6045, Phase II Studie als Zusatztherapie)	ja	abgeschlossen	24 Wochen	Teriflunomid 7 mg/ Interferon beta Teriflunomid 14 mg/ Interferon beta Placebo/ Interferon beta
HMR1726D/2004 (PDY6046)	ja	abgeschlossen	24 Wochen	Teriflunomid 7 mg/ Glatirameracetat Teriflunomid 14 mg/ Glatirameracetat Placebo/ Glatirameracetat
HMR1726D/2005 (LTS6047, Verlängerung von PDY6045 und PDY6046)	ja	abgeschlossen	24 Wochen	Teriflunomid 7 mg/ Beta-Interferon Teriflunomid 14 mg/ Interferon beta Placebo/ Interferon beta Teriflunomid 7 mg/ Glatirameracetat Teriflunomid 14 mg/ Glatirameracetat Placebo/ Glatirameracetat

Studienbezeichnung	Zulassungsstudie (ja/nein)	Status (abgeschlossen / abgebrochen / laufend)	Studiendauer ^b	Therapiearme
TEMZO (HMR1726D/3001, EFC6049)	ja	abgeschlossen	108 Wochen	Teriflunomid 7 mg, Teriflunomid 14 mg, Placebo
TOWER (EFC 10531)	ja	abgeschlossen	48 bis 154 Wochen	Teriflunomid 7 mg, Teriflunomid 14 mg, Placebo
TENERE (EFC10891)	ja	abgeschlossen	48 bis 118 Wochen	Teriflunomid 7 mg, Teriflunomid 14 mg, Interferon beta-1a, s.c. bis zu 44 µg
Verlängerung der TEMZO-Studie ^e (HMR1726D/3004, LTS6050)	ja	laufend	288 Wochen (oder bis Teriflunomid in dem Land, in dem er Patient lebt, kommerziell erhältlich ist)	Teriflunomid 7 mg Teriflunomid 14 mg
TOPIC (HMR1726D-3005, EFC6260)	nein	laufend ^c	24 bis 108 Wochen	Teriflunomid 7 mg Teriflunomid 14 mg Placebo
TERACLES (EFC6058)	nein	abgeschlossen ^d	48 – 152 Wochen	Teriflunomid 7 mg Teriflunomid 14 mg Placebo
Verlängerung der TOPIC-Studie	nein	laufend	bis Teriflunomid in dem Land, in dem der Patient lebt, kommerziell erhältlich ist	Teriflunomid 7 mg Teriflunomid 14 mg
Verlängerung der TOWER-Studie ^e	nein	laufend	bis Teriflunomid in dem Land, in dem der Patient lebt, kommerziell erhältlich ist	Teriflunomid 14 mg
Verlängerung der TENERE-Studie ^e	nein	laufend	bis Teriflunomid in dem Land, in dem der Patient lebt, kommerziell erhältlich ist	Teriflunomid 14 mg

^aStudien, die keine RCT sind und Phase I-Studien sind nicht mit aufgeführt, dies umfasst u. a. Dosisfindungsstudien an gesunden Probanden oder auch z. B. TERIVA, eine Studie, in der die Immunantwort auf eine Grippeimpfung bei Patienten unter Teriflunomid-Behandlung untersucht wird, bei der es sich aber nicht um eine randomisierte, kontrollierte Studie handelt.

^bbezieht sich in allen Fällen auf die reine Behandlungsdauer, d. h. auf die Studiendauer ohne die Randomisierungs- und Eliminationsphasen

^cDie aktuelle Revisionen der MS-Diagnosekriterien erlauben eine frühere Diagnose der RMS, was das medizinische Umfeld, in dem klinische Studien zu CIS und MS generell durchgeführt werden, verändert hat. Vor diesem Hintergrund hat das Steering Committee die Entscheidung getroffen, die Rekrutierung zu stoppen und so die Anzahl der Patienten, die einer Behandlung mit Placebo ausgesetzt sind, möglichst gering zu halten. [39]

^dFragestellung in der TERACLES-Studie war die Effektivität von Teriflunomid im Vergleich zu Placebo jeweils in Kombination mit Interferon beta bei Patienten mit RMS, die unter laufender Interferon-beta-Therapie weiterhin klinische Aktivität zeigten. Mit Blick auf die aktuellen Entwicklungen im Bereich der MS-Therapie wurde die Kombination einer oralen und einer parenteralen Darreichungsform als ein im klinischen Alltag wenig präferiertes Therapieschema eingestuft. Dies spiegelte sich auch in einer sehr schleppenden Patientenrekrutierung für die Studie wider. Aus diesen Gründen wurde die Studie vorzeitig beendet. [40]

^edie Verlängerungsstudien zu HMR1726/2001, TEMSO, TOWER und TENERE sind Verlängerungen, die unter Wegfall des vorherigen Komparators Placebo oder Interferon beta-1a weiterlaufen. Sie sind in der Liste der Studien des pharmazeutischen Unternehmers aufgeführt, da es sich um Verlängerungen der für die Zulassung und die Nutzenbewertung maßgeblich relevanten Studien handelt.

Geben Sie an, welchen Stand die Information in Tabelle 4-11 hat, d. h. zu welchem Datum der Studienstatus abgebildet wird. Das Datum des Studienstatus sollte nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Die Informationen in Tabelle 4-11 bilden den Studienstatus zum 29.07.2013 ab.

Geben Sie in der nachfolgenden Tabelle an, welche der in Tabelle 4-11 genannten Studien nicht für die Nutzenbewertung herangezogen wurden. Begründen Sie dabei jeweils die Nichtberücksichtigung. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-12: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	Begründung für die Nichtberücksichtigung der Studie
HMR1726D/2001 (Phase II Studie als Monotherapie)	Studiendauer zu kurz (36 Wochen)
HMR1726D/2002 (LTS6048, Verlängerung der Phase II Monotherapie-Studie)	kein passender Komparator (in der Verlängerung Wegfall des Placebo-Arms)
HMR1726D-2003 (PDY6045, Phase II Studie als Zusatztherapie)	Studie zur Kombinationstherapie (andere Indikation als in der Zulassung)
HMR1726D/2004 (PDY6046)	Studie zur Kombinationstherapie (andere Indikation als in der Zulassung)
HMR1726D/2005 (LTS6047, Verlängerung von PDY6045 und PDY6046)	Studie zur Kombinationstherapie (andere Indikation als in der Zulassung)
Verlängerung der TEMSO-Studie (HMR1726D/3004, LTS6050)	kein passender Komparator (in der Verlängerung Wegfall des Placebo-Arms)
TOPIC (HMR1726D-3005, EFC6260)	Studie bei CIS-Patienten (andere Indikation als in der Zulassung),
TERACLES (EFC6058)	Studie zur Kombinationstherapie (andere Indikation als in der Zulassung)

Studienbezeichnung	Begründung für die Nichtberücksichtigung der Studie
Verlängerung der TOPIC-Studie	Studie bei CIS-Patienten (andere Indikation als in der Zulassung), nicht kontrolliert
Verlängerung der TOWER-Studie	kein passender Komparator (in der Verlängerung Wegfall des Placebo-Arms)
Verlängerung der TENERE-Studie	kein passender Komparator (in der Verlängerung Wegfall des Interferon-Arms)

Stellen Sie die abgebrochenen und laufenden Studien, die nicht in die Bewertung eingehen, in Anhang 4-D und Anhang 4-E dar.

4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche

Sofern eine bibliografische Literaturrecherche durchgeführt wurde, beschreiben Sie nachfolgend das Ergebnis dieser Recherche. Illustrieren Sie den Selektionsprozess und das Ergebnis der Selektion mit einem Flussdiagramm. Geben Sie dabei an, wie viele Treffer sich insgesamt (d. h. über alle durchsuchten Datenbanken) aus der bibliografischen Literaturrecherche ergeben haben, wie viele Treffer sich nach Entfernung von Dubletten ergeben haben, wie viele Treffer nach Sichtung von Titel und, sofern vorhanden, Abstract als nicht relevant angesehen wurden, wie viele Treffer im Volltext gesichtet wurden, wie viele der im Volltext gesichteten Treffer nicht relevant waren (mit Angabe der Ausschlussgründe) und wie viele relevante Treffer verblieben. Geben Sie zu den relevanten Treffern an, wie vielen Einzelstudien diese zuzuordnen sind. Listen Sie die im Volltext gesichteten und ausgeschlossenen Studien unter Nennung des Ausschlussgrunds in Anhang 4-C.

Falls die Recherche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel gemeinsam mit der Recherche zu einer anderen Fragestellung (z. B. zu indirekten Vergleichen) durchgeführt wurde, unterteilen Sie die Angaben zu relevanten Treffern und Studien entsprechend diesen Fragestellungen.

[Anmerkung: „Relevanz“ bezieht sich in diesem Zusammenhang auf die im Abschnitt 4.2.2 genannten Kriterien für den Einschluss von Studien in die Nutzenbewertung.]

Orientieren Sie sich bei der Erstellung des Flussdiagramms an dem nachfolgenden Beispiel.

Für die Recherche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel wurde eine bibliographische Literaturrecherche durchgeführt. Die Zulassung für Teriflunomid liegt bei Einreichen des vorliegenden Dossiers zwar keine 12 Monate zurück, jedoch wurde ein indirekter Vergleich durchgeführt. Deswegen wurde eine systematische Literaturrecherche nach Studien, die Teriflunomid mit Placebo vergleichen, durchgeführt. Da potentielle Treffer dieser Literaturrecherche auch für Modul 4.3.1 relevant wären, wird die entsprechende Suche an dieser Stelle präsentiert.

Eine systematische Literaturrecherche, die die Identifizierung von Studien, die die zweckmäßige Vergleichstherapie Interferon beta-1a 44 µg, dreimal die Woche subkutan (Rebif®) gegen Placebo vergleichen, ist Bestandteil des indirekten Vergleichs und wird entsprechend in Kapitel 4.3.2 dargestellt.

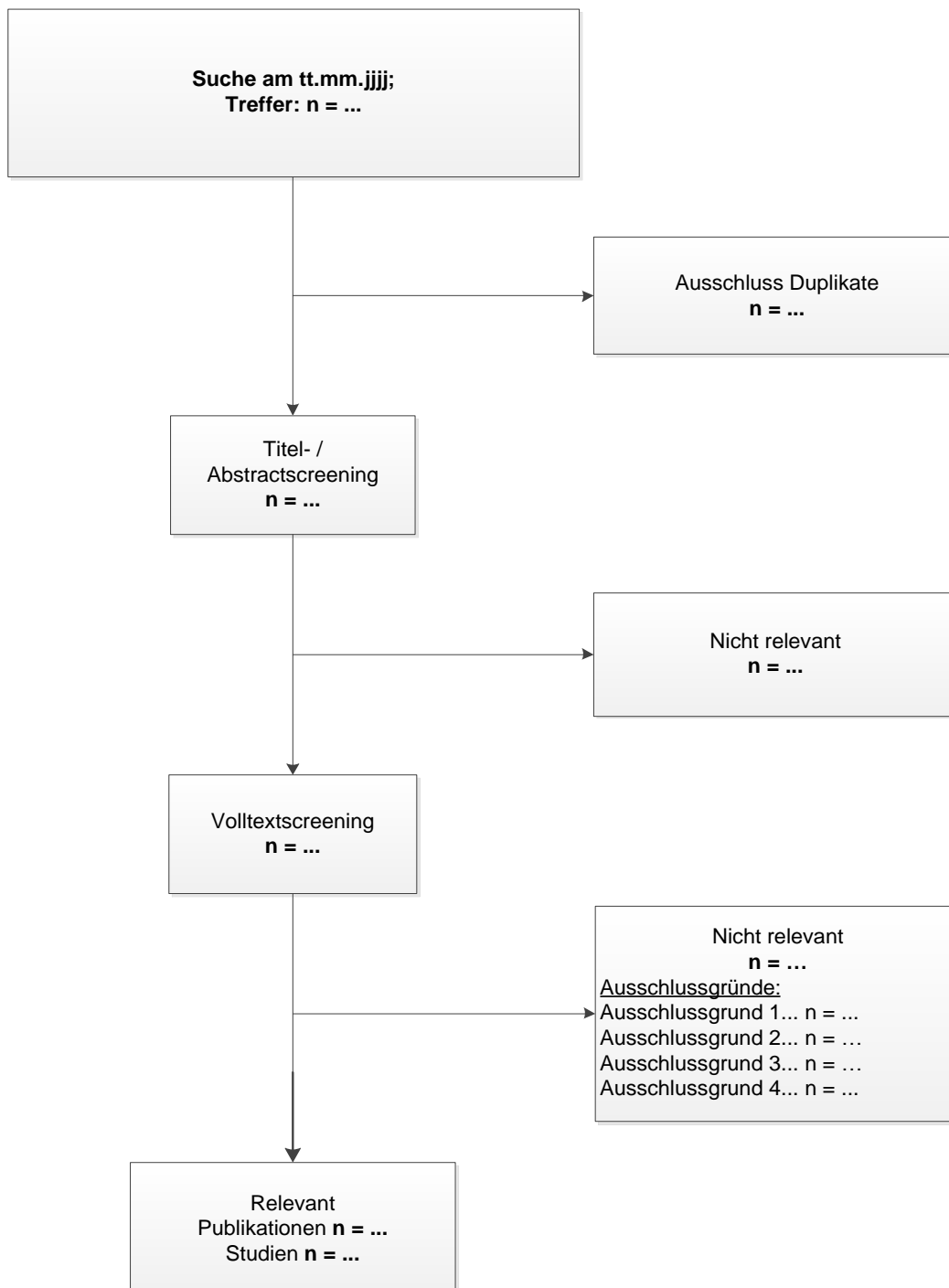
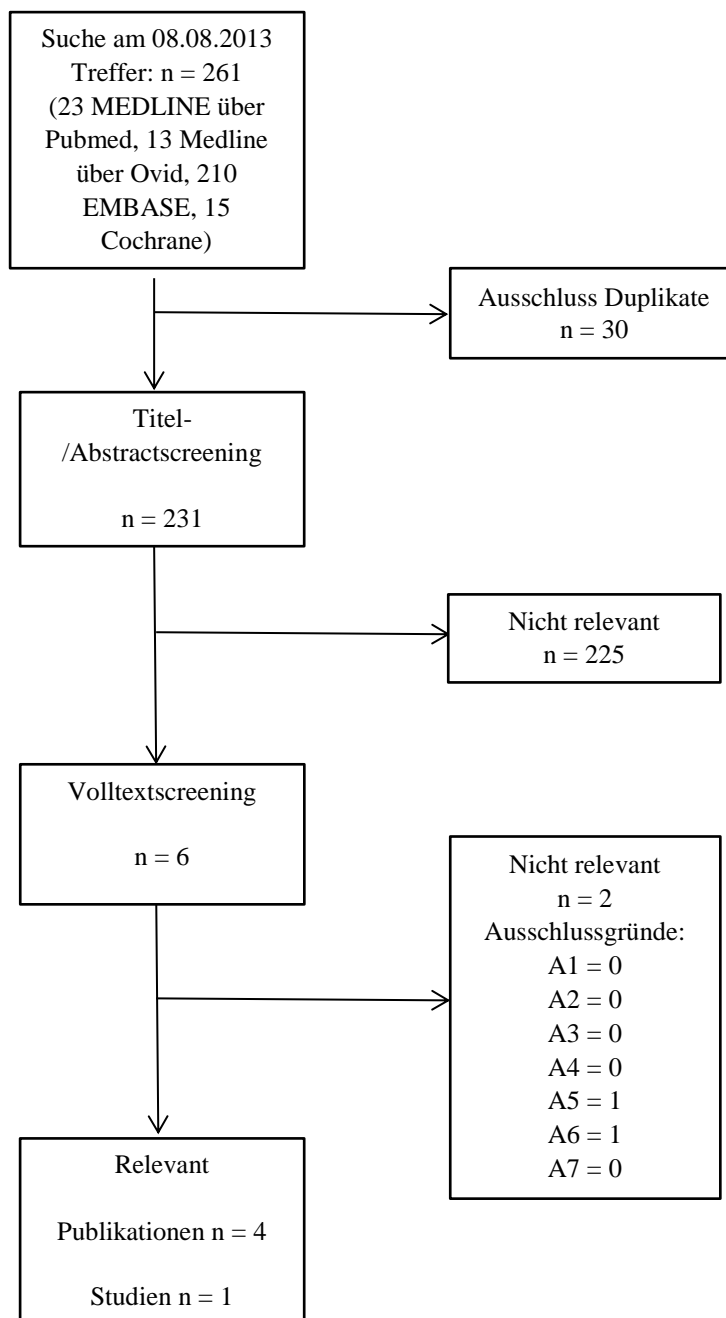


Abbildung 2: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten, kontrollierten Studien mit dem zu bewertenden Arzneimittel

Abbildung 4-3: Flussdiagramm für die systematische Literaturrecherche in MEDLINE, EMBASE und den Cochrane-Datenbanken



In der systematischen bibliographischen Literaturrecherche wurde eine relevante Studie für die Bewertung von Teriflunomid identifiziert, für diese Studie liegen vier relevante Publikationen vor. Bei der gefundenen Studie handelt es sich um die Zulassungsstudie TEMSO, die im Studienpool des pharmazeutischen Unternehmers bereits enthalten und in Tabelle 4-11 berücksichtigt ist. Die Suchstrategie ist in Anhang 4-A dokumentiert. Die Bewertung der Relevanz erfolgte anhand der in Kapitel 4.2.2.2 aufgeführten Ein- und Ausschlusskriterien A1 bis A7 (s. Tabelle 4-7), von den im Volltext gesichteten Treffern

wurde eine Studie ausgeschlossen, in der die placebokontrollierte Phase zu kurz war (Ausschlussgrund A6, Studie HMR1726D/2001 plus Verlängerung des pharmazeutischen Unternehmers), eine weitere Publikation wurde ausgeschlossen, weil es sich um ein Review-Artikel handelt (Ausschlussgrund A5).

Tabelle 4-13: Relevante Publikationen

Studie	Referenz
TEMSO	O'Connor P., Wolinsky J. S., Confavreux C., Comi G., Kappos L., Olsson T. P., et al. (2011): Randomized trial of oral teriflunomide for relapsing multiple sclerosis. The New England journal of medicine; 365(14):1293-303. [35]
TEMSO	O'Connor P. W., Lublin F. D., Wolinsky J. S., Confavreux C., Comi G., Freedman M. S., et al. (2013): Teriflunomide reduces relapse-related neurological sequelae, hospitalizations and steroid use. Journal of neurology 2013 [41]
TEMSO	Wolinsky J. S., Narayana P. A., Nelson F., Datta S., O'Connor P., Confavreux C., et al. (2013): Magnetic resonance imaging outcomes from a phase III trial of teriflunomide. Multiple sclerosis (Houndmills, Basingstoke, England);[42]
TEMSO	Miller A. E., O'Connor P., Wolinsky J. S., Confavreux C., Kappos L., Olsson T. P., et al. (2012): Pre-specified subgroup analyses of a placebo-controlled phase III trial (TEMSO) of oral teriflunomide in relapsing multiple sclerosis. Multiple sclerosis (Houndmills, Basingstoke, England); 18(11):1625-32. [43]

4.3.1.1.3 Studien aus der Suche in Studienregistern

Beschreiben Sie in der nachfolgenden Tabelle alle relevanten Studien, die durch die Suche in Studienregistern identifiziert wurden. Geben Sie dabei an, in welchem Studienregister die Studie identifiziert wurde und welche Dokumente dort zur Studie jeweils hinterlegt sind (z. B. Studienregistereintrag, Bericht über Studienergebnisse etc.). Geben Sie auch an, ob die Studie in der Liste der Studien des pharmazeutischen Unternehmers enthalten ist (siehe Tabelle 4-11) und ob die Studie auch durch die bibliografische Literaturrecherche identifiziert wurde (sofern eine solche durchgeführt wurde). Fügen Sie für jede Studie eine neue Zeile ein.

[Anmerkung: „Relevanz“ bezieht sich in diesem Zusammenhang auf die im Abschnitt 4.2.2 genannten Kriterien für den Einschluss von Studien in die Nutzenbewertung.]

Orientieren Sie sich bei Ihren Angaben an der beispielhaften ersten Tabellenzeile.

Tabelle 4-14: Relevante Studien aus der Suche in Studienregistern – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	Identifikationsorte (Name der Studienregister und Angabe der Zitate ^a)	Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja / nein)	Studie durch bibliografische Literaturrecherche identifiziert (ja / nein / n. d. ^b)
TEMSO	clinicaltrials.gov [NCT00134563], Registereintrag und Ergebnisse [44, 45] ICTRP Search Portal [EUCTR2004-000555-42-NO], Registereintrag [46]; EU ClinicalTrialsRegister, Registereintrag [47]	ja	ja
TOWER	clinicaltrials.gov [NCT00751881], Registereintrag und Ergebnisse [48, 49]; ICTRP Search Portal [EUCTR2007-004452-36-GB], Registereintrag [50]; EU ClinicalTrialsRegister [51], Registereintrag; PharmNet.Bund, Registereintrag [52]	ja	nein
TENERE	clinicaltrials.gov [NCT00883337], Registereintrag und Ergebnisse [53, 54]; ICTRP Search Portal [EUCTR2008-006226-34-ES], Registereintrag [55]; EU ClinicalTrialsRegister, Registereintrag [56]; PharmNet.Bund, Registereintrag [57]	ja	nein
a: Zitat des Studienregistereintrags sowie, falls vorhanden, der im Studienregister aufgelisteten Berichte über Studiendesign und/oder -ergebnisse. b: n. d. = bibliografische Literaturrecherche nicht durchgeführt.			

4.3.1.1.4 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel

Benennen Sie in der nachfolgenden Tabelle den aus den verschiedenen Suchschritten (Abschnitte 4.3.1.1.1, 4.3.1.1.2 und 4.3.1.1.3) resultierenden Pool relevanter Studien für das zu bewertende Arzneimittel, auch im direkten Vergleich zur zweckmäßigen Vergleichstherapie. Führen Sie außerdem alle für die Nutzenbewertung berücksichtigten Studien einschließlich der verfügbaren Quellen in Abschnitt 4.6 auf.

Folgende Informationen sind in der Tabelle darzulegen: Studienbezeichnung, Studienkategorie und verfügbare Datenquellen. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Hierbei sollen die Studien durch Zwischenzeilenüberschriften ggf. sinnvoll angeordnet werden, beispielsweise nach Therapieschema (Akut-/Langzeitstudien) und jeweils separat nach Art der Kontrolle (Placebo, zweckmäßige Vergleichstherapie, beides). Sollten Sie eine Strukturierung des Studienpools vornehmen, berücksichtigen Sie diese auch in den weiteren Tabellen in Modul 4.

Tabelle 4-15: Studienpool – RCT mit dem zu bewertenden Arzneimittel

Studie	Studienkategorie			verfügbare Datenquellen ^a		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	gesponserte Studie ^b (ja/nein)	Studie Dritter (ja/nein)	Studienbericht (ja/nein [Zitat])	Registereintrag ^c (ja/nein [Zitat])	Publikation (ja/nein [Zitat])
placebokontrolliert						
TEMZO	ja	ja	nein	ja [32]	ja [45, 47]	ja [35, 41-43]
TOWER	ja	ja	nein	ja [31]	ja [49, 50]	nein
aktivkontrolliert, zweckmäßige Vergleichstherapie(n)						
TENERE	ja	ja	nein	ja [22]	ja [54, 55]	nein
a: Bei Angabe „ja“ sind jeweils die Zitate der Datenquelle(n) mit anzugeben.						
b: Studie, für die der Unternehmer Sponsor war oder an der der Unternehmer anderweitig finanziell beteiligt war.						
c: Zitat der Studienregistereinträge sowie, falls vorhanden, der in den Studienregistern aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.						

4.3.1.2 Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel

4.3.1.2.1 Studiendesign und Studienpopulationen

Beschreiben Sie das Studiendesign und die Studienpopulation der in die Bewertung eingeschlossenen Studien mindestens mit den Informationen in den folgenden Tabellen. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Fügen Sie für jede Studie eine neue Zeile ein.

Weitere Informationen zu Studiendesign, Studienmethodik und Studienverlauf sind in Anhang 4-F zu hinterlegen.

Tabelle 4-16: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel

Studie	Studiendesign <RCT, doppelblind/einfach verblindet/offen, parallel/cross-over etc.>	Population <relevante Charakteristika, z. B. Schweregrad>	Interventionen (Zahl der randomisierten Patienten)	Studiendauer <ggf. Run-in, Behandlung, Nachbeobachtung>	Ort und Zeitraum der Durchführung	Primärer Endpunkt; patientenrelevante sekundäre Endpunkte
TEM SO	RCT, doppelblind, parallel	MS-Patienten (McDonald's Kriterien) Erwachsene (18-55 Jahre) EDSS ≤5,5 Schubförmiger Verlauf mit oder ohne Progression (RRMS, SPMS, PRMS) Mindestens 1 Schub im vorangegangenen Jahr oder 2 Schübe in den 2 vorangegangenen Jahren	Teriflunomid 7 mg (n = 365) Teriflunomid 14 mg (n = 358) Placebo (n = 363)	Behandlungsdauer: 108 Wochen Beobachtungsdauer: 128 Wochen (beinhaltet eine 4-Wochen- Screening-Periode, die Behandlungsphase und eine bis zu 16 Wochen dauernde Eliminationsphase (Wash-out))	126 aktive Zentren in 21 Ländern: Österreich, Kanada, Chile, Tschechien, Dänemark, Estland, Finnland, Frankreich, Deutschland, Italien, Niederlande, Norwegen, Polen, Portugal, Russland, Schweden, Schweiz, Türkei, Ukraine, Großbritannien, USA 09/2004 – 9/2011	Jährliche Schubrate; Zeit bis zur Behinderungsprogressi on (EDSS), MRT- bezogene Endpunkte, Fatigue (FIS), unerwünschte Ereignisse
TOWER	RCT, doppelblind, parallel	MS-Patienten (McDonald's Kriterien) Erwachsene (18-55 Jahre) EDSS ≤5,5 Schubförmiger Verlauf mit oder ohne Progression	Teriflunomid 7 mg (n = 408) Teriflunomid 14 mg (n = 372) Placebo (n = 389)	Behandlungsdauer : 48 bis 152 Wochen (je nach Zeitpunkt der Aufnahme in die Studie) Beobachtungsdauer: 56 bis 160 Wochen (beinhaltet eine 4- Wochen-Screening- Periode, die Behandlungsphase und	189 aktive Zentren in 26 Ländern: Australien, Österreich, Weißrussland, Belgien, Kanada, Chile, China, Tschechien, Estland, Frankreich, Deutschland, Griechenland, Mexiko, Niederlande,	Jährliche Schubrate; Zeit bis zur Behinderungsprogressi on (EDSS), Fatigue (FIS), Zeit bis zum ersten Schub, Anteil schubfreier Patienten, Anteil Patienten ohne Behinderungsprogressi on (nach 6 Monaten, 1

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

		(RRMS, SPMS, PRMS) Mindestens 1 Schub im vorangegangenen Jahr oder 2 Schübe in den 2 vorangegangenen Jahren		eine 4 Wochen dauernde Eliminationsphase)	Philippinen, Polen, Rumänien, Slowakei, Spanien, Schweden, Thailand, Tunesien, Türkei, Ukraine, Großbritannien, USA 08/2008 – 04/2012	Jahr, 2 Jahren), Mittlere Änderung des EDSS, Lebensqualität (SF-36), unerwünschte Ereignisse
TENERE	RCT, in Bezug auf den auswertenden Neurologen verblindet, parallel	MS-Patienten (McDonald's Kriterien) Erwachsene (≥18 Jahre) EDSS ≤5,5	Teriflunomid 7 mg (n= 109) Teriflunomid 14 mg (n= 111) Beta-Interferon 1a s. c. 44 µg ^a (n = 104)	Behandlungsdauer: 48 – 118 Wochen Beobachtungsdauer: 56 - 126 Wochen (beinhaltet eine 4-Wochen-Screening-Periode, die Behandlungsphase und eine bis zu 16 Wochen dauernde Eliminationsphase)	53 aktive Zentren in 13 Ländern: Belgien, Kanada, Tschechien, Frankreich, Deutschland, Griechenland, Ungarn, Italien, Polen, Spanien, Schweiz, Tunesien, Großbritannien 4/2009 – 9/2011	Therapieversagen („Time to Failure“); Schubrate, Fatigue (FIS), Patientenzufriedenheit (TSQM), unerwünschte Ereignisse

^aDie Administration von Rebif[®] erfolgte gemäß Fachinformation [34] mit einer Auftitrationsphase von 2 Wochen mit 8,8 µg Rebif, anschließenden 2 Wochen mit 22 µg Rebif und ab Woche 5 einer Erhaltungsdosis von 44 µg. Nach Studienprotokoll und entsprechend der Fachinformation war bei schlechter Verträglichkeit eine Reduktion auf die niedrigere 22 µg-Dosis erlaubt. Im Laufe der Studie wurde bei 12 Patienten eine entsprechende Dosisreduktion vorgenommen., 88,5 % der Patienten haben über die gesamte Studiendauer die höhere Dosierung von 44 µg erhalten. Die Dosis in der Vergleichsgruppe wird deshalb vereinfacht mit 44 µg angegeben.

Tabelle 4-17: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel

Studie	<Gruppe 1>	<Gruppe 2>	<Gruppe 3>	<i>ggf. weitere Spalten mit Behandlungscharakteristika z. B. Vorbehandlung, Behandlung in der Run-in- Phase etc.</i>
<Studie 1>	xxx 250 µg, 1 Inhalation bid + Placebo 2 Inhalationen bid	yyy 200 µg, 2 Inhalationen bid + Placebo 1 Inhalation bid		Vorbehandlung: zzz 1000 µg pro Tag, 4 Wochen vor Studienbeginn Bedarfsmedikation: aaa
TEMPO	Teriflunomid 7 mg qd	Teriflunomid 14 mg qd	Placebo qd	nicht zutreffend
TOWER	Teriflunomid 7 mg qd	Teriflunomid 14 mg qd	Placebo qd	nicht zutreffend
TENERE	Teriflunomid 7 mg qd	Teriflunomid 14 mg qd	Interferon- beta-1a s. c. 44 µg tiw	nicht zutreffend

qd = einmal täglich, tiw = dreimal wöchentlich

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-18: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel

Studie Gruppe	N (randomisierte Patienten)	Alter (Jahre) Mittelwert ±Standard- abweichung	Geschlecht w /m (%)	Anzahl Schübe im vorangegangenen Jahr Mittelwert ±Standard- abweichung	Zeit seit der ersten MS-Diagnose (in Jahren) Mittelwert ±Standard- abweichung	Zeit seit der ersten MS-Diagnose (in Jahren) Median	Baseline EDSS Mittelwert ±Standard- abweichung	Verlaufsform RRMS/SPMS /PRMS (%)
TEM SO								
T7	366 ^a	37,4 ±9,0	69,7/30,3	1,4 ±0,7	5,28 ±5,36	3,75 (0,1 – 27,6)	2,68 ±1,34	91,0/4,6/4,4
T14	359 ^a	37,8 ±8,2	71,0/29,0	1,3 ±0,7	5,58 ±5,48	3,67 (0,1 – 30,1)	2,67 ±1,24	92,8/3,3/3,9
Pb	363	38,4 ±9,0	75,8/24,2	1,4 ±0,7	5,13 ±5,59	3,50 (0,1 – 31,6)	2,68 ±1,34	90,6/6,1/3,3
TOWER								
T7	408	37,4 ±9,4	73,5/26,5	1,4 ±0,7	5,30 ±5,45	3,50 (0,0 – 29,1)	2,71 ±1,39	96,3/0,7/2,9
T14	372	38,2 ±9,4	69,4/30,6	1,4 ±0,7	5,27 ±5,90	3,25 (0,0 – 32,7)	2,71 ±1,35	98,9/0,5/0,5
Pb	389	37,9 ±9,3	70,2/29,8	1,4 ±0,0	4,92 ±5,66	3,17 (0,0 – 33,8)	2,69 ±1,36	97,4/1,0/1,5
TENERE								
T7	109	35,2 ±9,2	64,2/35,8	1,3 ±0,8	3,72 ±5,19	0,67 (0,1 – 23,4)	2,04 ±1,22	100/0/0
T14	111	36,8 ±10,3	70,3/29,7	1,4 ±0,8	3,68 ±6,24	0,75 (0,1 – 36,5)	2,33 ±1,35	97,3/0,9/1,8
Rb	104	37,0 ±10,6	68,3/31,7	1,2 ±1,0	3,82 ±5,69	1,00 (0,1 – 30,3)	2,04 ±1,19	100/0/0

T7 = Teriflunomid 7 mg, T14 = Teriflunomid 14 mg, Pb = Placebo, Rb = Rebif[®] (Beta-Interferon-1a s. c. 44 µg)^asowohl in der 7mg- als auch in der 14mg- Teriflunomid-Gruppe gab es jeweils einen Patienten, der keine Studienmedikation erhalten hat und aus der ITT-Analyse ausgeschlossen wurde (s. Definition der ITT-Population in Kapitel 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien)

Beschreiben Sie die Studien zusammenfassend. Sollte es Unterschiede zwischen den Studien geben, weisen Sie in einem erläuternden Text darauf hin.

Es konnten drei randomisierte, kontrollierte Studien identifiziert werden, die die unter Kapitel 4.2.2.1 definierten Einschlusskriterien erfüllen (TEMSO, TOWER und TENERE). Alle drei Studien bewerten die Effektivität und Verträglichkeit von Teriflunomid in zwei Dosierungen (7 mg und 14 mg) bei der Therapie von Patienten mit schubförmigen Formen der MS (>90% RRMS). Die Patientenpopulationen wiesen bei Studienbeginn eine ähnliche Schwere der Erkrankung auf. Die Patienten hatten bei Studienbeginn im Schnitt 1,4 Schübe im vorangegangenen Jahr und in den Studien TEMSO und TOWER einen mittleren EDSS von 2,7, in der Studie TENERE lag der mittlere EDSS mit 2,1 ein wenig niedriger. Die Geschlechterverteilung war in allen drei Studien ähnlich und entsprach der Geschlechterverteilung in der Zielpopulation, ca. 70 % der MS-Patienten sind weiblich (s. Modul 3). Auch das mittlere Alter der Patienten zwischen 30 und 40 Jahren ist für Studien im Anwendungsgebiet A als geeignet anzusehen (s. Modul 3).

Bei den beiden Studien TEMSO und TOWER handelt es sich um doppelblinde, placebo-kontrollierte Studien. Beide Studien sind von ähnlicher Größe (ca. 1.000 eingeschlossene Patienten) und haben die jährliche Schubrate als primären Endpunkt und die Zeit bis zur Behinderungsprogression als sekundären Endpunkt. Der wesentliche Unterschied der beiden Studien besteht in der geplanten Dauer der placebokontrollierten Behandlungsphase. Bei TEMSO ist die placebokontrollierte Phase auf 108 Wochen angelegt. Bei TOWER setzt sich die Dauer der placebokontrollierten Phase der Studie aus Rekrutierungszeitraum plus 48 Wochen Behandlungszeitraum des letzten in die Studie eingeschlossenen Patienten zusammen. Entsprechend betrug die Behandlungsdauer abhängig vom Zeitpunkt des Einschlusses in die Studie zwischen 48 und 152 Wochen.

TENERE unterscheidet sich maßgeblich von den beiden erstgenannten Studien. Die Studienpopulation war kleiner (324 Patienten) und die Studie war aktivkontrolliert (Vergleichstherapie Rebif[®] (Interferon beta-1a, subkutan, 44 µg, dreimal wöchentlich)). Aufgrund des Vergleiches einer oralen mit einer parenteralen Darreichungsform, wurde die Studie nur einfach verblindet durchgeführt (Verblindung des auswertenden Neurologen, nicht aber des behandelnden Arztes). Darüber hinaus wurde mit „Zeit bis zum Therapieversagen“ („Time to Failure“), definiert als das Auftreten eines bestätigten Schubes oder Beendigung der Studie aus jedwedem Grund, ein primärer Endpunkt gewählt, der mehr die klinische Realität abbildet. Die jährliche Schubrate wurde als wesentlicher sekundärer Endpunkt erfasst. Ähnlich wie in der TOWER-Studie setzte sich die Behandlungsdauer im Rahmen der randomisierten Phase aus Rekrutierungszeitraum plus 48 Wochen Behandlungszeitraum des letzten in die Studie eingeschlossenen Patienten zusammen. Sie betrug entsprechend zwischen minimal 48 Wochen und maximal 118 Wochen.

4.3.1.2.2 Verzerrungspotenzial auf Studienebene

Bewerten Sie das Verzerrungspotenzial der RCT auf Studienebene mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-19: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung		Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial auf Studienebene
			Patient	Bewerter			
<Studie 1>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TEM SO	ja	ja	ja	ja	nein	nein	niedrig
TOWER	ja	ja	ja	ja	nein	nein	niedrig
TENERE	ja	ja	nein	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Bei TEMSO und TOWER handelt es sich um randomisierte, doppelblinde, Placebo-kontrollierte Studien. Es erfolgte eine adäquate Erzeugung der Randomisierungssequenz und Verdeckung der Gruppenzuteilung. Da die Studien doppelblind durchgeführt worden sind, ist diesbezüglich keine Verzerrung des Studienergebnisses zu erwarten. Die Patientenpopulationen der verschiedenen Arme waren hinsichtlich wichtiger Faktoren wie Alter, Geschlecht, Krankheitsschwere (EDSS-Wert und Schubaktivität) und Vorbehandlung vergleichbar. Es liegt kein Hinweis auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotenzial erhöhende Faktoren vor. Damit ist das Verzerrungspotenzial auf Studienebene für die Studien TEMSO und TOWER als niedrig einzustufen.

Bei TENERE handelt es sich ebenfalls um eine randomisierte kontrollierte Studie. Es erfolgte eine adäquate Erzeugung der Randomisierungssequenz und Verdeckung der Gruppenzuteilung (Verblindung des auswertenden Neurologen). Die Studie war aktiv kontrolliert mit Interferon beta-1a (Rebif® 44 µg) als Komparator. Weder die Patienten noch der behandelnde Neurologe waren hinsichtlich der Studienmedikation verblindet. In der TENERE-Studie wurde Teriflunomid als Filmtablette mit einem subkutan zu applizierenden Interferon beta verglichen. Eine Verblindung wäre in diesem Falle nur durch eine Double-Dummy-Technik zu erreichen gewesen. Auch ein solches Design würde jedoch nicht zu einer

verlässlichen Verblindung führen, da die Applikation des Interferons beta-1a mit typischen Reaktionen an der Injektionsstelle verbunden ist. Eine Simulation dieser Reaktionen an der Injektionsstelle in entsprechend präparierten Placeboinjektionen ist ethisch fragwürdig. Um die Verzerrung durch die mangelnde Verblindung gering zu halten, wurden die wesentlichen Zielkriterien (Schub) durch einen bezüglich der Behandlung verblindeten untersuchenden Neurologen (in den Tabellen zur Bewertung des Verzerrungspotentials als „Bewerter“ definiert) erfasst. Die Patienten wurden angewiesen, die Injektionsstelle adäquat zu verdecken [35]. Dies entspricht den Empfehlungen des CHMP [2]. Durch die getroffenen Maßnahmen ist das Verzerrungspotential gegenüber doppelblinden Studien nicht wesentlich erhöht.

Die Patientenpopulationen der verschiedenen Arme waren auch in der TENERE-Studie hinsichtlich wichtiger Faktoren wie Alter, Geschlecht, Krankheitsschwere (EDSS-Wert und Schubaktivität) und Vorbehandlung vergleichbar. Es liegt kein Hinweis auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential beeinflussende Faktoren vor. Damit ist das Verzerrungspotential auf Studienebene für die Studie TENERE als niedrig einzustufen.

4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

Geben Sie in der folgenden Tabelle einen Überblick über die patientenrelevanten Endpunkte, auf denen Ihre Bewertung des medizinischen Nutzens und Zusatznutzens beruht. Geben Sie dabei an, welche dieser Endpunkte in den relevanten Studien jeweils untersucht wurden. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Fügen Sie für jede Studie eine neue Zeile ein.

Die Studien TEMSO, TOWER und TENERE beinhalten jeweils zwei Teriflunomid-Arme mit unterschiedlichen Dosierungen des Wirkstoffs (14 mg und 7 mg). Da Teriflunomid in Deutschland lediglich in der Dosierung von 14 mg zugelassen ist, ist der 7 mg Teriflunomid-Behandlungsarm für die Nutzenbewertung nicht relevant. Im Folgenden werden daher lediglich die Ergebnisse für den 14 mg Behandlungsarm dargestellt.

Patientenrelevante Endpunkte, die der Bewertung des Nutzens/Zusatznutzens zugrunde liegen

Tabelle 4-20: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel

Endpunkt	Morbidität				Krankheitslast	Patienten-zufriedenheit	Lebensqualität	Mortalität	Unerwünschte Ereignisse
	Schübe	Behinderungsprogression	Therapieversagen	Fatigue					
Operationalisierung	Jährliche Schubrate Zeit bis zum Schub Schubfreie Patienten	Zeit bis zur Behinderungsprogression (inklusive Anteil Patienten mit Behinderungsprogression zu einem bestimmten Zeitpunkt)	Mittlere Änderung des MSFC-Z-Scores Mittlere Änderung des EDSS-Scores	Therapieversagen („Time to Failure“) ^a	FIS ^b	„Burden of Disease“ ^c Unique active lesions	TSQM ^d SF-36 ^e EQ-5D ^f	Während der Studie verstorbene Patienten	Gesamtrate unerwünschter Ereignisse Gesamtrate schwerwiegender unerwünschter Ereignisse Gesamtrate unerwünschter Ereignisse, die zum Behandlungsabbruch führten Unerwünschte Ereignisse von besonderem Interesse
Studie									
TEM SO	ja ja ja	ja	ja ja	nein	ja	ja ja	nein	ja Ja	ja ja ja ja ja
TOWER	ja ja ja	ja	nein ja	nein	ja	nein nein	nein	ja nein	ja ja ja ja ja
TENERE	ja ja ja	ja	nein ja	ja	ja	nein nein	ja	nein nein	ja ja ja ja ja

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

^a„Therapieversagen“ ist definiert als bestätigter Schub oder permanenter Therapieabbruch aus jedwedem Grund- was immer zuerst auftrat.

^bFIS = Fatigue Impact Scale

^c„Burden of Disease“ ist hier definiert als Zunahme des Gesamtvolumens an Läsionen, d. h. Gesamtvolumen hyperintenser T2-Läsionen plus Gesamtvolumen hypointenser T1-Läsionen

^dTSQM = Treatment Satisfaction Questionnaire for Medication

^eSF-36 = Short Form 36 (Fragebogen zur Erfassung der Lebensqualität)

^fEQ-5D = EuroQoL-5D (Fragebogen zur Erfassung der Lebensqualität)

Entsprechend der Angaben in Tabelle 4-20 sind die patientenrelevanten Endpunkte Schübe, Behinderungsprogression, Therapieversagen, Fatigue, Patientenzufriedenheit, Lebensqualität, Mortalität und unerwünschte Ereignisse bei der Auswertung der Ergebnisse aus randomisierten, kontrollierten Studien berücksichtigt und jeweils in einem eigenen Kapitel beschrieben (s. Abschnitt 4.3.1.3.1 bis 4.3.1.3.11). Die verschiedenen Operationalisierungen der Schübe („Jährliche Schubrate“, „Zeit bis zum Schub“ und „Schubfreie Patienten“) sind jeweils in einem eigenen Kapitel beschrieben und in Abschnitt 4.3.1.3.1 zusammengefasst, analog sind die verschiedenen Auswertungen zur Behinderungsprogression („Zeit bis zur Behinderungsprogression“, „Mittlere Änderung des MSFC-Z-Scores“ und „Mittlere Änderung des EDSS-Scores“) dargestellt, die in Abschnitt 4.3.1.3.2 zusammengefasst sind.

Ausgewertete Patientenpopulationen

Für die Auswertung der Effektivitätseindpunkte, d. h. alle Endpunkte, die der Kategorie Morbidität, Lebensqualität oder Patientenzufriedenheit zugeordnet sind, wird die ITT-Population für die Auswertung herangezogen. Bei ausgewählten Endpunkten, wurde zusätzlich eine Auswertung der PP-Population durchgeführt. Für die Auswertung der Morbidität und der Verträglichkeitseindpunkte (unerwünschte Ereignisse) wurde die Safety-Population herangezogen.

Dabei folgt die Definition der ITT-Population nicht in allen Studien der strengen Definition dieser Population („alle randomisierten Patienten“), sondern in einigen Studien wird eine leicht modifizierte ITT-Population herangezogen, diese Populationen sind in den Studien wie folgt definiert:

Tabelle 4-21: Definitionen der Studienpopulationen in den Studienberichten von TEMSO, TOWER und TENERE

Studie	Population	Definition	Auswertung
TEMSO	ITT-Population (Intent-to-treat-Population)	Definiert als alle randomisierten Patienten, die mindestens 1 Tag der Studienmedikation ausgesetzt waren. Alle Patienten wurden in dem Behandlungsarm ausgewertet, in den sie randomisiert wurden. Streng genommen handelt es sich um eine modifizierte ITT-Population (mITT), dennoch wird sie im folgenden Dokument als ITT Population bezeichnet.	Alle Wirksamkeitseindpunkte und Lebensqualität: <ul style="list-style-type: none"> • Schübe • Behinderungsprogression • Fatigue • Burden of Disease • Lebensqualität
	PP-Population (Per-protocol-Population)	Definiert als Teilmenge der ITT-Population, die aus allen Patienten besteht, bei denen keine schwerwiegende Abweichung vom Studienprotokoll festgestellt werden konnte.	Wurde im Rahmen der Studie für zusätzliche Analysen herangezogen.
	Safety-Population	Die Safety-Population ist definiert als alle randomisierten Patienten, die der Studienmedikation ausgesetzt waren, ungeachtet der verabreichten Menge. Sollte ein Patient eine von der Randomisierung abweichende	Mortalität und Verträglichkeitseindpunkte: <ul style="list-style-type: none"> • Gesamtrate unerwünschter

		<p>Behandlung bekommen, so erfolgte die Auswertung nach der tatsächlich erfolgten Behandlung.</p> <p>Randomisierte Patienten, bei denen eine tatsächliche Einnahme der Studienmedikation nicht nachgewiesen werden konnte, wurden in die Safety-Population eingeschlossen.</p> <p>Patienten, die behandelt, aber nicht randomisiert wurden, sind nicht Teil der Safety-Population.</p> <p>Patienten, die im Laufe der Studie verschiedene Studienmedikationen erhalten haben, sind der niedrigeren erhaltenen Dosis zugeordnet.</p>	<p>Ereignisse</p> <ul style="list-style-type: none"> • Gesamtrate schwerwiegender unerwünschter Ereignisse • Unerwünschte Ereignisse, die zum Behandlungsabbruch führten • Unerwünschte Ereignisse von besonderem Interesse
TOWER	ITT-Population (Intent-to-treat-Population)	<p>Definiert als alle randomisierten Patienten, die mindestens 1 Tag der Studienmedikation ausgesetzt waren. Alle Patienten wurden in dem Behandlungsarm ausgewertet, in den sie randomisiert wurden.</p> <p>Streng genommen handelt es sich um eine modifizierte ITT-Population (mITT), im folgenden Dokument wird sie dennoch als ITT Population bezeichnet.</p>	<p>Alle Wirksamkeitsendpunkte und Lebensqualität:</p> <ul style="list-style-type: none"> • Schübe • Behinderungsprogression • Fatigue • Lebensqualität
	PP-Population (Per-protocol-Population)	<p>Definiert als Teilmenge der ITT-Population, die aus allen Patienten besteht, bei denen keine schwerwiegende Abweichung vom Studienprotokoll festgestellt werden konnte.</p>	<p>Wurde im Rahmen der Studie für zusätzliche Analysen herangezogen.</p>
	Safety-Population	<p>Die Safety-Population war eine Teilmenge aller randomisierten Patienten, die mindestens eine Dosis an Studienmedikation erhalten haben. Sollte ein Patient eine von der Randomisierung abweichende Behandlung bekommen, so erfolgte die Auswertung nach der tatsächlich erfolgten Behandlung.</p> <p>Randomisierte Patienten, bei denen eine tatsächliche Einnahme der Studienmedikation nicht nachgewiesen werden konnte, wurden in die Safety-Population eingeschlossen.</p> <p>Patienten, die behandelt, aber nicht randomisiert wurden, sind nicht Teil der Safety-Population.</p> <p>Patienten, die im Laufe der Studie verschiedene Studienmedikationen erhalten haben, sind der niedrigeren erhaltenen Dosis zugeordnet.</p>	<p>Mortalität und Verträglichkeitsendpunkte:</p> <ul style="list-style-type: none"> • Gesamtrate unerwünschter Ereignisse • Gesamtrate schwerwiegender unerwünschter Ereignisse • Unerwünschte Ereignisse, die zum Behandlungsabbruch führten • Unerwünschte Ereignisse von besonderem Interesse
TENERE	ITT-Population (Intent-to-treat-Population)	<p>Definiert als alle randomisierten Patienten. Alle Patienten wurden in dem Behandlungsarm ausgewertet, in den sie randomisiert wurden.</p>	<p>Alle Wirksamkeitsendpunkte:</p> <ul style="list-style-type: none"> • Schübe • Behinderungsprogression • Therapieversagen • Fatigue

Safety-Population	Die Safety-Population ist definiert als alle randomisierten Patienten, die der Studienmedikation ausgesetzt waren, ungeachtet der verabreichten Menge. Randomisierte Patienten, bei denen eine tatsächliche Einnahme der Studienmedikation nicht nachgewiesen werden konnte, wurden in die Safety-Population eingeschlossen.	Mortalität und Verträglichkeitsendpunkte: <ul style="list-style-type: none"> • Gesamtrate unerwünschter Ereignisse • Gesamtrate schwerwiegender unerwünschter Ereignisse • Unerwünschte Ereignisse, die zum Behandlungsabbruch führten • Unerwünschte Ereignisse von besonderem Interesse
--------------------------	---	---

Die Konsequenz aus der von der strengen Definition abweichenden Festlegung der „ITT-Population“ ist, dass in der Studie TEMSO zwei der insgesamt 1088 Patienten nicht in die Auswertung der Wirksamkeitsendpunkte einfließen. Dies wird nicht als ein das Ergebnis verzerrender Faktor gewertet.

Die Zulassung von Teriflunomid beschränkt sich auf Patienten mit RRMS. In den Studien TEMSO, TOWER und TENERE waren jedoch auch für Patienten mit anderen schubförmigen Formen der MS die Einschlusskriterien erfüllt. Patienten mit schubförmig verlaufender SPMS oder PRMS, d. h. Patienten mit progressiv-schubförmigen Verlauf, konnten in die Studien eingeschlossen werden. De facto ist der Anteil der letztgenannten Patientengruppen in den Studien sehr gering (<10 % in jeder Gruppe), so dass der Einfluss auf das Studienergebnis wahrscheinlich nicht groß, aber auch nicht gänzlich auszuschließen, ist. Deswegen werden Angaben, die in den Studienberichten von TEMSO und TOWER auch für die Subgruppe der RRMS-Patienten gemacht werden, sowohl für diese Patientengruppe als auch für die Gesamtstudienpopulation angegeben, dies betrifft vor allem die jährliche Schubrate und die Behinderungsprogression. Der Einfluss der Verlaufsform der Erkrankung auf Verträglichkeitsendpunkte wird als gering eingestuft. In der TENERE-Studie waren lediglich drei Patienten eingeschlossen, die an SPMS oder PRMS litten (alle im 14 mg –Teriflunomid-Behandlungsarm) eine gesonderte Darstellung ohne diese Patienten liegt nicht vor, wird aufgrund des geringen Anteils aber nicht als ein das Ergebnis verzerrender Faktor gewertet.

Zusammenfassend werden für die verschiedenen Endpunkte folgende Patientenpopulationen für die Bewertung herangezogen:

Tabelle 4-22: Für die Bewertung ausgewertete Patientenpopulationen

Endpunkt	Patientenpopulation
Mortalität	
Während der Studie verstorbene Patienten	Für alle Studien (TEMSO, TOWER, TENERE) wird die gesamte Studienpopulation ausgewertet. Ein Einfluss der Verlaufsform

	auf die Mortalität wird als sehr unwahrscheinlich eingestuft.
Morbidität	
Schübe	Für die Studien TEMSO und TOWER wird sowohl die gesamte Studienpopulation als auch die RRMS-Subpopulation ausgewertet. In der TENERE-Studie wird die gesamte Studienpopulation für die Auswertung herangezogen. ^a
Behinderungs- progression	Für die Studien TEMSO und TOWER wird für die „Zeit bis zur Behinderungsprogression“ sowohl die gesamte Studienpopulation als auch die RRMS-Subpopulation ausgewertet. In der TENERE-Studie wird die gesamte Studienpopulation für die Auswertung herangezogen. ^a Für die Operationalisierungen, die die Veränderung eines Scores (EDSS, MSFC) im Vergleich zur Baseline sind, wird keine differenzierte Analyse nach verschiedenen Patientengruppen durchgeführt.
Therapieversagen	Der Endpunkt ist nur in der TENERE-Studie erhoben, eine Auswertung findet für die gesamte Studienpopulation statt. ^a
Fatigue	Eine Auswertung der Subpopulation der RRMS-Patienten liegt nicht vor, für TEMSO, TOWER und TENERE wurde die gesamte Studienpopulation zur Auswertung herangezogen.
Unerwünschte Ereignisse	
AEs, SAEs, Behandlungsabbrüche, relevante spezifische AEs	Eine Analyse der unerwünschten Ereignisse nach MS-Subtyp liegt nicht vor. Aufgrund der geringen Anzahl der Patienten, die nicht der Zielpopulation entsprechen (<10 %) und der Tatsache, dass in den Studien MS-Schübe nicht als unerwünschtes Ereignis gewertet wurden, wird die Verzerrung des Ergebnis in Bezug auf die Gesamtraten als niedrig eingestuft. Für unerwünschte Ereignisse von besonderem Interesse wird zudem davon ausgegangen, dass diese Ereignisse entweder Zufall oder auf die Intervention zurückzuführen sind, nicht aber von der Verlaufsform der Erkrankung abhängen. Da keine der als „unerwünschte Ereignisse von besonderem Interesse“ prädefinierten Ereignisse typische MS-Symptome darstellen, ist diese Annahme zulässig.
Lebensqualität	
SF-36, EQ-5D	Eine Auswertung der Subpopulation der RRMS-Patienten liegt nicht vor, für TEMSO und TOWER wurde die gesamte Studienpopulation zur Auswertung herangezogen.
Patientenzufriedenheit	
TSQM	Der Endpunkt ist nur in der TENERE-Studie erhoben, eine

Auswertung findet nur für die gesamte Studienpopulation statt

^aIn der TENERE-Studie sind fast ausschließlich RRMS-Patienten in die Studie eingeschlossen worden (lediglich in einem Studienarm waren Patienten mit anderen MS-Unterformen eingeschlossen, die dort einen Anteil von < 3% haben), so dass die gesamte Studienpopulation zur Auswertung herangezogen werden kann.

4.3.1.3.1 Schübe

Bei der Verlaufsform RRMS ist das Krankheitsbild charakterisiert durch zeitlich begrenzte Phasen neurologischer Dysfunktion (Schübe). Dementsprechend sind die auftretenden MS-Schübe ein wesentlicher Morbiditäts-Endpunkt für die Bewertung eines krankheitsmodifizierenden MS-Therapeutikums. Im vorliegenden Dossier werden verschiedene Operationalisierungen der Schübe, die jährliche Schubrate (4.3.1.3.1.1), die Zeit bis zum Schub (4.3.1.3.1.2) und der Anteil schubfreier Patienten (4.3.1.3.1.3) ausgewertet, die jeweils in einem eigenen Kapitel dargestellt sind.

Übersicht der während der Studien aufgetretenen Schübe

Die MS-Schübe sind einer der zwei wesentlichen Endpunkte in Studien, die krankheitsmodifizierende MS-Wirkstoffe bewerten [1, 2]. Zur besseren Übersicht wird der Auswertung der Schübe entsprechend der oben genannten Operationalisierung eine tabellarische Darstellung der aufgetretenen Schübe vorangestellt.

Tabelle 4-23: Übersicht der während der Studien aufgetretenen Schübe – in der gesamten Studienpopulation

Studie	Patientenzahl (ITT ^a)	Schübe (gesamt)	Patientenjahre	Anzahl Patienten mit ≥ 1 Schub	Anzahl Patienten ohne Schub
TEMSO	T14 (n=358)	227	615,0	141 (39,4 %)	217 (60,6 %)
	Pb (n=363)	335	627,7	184 (50,7 %)	179 (49,3 %)
TOWER	T14 (n=370)	177	573,6	122 (33,0 %)	248 (67,0 %)
	Pb (n=388)	296	608,4	186 (47,9 %)	202 (52,1 %)
TENERE ^b	T14 (n=111)	35	132,2	26 (23,4 %)	85 (76,6 %)
	Rb44 (n=104)	25	112,1	16 (15,4 %)	88 (84,6 %)

T14 = Teriflunomid 14 mg;; Pb = Placebo; Rb44 = Rebif 44 µg

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^b99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Die Angaben in Tabelle 4-23 beziehen sich jeweils auf die gesamte Studienpopulation, eine Differenzierung nach MS-Verlaufsformen, bzw. eine gesonderte Angabe für die Patienten in der Studienpopulation, die zur Zielpopulation von 14 mg Teriflunomid gehören, folgt in Tabelle 4-24.

Tabelle 4-24: Übersicht der während der Studien aufgetretenen Schübe in der RRMS-Subpopulation

Studie	Patientenzahl (ITT ^a)	Schübe (gesamt)	Patientenjahre	Anzahl Patienten mit ≥ 1 Schub	Anzahl Patienten ohne Schub
TEMSO	T14 (n=332)	205	576,5	130 (39,2 %)	202 (60,8%)
	Pb (n=329)	305	574,9	169 (51,4 %)	160 (48,6 %)
TOWER	T14 (n=364)	169	564,2	120 (33,0 %)	244 (67,0 %)
	Pb (n=378)	286	596,4	179 (47,4 %)	199 (52,6 %)

T14 = Teriflunomid 14 mg;; Pb = Placebo

4.3.1.3.1 Jährliche Schubrate – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-25: Operationalisierung des Endpunktes „Jährliche Schubrate“

Studie	Operationalisierung
Schübe	
TEMSO	Jährliche Schubrate wird definiert als <ul style="list-style-type: none"> Anzahl bestätigter Schübe pro Patientenjahr („jährliche Schubrate“; ARR = annualized relapse rate) Die Definition eines Schubes ist unten angegeben.
TOWER	Jährliche Schubrate wird definiert als <ul style="list-style-type: none"> Anzahl bestätigter Schübe pro Patientenjahr („jährliche Schubrate“; ARR = annualized relapse rate) Die Definition eines Schubes ist unten angegeben.
TENERE	Jährliche Schubrate wird definiert als <ul style="list-style-type: none"> Anzahl bestätigter Schübe pro Patientenjahr („jährliche Schubrate“; ARR = annualized relapse rate) Die Definition eines Schubes ist unten angegeben.
<p>Ein Schub war in allen Studien wie folgt definiert:</p> <p>Auftreten eines neuen klinischen Zeichens/Symptoms oder die klinische Verschlechterung eines vorhandenen Symptoms (eines, das mindestens 30 Tage lang stabil war), das mindestens 24 h anhielt und nicht von Fieber begleitet war. Die Schübe wurden durch den behandelnden Neurologen, basierend auf einer objektiven Bewertung (EDSS/FS scores) eines unabhängigen Bewerter (der untersuchende Neurologe), bestätigt, wenn eines der folgenden Kriterien erfüllt war:</p> <ul style="list-style-type: none"> Erhöhung um 1 Punkt in mindestens 2 FS-Funktionen oder eine Erhöhung um 2 Punkte in mindestens einer FS-Funktion (ausgenommen Darm/Blase und Gehirn) Eine Erhöhung um mindestens 0,5 Punkte im EDSS score (es sei denn, der EDSS war vorher 0, dann war eine Erhöhung um 1,0 Punkte gefordert) im Vergleich zu der vorherigen stabilen Einstufung 	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-26: Bewertung des Verzerrungspotenzials für „Jährliche Schubrate“ in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
TEMPO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
TENERE	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wurde für alle drei Studien TEMPO, TOWER und TENERE auf Studienebene als niedrig eingestuft. Es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren. Der Endpunkterheber, d. h. der auswertende Neurologe war in allen Studien (auch in der nicht doppelblind durchgeführten TENERE-Studie) verblindet, so dass dies nicht als ein das Verzerrungspotential erhöhender Faktor für den Endpunkt „Jährliche Schubrate“ gewertet werden kann.

Stellen Sie die Ergebnisse für den Endpunkt „Jährliche Schubrate“ für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Tabelle 4-27: Ergebnisse für „Jährliche Schubrate“ aus RCT mit dem zu bewertenden Arzneimittel - für die gesamte Studienpopulation

Studie	Patienten- zahl (ITT ^e)	Schübe (gesamt)	Patienten- jahre	Nicht-adjustierte jährliche Schubrate ^a	Adjustierte jährliche Schubrate ^b 95 % KI []	Relative Risk ^d 95 % KI [] p-Wert ^f	Relative risk reduction vs. Placebo bzw. Rebif [®] 44 µg
TEMSO	T14 (n=358)	227	615,0	0,369	0,369 [0,308; 0,441]	0,685 [0,554; 0,847] p = 0,0005	31,5 %
	Pb (n=363)	335	627,7	0,534	0,539 [0,466; 0,623]		
TOWER	T14 (n=370)	177	573,6	0,309	0,319 [0,267; 0,381]	0,637 [0,512, 0,793] p < 0,0001	36,3 %
	Pb (n=388)	296	608,4	0,487	0,501 [0,432; 0,581]		
TENERE ^g	T14 (n=111)	35	132,2	0,265	0,259 [0,153; 0,438]	1,197 [0,623; 2,299] p = 0,5896 ^c	-19,7 %
	Rb44 (n=104)	25	112,1	0,223	0,216 [0,113; 0,415]		

T14 = Teriflunomid 14 mg; Pb = Placebo; Rb44 = Rebif[®] 44 µg^aGesamtzahl der während der Behandlung aufgetretenen Schübe geteilt durch die Anzahl an Patientenjahren in der Studie^bErhalten aus einem Poisson-Modell mit der Gesamtzahl bestätigter Schübe zwischen Randomisierung und letzter Dosierung als abhängiger Variable, Behandlung, Baseline EDSS-Wert ($\leq 3,5$ / $> 3,5$) und Region als unabhängigen Variablen und log-transformierte Behandlungsdauer als Offset-Variable^cChi-Quadrat-Test aus der Schätzung der Rate Ratios^drelatives Risiko im Vergleich zu Placebo (TEMSO, TOWER), bzw. Rebif 44 µg (TENERE)^es. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

^fH₀: RR = 1

^g99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Tabelle 4-28: Ergebnisse für jährliche Schubrate aus TEMSO und TOWER - für die Subgruppe der RRMS-Patienten

Studie	Patientenzahl (ITT ^d)	Schübe (gesamt)	Patienten-jahre	Nicht-adjustierte jährliche Schubrate ^a	Adjustierte jährliche Schubrate ^b 95 % KI []	Relative Risk ^c 95 % KI [] p-Wert ^e	Relative risk reduction vs. Placebo bzw. Rebif [®] 44 µg
TEMSO	T14 (n=332)	205	576,5	0,356	0,355 [0,294; 0,429]	0,661 [0,531; 0,824] p = 0,0002	33,9 %
	Pb (n=329)	305	574,9	0,531	0,537 [0,462; 0,623]		
TOWER	T14 (n=364)	169	564,2	0,300	0,310 [0,261; 0,369]	0,628 [0,506; 0,779] p < 0,0001	37,2 %
	Pb (n=378)	286	596,4	0,480	0,495 [0,425; 0,576]		

T14 = Teriflunomid 14 mg; Pb = Placebo; Rb44 = Rebif[®] 44 µg

^aGesamtzahl der während der Behandlung aufgetretenen Schübe geteilt durch die Anzahl an Patientenjahren in der Studie

^bErhalten aus einem Poisson-Modell mit der Gesamtzahl bestätigter Schübe zwischen Randomisierung und letzter Dosierung als abhängiger Variable, Behandlung, Baseline EDSS-Wert ($\leq 3,5$ / $> 3,5$) und Region als unabhängigen Variablen und log-transformierte Behandlungsdauer als Offset-Variable

^crelatives Risiko im Vergleich zu Placebo (TEMSO, TOWER)

^ds. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^eH₀: RR = 1

TEMESO

Die jährliche Schubrate war der primäre Endpunkt in der TEMSO-Studie. In der TEMSO-Studie betrug die adjustierte jährliche Schubrate in der Placebo-Gruppe 0,539 (95 % KI: [0,466; 0,623]) und in der 14 mg Teriflunomid-Gruppe 0,369 (95 % KI: [0,308; 0,441]). Diese Ergebnisse entsprachen einer Reduktion des relativen Risikos im Vergleich zu Placebo im 14 mg Teriflunomid-Behandlungsarm von 31,5 % (95 % KI: [15,3 %; 44,6 %]). Das Ergebnis ist mit einem p-Wert von 0,0005 statistisch signifikant von Null verschieden.

Für die Zielpopulation RRMS-Patienten ergab sich mit einer relativen Risikoreduktion von 33,9 % (95 % KI: [17,6 %; 46,9 %]) ein vergleichbares Ergebnis, auch hier war das Ergebnis statistisch signifikant.

Der Vergleich mit der nicht adjustierten jährlichen Schubrate und der individuellen jährlichen Schubrate führte zu einem vergleichbaren Ergebnis hinsichtlich der Effektivität für Teriflunomid. In einer sekundären Analyse wurden die MS-Schübe auch für die PP-Population (per protocol) ausgewertet, es ergab sich kein wesentlicher Unterschied im Vergleich zu der Auswertung der ITT-Population [32]. Die Ergebnisse waren robust und veränderten sich nur marginal. Des Weiteren wurde eine Sensitivitätsanalyse mit zusätzlichen Daten, die während des Follow-up Zeitraums erhoben worden sind [32], und Subgruppenanalysen (s. Modul 4.3.1.3.11.2) durchgeführt. Auch durch diese Auswertungen ergab sich kein Hinweis auf ein verzerrtes Ergebnis.

TOWER

Die jährliche Schubrate war der primäre Endpunkt in der TOWER-Studie. Die adjustierte jährliche Schubrate betrug in der Placebo-Gruppe 0,501 (95 % KI: [0,432; 0,581]) und in der 14 mg Teriflunomid-Gruppe 0,319 (95 % KI: [0,267; 0,381]). Diese Ergebnisse entsprachen einer statistisch signifikanten (p-Wert kleiner als 0,0001) Reduktion des relativen Risikos im Vergleich zu Placebo in dem 14 mg Teriflunomid-Behandlungsarm von 36,3 % (95 % KI: [20,7 %; 48,8 %]). Sensitivitätsanalysen [31] und Subgruppenanalysen (s. Modul 4.3.1.3.11.3) unterstützten die Ergebnisse der Primäranalyse.

Auch in der TOWER-Studie war die Reduktion des relativen Risikos für die RRMS-Subpopulation mit einem Wert von 37,2 % [22,1 %; 49,4 %] mit dem Ergebnis der Gesamtstudienpopulation vergleichbar, auch hier ist das Ergebnis statistisch signifikant.

TENERE

Die jährliche Schubrate war der wichtigste sekundäre Endpunkt in der TENERE-Studie. Die Gesamtzahl bestätigter Schübe betrug in der TENERE-Studie 35 Schübe in der 14 mg Teriflunomid-Gruppe und 25 Schübe in der Rebif[®]-Gruppe. Dies entspricht adjustierten jährlichen Schubraten von 0,259 (95% KI: [0,153; 0,438]) in der 14 mg Teriflunomid-Gruppe und 0,216 (95% KI: [0,113; 0,415]) in der Rebif[®]-Gruppe.

Diese Ergebnisse entsprachen einer Erhöhung des relativen Risikos um 19,7% (95 % KI: [-37,7 %; 129,9 %]) in der 14 mg Teriflunomid-Gruppe im Vergleich mit der Rebif[®]-Gruppe, wobei der Unterschied zwischen der 14 mg Teriflunomid-Gruppe und der Rebif[®]-Gruppe statistisch nicht signifikant war ($p = 0,5896$).

Jährliche Schubrate - Zusammenfassung

In den Studien TEMSO und TOWER wurde das Studienziel erreicht und es konnte eine signifikante Reduktion hinsichtlich des primären Endpunkts (jährliche Schubrate) im Vergleich zu Placebo gezeigt werden. Die Schubrate wurde durch die Behandlung mit Teriflunomid im Vergleich mit Placebo um **31,5 %** in der TEMSO-Studie und um **36,3 %** in der TOWER-Studie reduziert.

Bei Betrachtung der Zielpopulation von 14 mg Teriflunomid (erwachsene Patienten mit RRMS) kann eine vergleichbare Reduktion des relativen Risikos um **33,9 %** bzw. **37,2 %** erreicht werden. Tendenziell ergibt sich für die Subgruppe der RRMS-Patienten bezogen auf diesen Zielparameter sogar ein leicht besseres Ergebnis als für die gesamte Studienpopulation, da aber die Subgruppe der Patienten mit SPMS und PRMS sehr klein ist, kann eine solche Aussage basierend auf den vorliegenden Daten nicht getroffen werden.

In Bezug auf diesen Effektivitätsparameter konnte in der direkt vergleichenden Studie TENERE (Teriflunomid versus Rebif[®] 44 µg) kein signifikanter Unterschied der beiden Behandlungen nachgewiesen werden.

4.3.1.3.1.2 Zeit bis zum Schub und Patienten mit bestätigtem Schub - RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-29: Operationalisierung des Endpunktes „Zeit bis zum Schub“

Studie	Operationalisierung
Schübe	
TEMESO	Die Zeit bis zum Schub wird operationalisiert als Wahrscheinlichkeit einen Schub zu erleiden, bestimmt durch eine Ereigniszeitanalyse. Die Definition eines Schubes ist unten angegeben.
TOWER	Die Zeit bis zum Schub wird operationalisiert als Wahrscheinlichkeit einen Schub zu erleiden, bestimmt durch eine Ereigniszeitanalyse. Die Definition eines Schubes ist unten angegeben.
TENERE	Die Zeit bis zum Schub wird operationalisiert als Wahrscheinlichkeit einen Schub zu erleiden, bestimmt durch eine Ereigniszeitanalyse. Die Definition eines Schubes ist unten angegeben.
<p>Ein Schub war in allen Studien wie folgt definiert: Auftreten eines neuen klinischen Zeichens/Symptoms oder die klinische Verschlechterung eines vorhandenen Symptoms (eines, das mindestens 30 Tage lang stabil war), das mindestens 24 h anhielt und nicht von Fieber begleitet war. Die Schübe wurden durch den behandelnden Neurologen, basierend auf einer objektiven Bewertung (EDSS/FS scores) eines unabhängigen Bewerter (der untersuchende Neurologe), bestätigt, wenn eines der folgenden Kriterien erfüllt war:</p> <ul style="list-style-type: none"> • Erhöhung um 1 Punkt in mindestens 2 FS-Funktionen oder eine Erhöhung um 2 Punkte in mindestens einer FS-Funktion (ausgenommen Darm/Blase und Gehirn) • Eine Erhöhung um mindestens 0,5 Punkte im EDSS score (es sei denn, der EDSS war vorher 0, dann war eine Erhöhung um 1,0 Punkte gefordert) im Vergleich zu der vorherigen stabilen Einstufung 	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-30: Bewertung des Verzerrungspotenzials für „Zeit bis zum Schub“ in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
TEMZO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
TENERE	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wurde für alle drei Studien TEMZO, TOWER und TENERE auf Studienebene als niedrig eingestuft. Es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren. Der Endpunkterheber, d. h. der auswertende Neurologe war in allen Studien (auch in der nicht doppelblind durchgeführten TENERE-Studie) verblindet, so dass dies nicht als ein das Verzerrungspotential erhöhender Faktor für den Endpunkt „Zeit bis zum Schub“ gewertet werden kann.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Stellen Sie die Ergebnisse für den Endpunkt Zeit bis zum Schub für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-31: Ergebnisse für „Zeit bis zum Schub“ aus RCT mit dem zu bewertenden Arzneimittel - für die gesamte Studienpopulation

Studie	Anzahl Patienten (ITT ^b)	Anzahl Patienten mit ≥1 Schub	Relatives Risiko 95 % KI p-Wert ^c	Odds Ratio 95 % KI p-Wert ^c	Wahrscheinlichkeit für Patienten ohne bestätigten Schub (W48) 95 % KI []	Wahrscheinlichkeit für Patienten ohne bestätigten Schub (W108) 95 % KI []	Zeit bis zum Schub Hazard Ratio ^a 95 % KI [] p-Wert ^c
TEM SO	T14 (n=358)	141 (39,4%)	0,777 [0,660; 0,915] p = 0,0025	0,632 [0,470; 0,849] p = 0,0023	0,677 [0,626; 0,727]	0,565 [0,510; 0,620]	0,719 [0,577; 0,895] p = 0,0030
	Pb (n=363)	184 (50,7%)			0,601 [0,549; 0,653]	0,456 [0,402; 0,510]	
TOWER	T14 (n=370)	122 (33,0 %)	0,688 [0,575; 0,822] p < 0,0001	0,534 [0,398; 0,717] p < 0,0001	0,763 [0,717; 0,810]	0,571 [0,505; 0,637]	0,631 [0,502; 0,794] p < 0,0001
	Pb (n=388)	186 (47,9 %)			0,606 [0,555; 0,656]	0,468 [0,410; 0,526]	
Wahrscheinlichkeit für ein Therapieversagen aufgrund eines Schubes (s. Teilauswertungen des kombinierten Endpunktes „Time to Failure“)							
Studie	Anzahl Patienten (ITT)	Anzahl Patienten mit ≥1 Schub	Relatives Risiko 95 % KI p-Wert ^c	Odds Ratio 95 % KI p-Wert ^c	Woche 48		Hazard Ratio 95 % KI [] p-Wert ^c
TENERE ^e	T14 (n=111)	26 (23,4%) ^d	1,523 [0,868; 2,672] p = 0,1429	1,682 [0,843; 3,356] p = 0,1398	0,213 [0,133; 0,292]		1,463 [0,784; 2,728] p = 0,2292
	Rb (n=104)	16 (15,4%) ^d			0,164 [0,087; 0,240]		

^aerhalten aus einem COX proportional Hazards- Modell mit Behandlung, Baseline-EDSS-Score ($\leq 3,5 / > 3,5$) und Region als unabhängige Variablen.

^bs. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^cEs gelten die folgenden Nullhypothesen: $H_0: RR=1$, $H_0: OR=1$ und $H_0: HR=1$

^dAusgewertet sind die Patienten mit einem bestätigten Schub, der Anteil Patienten mit Schub (egal ob bestätigt oder nicht) liegt mit 29 (26,1%) in der Teriflunomid- und 19 (18,3%) in der Rebif-Gruppe etwas höher

^e99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Abbildung 4-4: Kaplan-Meier-Kurven für Zeit bis zum ersten Schub in TEMSO (links) und TOWER (rechts)

Figure 7 - Kaplan-Meier plot of time to first multiple sclerosis relapse - intent-to-treat population

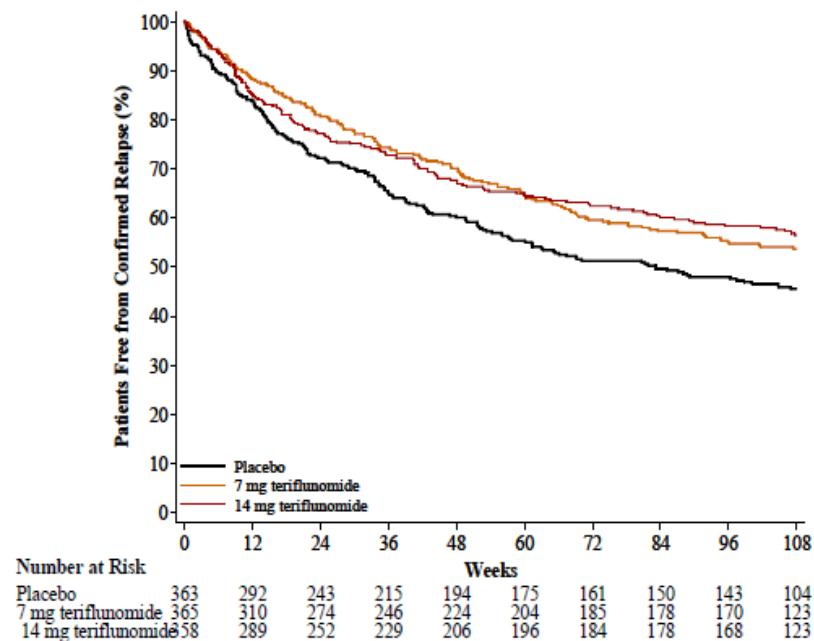


Figure 11 - Kaplan-Meier plot of time to first MS relapse - ITT population

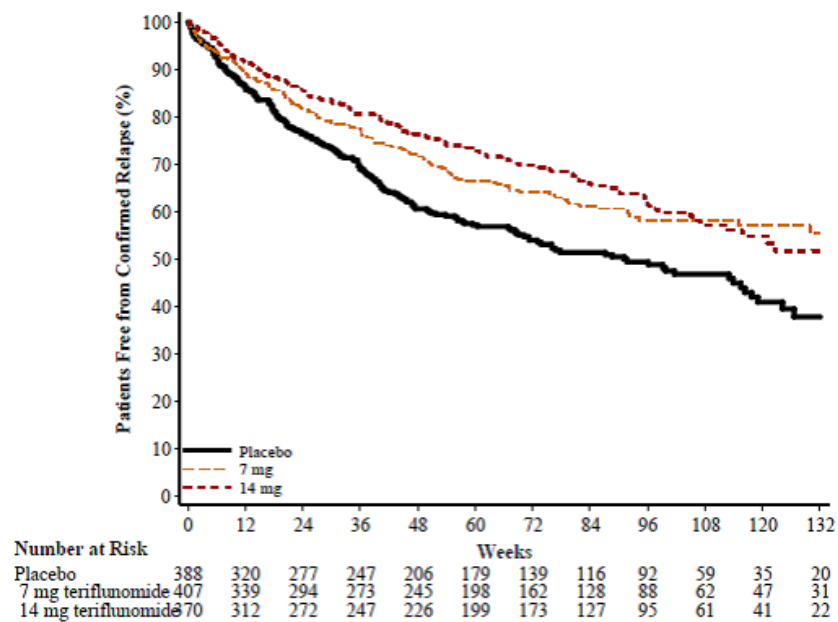


Tabelle 4-32: Ergebnisse für „Patienten mit Schub“ aus TEMSO und TOWER - für die RRMS-Subpopulation

Studie	Anzahl Patienten (ITT ^a)	Anzahl Patienten mit ≥ 1 Schub	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
TEMSO	T14 (n=332)	130 (39,2 %)	0,762 [0,643; 0,904] p = 0,0018	0,609 [0,447; 0,830] p = 0,0017
	Pb (n=329)	169 (51,4 %)		
TOWER	T14 (n=364)	120 (33,0 %)	0,696 [0,581; 0,834] p < 0,0001	0,547 [0,406; 0,736] p < 0,0001
	Pb (n=378)	179 (47,4 %)		

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bEs gelten die folgenden Nullhypothesen: $H_0: RR=1$, $H_0: OR=1$ und $H_0: HR=1$

TEMSSO

In der Ereigniszeitanalyse „Zeit bis zum Schub“ wurde das Risiko, einen Schub zu erleiden, bestimmt. Das entsprechende Hazard ist für Patienten, die mit 14 mg Teriflunomid behandelt wurden, um **28,1 %** geringer als für Patienten, die mit Placebo behandelt wurden (95 % KI: [10,5 %; 42,3 %]).

Der Anteil der Patienten, die während der Dauer der Studie einen Schub erlitten haben, war in der 14 mg Teriflunomid-Gruppe signifikant niedriger als in der Placebo-Gruppe (relatives Risiko 0,762; 95 % KI: [0,643; 0,904]).

TOWER

Das Hazard für die Zeit bis zum Schub ist für Patienten, die mit 14 mg Teriflunomid behandelt wurden, um **36,9 %** geringer als für Patienten, die mit Placebo behandelt wurden (95 % KI: [20,6 %; 49,8 %]).

Der Anteil der Patienten, die während der Dauer der Studie einen Schub erlitten haben, war in der 14 mg Teriflunomid-Gruppe signifikant niedriger als in der Placebo-Gruppe (relatives Risiko 0,696; 95 % KI: [0,581; 0,834]).

TENERE

Für den Endpunkt „Zeit bis zum ersten Schub“ konnte kein Vorteil für Teriflunomid 14 mg im Vergleich zu Rebif[®] 44 µg gezeigt werden.

4.3.1.3.1.3 Schubfreie Patienten - RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-33: Operationalisierung des Endpunktes „Schubfreie Patienten“

Studie	Operationalisierung
TEMSO	Definiert als der Anteil an Patienten, die im Verlauf der Studie keinen Schub erlitten haben. Die Definition eines Schubes ist unten angegeben.
TOWER	Definiert als der Anteil an Patienten, die im Verlauf der Studie keinen Schub erlitten haben. Die Definition eines Schubes ist unten angegeben.
TENERE	Definiert als der Anteil an Patienten, die im Verlauf der Studie keinen Schub erlitten haben. Die Definition eines Schubes ist unten angegeben.
<p>Ein Schub war in allen Studien wie folgt definiert:</p> <p>Auftreten eines neuen klinischen Zeichens/Symptoms oder die klinische Verschlechterung eines vorhandenen Symptoms (eines, das mindestens 30 Tage lang stabil war), das mindestens 24 h anhielt und nicht von Fieber begleitet war. Die Schübe wurden durch den behandelnden Neurologen, basierend auf einer objektiven Bewertung (EDSS/FS scores) eines unabhängigen Bewerter (der untersuchende Neurologe), bestätigt, wenn eines der folgenden Kriterien erfüllt war:</p> <ul style="list-style-type: none"> • Erhöhung um 1 Punkt in mindestens 2 FS-Funktionen oder eine Erhöhung um 2 Punkte in mindestens einer FS-Funktion (ausgenommen Darm/Blase und Gehirn) • Eine Erhöhung um mindestens 0,5 Punkte im EDSS score (es sei denn, der EDSS war vorher 0, dann war eine Erhöhung um 1,0 Punkte gefordert) im Vergleich zu der vorherigen stabilen Einstufung 	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-34: Bewertung des Verzerrungspotenzials für „Schubfreie Patienten“ in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
TEMSO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
TENERE	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wurde für alle drei Studien TEMSO, TOWER und TENERE auf Studienebene als niedrig eingestuft. Es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren. Der Endpunkterheber, d. h. der auswertende Neurologe war in allen Studien (auch in der nicht doppelblind durchgeführten TENERE-Studie) verblindet, so dass dies nicht als ein das Verzerrungspotential erhöhender Faktor für den Endpunkt „Schubfreie Patienten“ gewertet werden kann.

Stellen Sie die Ergebnisse für den Endpunkt „Schubfreie Patienten“ für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-35: „Schubfreie Patienten“ aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

Studie	Anzahl Patienten (ITT ^a)	Anzahl Patienten, die während der Studiendauer keinen Schub erlitten haben Prozentualer Anteil	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
TEMSO	T14 (n=358)	217 (60,6 %)	1,229 [1,075; 1,405] p = 0,0025	1,582 [1,177; 2,126] p = 0,0023
	Pb (n=363)	179 (49,3 %)		
TOWER	T14 (n=370)	248 (67,0 %)	1,287 [1,143; 1,451] p < 0,0001	1,872 [1,394; 2,512] p < 0,0001
	Pb (n=388)	202 (52,1 %)		
TENERE ^c	T14 (n=111)	85 (76,6 %)	0,905 [0,793; 1,032] p = 0,1369	0,594 [0,298; 1,186] p = 0,1398
	Rb (n=104)	88 (84,6 %)		

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bH₀: RR=1 bzw. H₀: OR=1

^c99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Tabelle 4-36: „Schubfreie Patienten“ aus TEMSO und TOWER - für die RRMS-Subpopulation

Studie	Anzahl Patienten (ITT ^a)	Anzahl Patienten, die während der Studiendauer keinen Schub erlitten haben Prozentualer Anteil	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
TEMSO	T14 (n=332)	202 (60,8%)	1,251 [1,087; 1,440] p = 0,0018	1,641 [1,205; 2,235] p = 0,0017
	Pb (n=329)	160 (48,6 %)		
TOWER	T14 (n=364)	244 (67,0 %)	1,273 [1,130; 1,435] p < 0,0001	1,829 [1,358; 2,463] p < 0,0001
	Pb (n=378)	199 (52,6 %)		

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bH₀: RR=1 bzw. H₀: OR=1

TEMSO

Der Anteil an schubfreien Patienten ist in der 14 mg Teriflunomid-Gruppe signifikant höher als in der Placebo-Gruppe (relatives Risiko von 1,229; 95% KI: [1,075; 1,405]).

Der Anteil an schubfreien Patienten ist in der RRMS-Subpopulation mit einem relativen Risiko von 1,251 und einem zugehörigen Konfidenzintervall von 95 % KI: [1,087; 1,440] mit dem Ergebnis der gesamten Studienpopulation vergleichbar.

TOWER

Der Anteil an schubfreien Patienten ist in der 14 mg Teriflunomid-Gruppe signifikant höher als in der Placebo-Gruppe (relatives Risiko von 1,287; 95% KI: [1,143; 1,451]).

Der Anteil an schubfreien Patienten ist in der RRMS-Subpopulation mit einem relativen Risiko von 1,273 und einem zugehörigen Konfidenzintervall von 95 % KI: [1,130; 1,435] mit dem Ergebnis der gesamten Studienpopulation vergleichbar.

TENERE

Für den Endpunkt „Anteil schubfreier Patienten“ konnte kein Vorteil für Teriflunomid 14 mg im Vergleich zu Rebif[®] 44 µg gezeigt werden.

4.3.1.3.1.4 Zusätzliche Auswertungen bezüglich der Schübe, die für die TEMSO-Studie (und teilweise für die TOWER-Studie) durchgeführt worden sind

Bei klinischer Notwendigkeit wurden die während der Studie aufgetretenen Schübe mit Kortikosteroiden behandelt. Die bevorzugte standardisierte Behandlung bestand aus der intravenösen Verabreichung von Natrium-Methylprednisolon-Succinat 1g/Tag für 3 bis 5 Tage. Der Großteil der bestätigten Schübe wurde auch tatsächlich mit Kortikosteroiden behandelt (s. Tabelle 4-37). Insgesamt waren die meisten Schübe von einem mittleren Schweregrad, nur wenige Schübe waren in ihrer Ausprägung schwer (s. Tabelle).

Tabelle 4-37: Schwere der Schübe und andere Schub-Effekte – Teil 1 Auswertungen der TEMSO Studie

Studie	Patientenzahl (ITT ^a)	Gesamtzahl, Patienten mit bestätigten Schüben	Patienten mit milden Schüben (prozentualer Anteil)	Patienten mit moderaten Schüben (prozentualer Anteil)	Patienten mit schweren Schüben (prozentualer Anteil)
TEMSO	T14 (n=358)	141	43 (30,5%)	87 (61,7%)	11 (7,8%)
	Pb (n=363)	184	50 (27,2%)	125 (67,9%)	9 (4,9%)

T14 = Teriflunomid 14 mg; Pb = Placebo

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

Tabelle 4-38: Schwere der Schübe und andere Schub-Effekte – Teil 2 Auswertungen der TEMSO Studie

Studie		Gesamtzahl bestätigter Schübe	Milde Schübe (prozentualer Anteil)	Moderate Schübe (prozentualer Anteil)	Schwere Schübe (prozentualer Anteil)
TEMSO	T14	227	90 (39,6%)	124 (54,6%)	13 (5,7%)
	Pb	335	117 (34,9%)	208 (62,1%)	10 (3,0%)

T14 = Teriflunomid 14 mg; Pb = Placebo

Tabelle 4-39: Schwere der Schübe und andere Schub-Effekte – Auswertungen der TEMSO - Studie- Teil 3 plus Auswertungen der TOWER-Studie^a

Studie	Patientenzahl (ITT ^c)	Schübe mit Spätfolgen		Hospitalisierungspflichtige Schübe		Kortikosteroid-pflichtige Schübe		Notfall-Einweisungen ins Krankenhaus (TEMSO), bzw. Hospitalisierungen aus irgendeinem Grund (TOWER)	
		Adjustierte jährliche Schubrate ^b	Relative Risiko- reduktion gegenüber Placebo	Adjustierte jährliche Schubrate ^b	Relative Risiko- reduktion gegenüber Placebo	Adjustierte jährliche Schubrate ^b	Relative Risiko- reduktion gegenüber Placebo	Adjustierte jährliche Schubrate ^b	Relative Risiko- reduktion gegenüber Placebo
TEMSO	T14 (n=358)	0,080	53 % p < 0,0001	0,057	59 % p < 0,0001	0,284	34 % p = 0,0003	0,251	42 % p = 0,0041
	Pb (n=363)	0,168		0,139		0,428		0,436	
TOWER	T14 (n=370)	0,081	53,5 % p = 0,0004	0,100	33,6 % p = 0,0155	0,274	35,7 % p = 0,0002	0,210	27,8% p = 0,0299
	Pb (n=388)	0,173		0,151		0,426		0,290	

^aDie Auswertung ist Bestandteil des Abschnitts 2.7.3 des kanadischen Zulassungsdokuments [36]

^bDie adjustierten jährlichen Raten leiten sich aus einem Poisson-Modell ab, welches die Gesamtzahl der Ereignisse von Interesse als jeweils abhängige Variable enthielt und Behandlung, Baseline-EDSS-Grades ($\leq 3,5 / > 3,5$), sowie Region als unabhängige Variablen und die log-transformierte standardisierte Studiendauer als Offset-Variable.

^cs. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

Hinweis: Patienten mit einem Ereignis, das als in seiner Intensität verschlechtert, anhaltend oder unbekannt berichtet war, (n=6) wurden der Kategorie „Schübe mit Spätfolgen“ zugerechnet.

TEM SO

Das Risiko, Schübe, die sich nur unvollständig zurückbilden, zu erleiden, war für Patienten unter Teriflunomid 14 mg-Behandlung gegenüber Patienten unter Placebo-Behandlung um 53 % signifikant reduziert.

Das Risiko, einen hospitalisierungspflichtigen Schub zu erleiden, war für Patienten unter Teriflunomid 14 mg-Behandlung gegenüber Patienten unter Placebo-Behandlung um 59 % signifikant reduziert.

Das Risiko, einen Kortikosteroid-pflichtigen Schub zu erleiden, war für Patienten unter Teriflunomid-Behandlung gegenüber Patienten unter Placebo-Behandlung um 34 % signifikant reduziert.

Das Risiko, einer Notfalleinweisung ins Krankenhaus, war für Patienten unter Teriflunomid 14 mg-Behandlung gegenüber Patienten unter Placebo-Behandlung um 42 % signifikant reduziert.

Diese Ergebnisse sind in Tabelle 4-39 zusammengestellt.

TOWER

Das Risiko, Schübe mit Spätfolgen zu erleiden, war für Patienten unter Teriflunomid 14 mg-Behandlung gegenüber Patienten unter Placebo-Behandlung um 53,5 % signifikant reduziert.

Das Risiko, einen hospitalisierungspflichtigen Schub zu erleiden, war für Patienten unter Teriflunomid 14 mg-Behandlung gegenüber Patienten unter Placebo-Behandlung um 33,6 % signifikant reduziert.

Das Risiko, einen Kortikosteroid-pflichtigen Schub zu erleiden, war für Patienten unter Teriflunomid-Behandlung) gegenüber Patienten unter Placebo-Behandlung um 35,7 % signifikant reduziert.

Das Risiko, einer Notfalleinweisung ins Krankenhaus, war für Patienten unter Teriflunomid 14 mg-Behandlung gegenüber Patienten unter Placebo-Behandlung um 27,8 % signifikant reduziert.

Diese Ergebnisse sind in Tabelle 4-39 zusammengestellt.

4.3.1.3.1.5 Metaanalysen der schubbezogenen Endpunkte

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

TEMZO und TOWER lassen sich bezüglich des Endpunktes Schübe metaanalytisch zusammenfassen. Die Studie TENERE (anderer Komparator) lässt sich nicht im Sinne einer Metaanalyse mit TEMZO und TOWER zusammenfassen. Die Metaanalysen der jährlichen Schubrate, der Patienten mit ≥ 1 Schub und der schubfreien Patienten sind allesamt für die Zielpopulation von 14 mg Teriflunomid, d. h. für RRMS Patienten durchgeführt worden, Metaanalysen der Ergebnisse dieser Operationalisierungen des Endpunktes Schübe bezogen auf die gesamte Studienpopulation ergaben kein in seiner Aussage abweichendes Ergebnis und sind im Rahmen des vorliegenden Dossiers nicht dargestellt.

Jährliche Schubrate

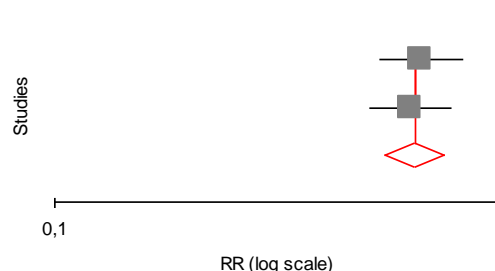
Die Metaanalyse der Ergebnisse aus TEMZO und TOWER für RRMS-Patienten ergibt für die jährliche Schubrate ein homogenes Bild ($I^2=0\%$, $p_Q=0,7444$). Die relative Risikoreduktion, die mit 14 mg Teriflunomid erreicht werden kann, beträgt **35,6 %**.

Tabelle 4-40: Ergebnis der Metaanalyse von TEMZO und TOWER für die jährliche Schubrate – für die RRMS-Subpopulation

Analyse	Komparatoren	Rate Ratio 95 % KI [] p-Wert ^a	Relative Risiko- Reduktion
Metaanalyse TEMZO & TOWER	Teriflunomid 14 mg versus Placebo	0,644 [0,552; 0,751] $p < 0,0001$ $I^2=0\%$, $p_Q=0,7444$	35,6 %

Abbildung 4-5: Forest-Plot der Metaanalyse von TEMZO und TOWER für die jährliche Schubrate

Study ID	Year	Exposed AM[e]/SE[e]	Control AM[c]/SE[c]
TEMZO (RRMS-Pat.)	2011	-0,414/0,1121	-0,414/0,1121
TOWER (RRMS-Pat.)	2012	-0,4652/0,1101	-0,4652/0,1101
META-ANALYSIS:			

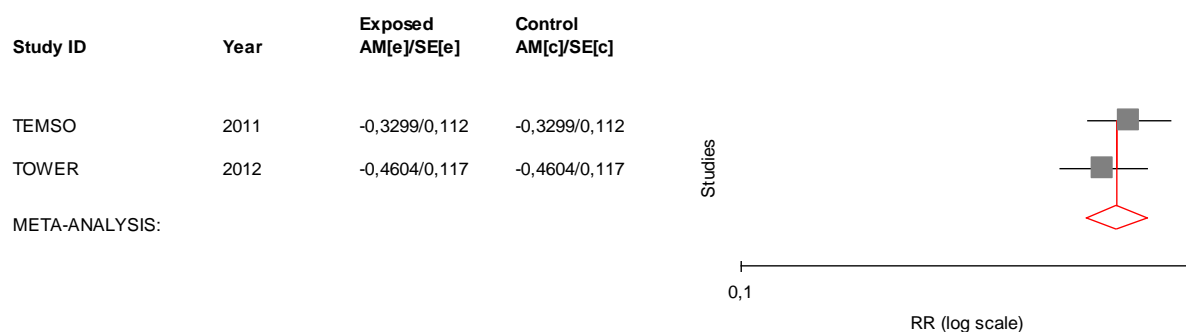


Zeit bis zum Schub

Tabelle 4-41: Ergebnis der Metaanalyse von TEMSO und TOWER für die Zeit bis zum Schub – für die gesamte Studienpopulation

Analyse	Komparatoren	HR 95 % KI [] p-Wert ^a
Metaanalyse TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	0,676 [0,576; 0,792] p < 0,0001 I ² =0%, p _Q =0,4201

Abbildung 4-6: Forest-Plot der Metaanalyse von TEMSO und TOWER für die Zeit bis zum Schub



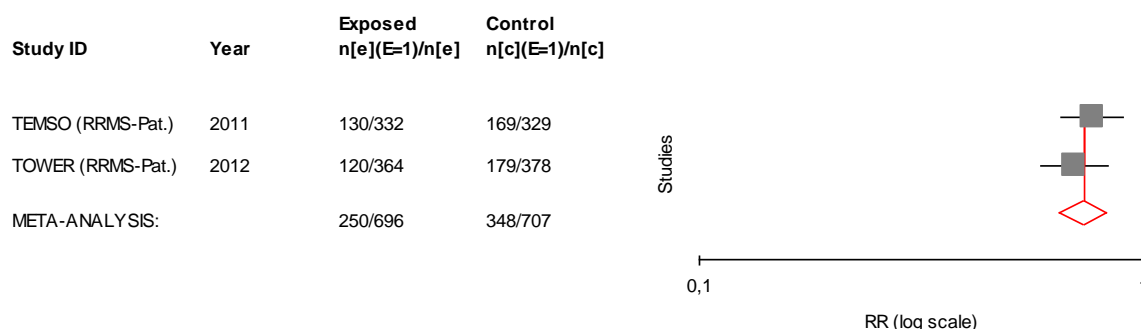
Patienten mit Schub und Schubfreie Patienten

Auch für die beiden Operationalisierungen „Patienten mit Schub“ und „schubfreie Patienten“ ergibt sich in den Metaanalysen ein homogenes Bild. Das Risiko einen Schub zu erleiden, ist unter Teriflunomid-Therapie signifikant niedriger als unter Placebo. Analog dazu ist die Chance, schubfrei zu bleiben, signifikant höher.

Tabelle 4-42: Ergebnis der Metaanalyse von TEMSO und TOWER für „Patienten mit Schub“ – für die RRMS-Subpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	0,731 [0,645; 0,827] p < 0,0001 I ² =0%, p _Q =0,4745	0,576 [0,465; 0,714] p < 0,0001 I ² =0%, p _Q =0,6205

Abbildung 4-7: Forest-Plot der Metaanalyse (basierend auf den relativen Risiken) von TEMSO und TOWER für die „Patienten mit Schub“

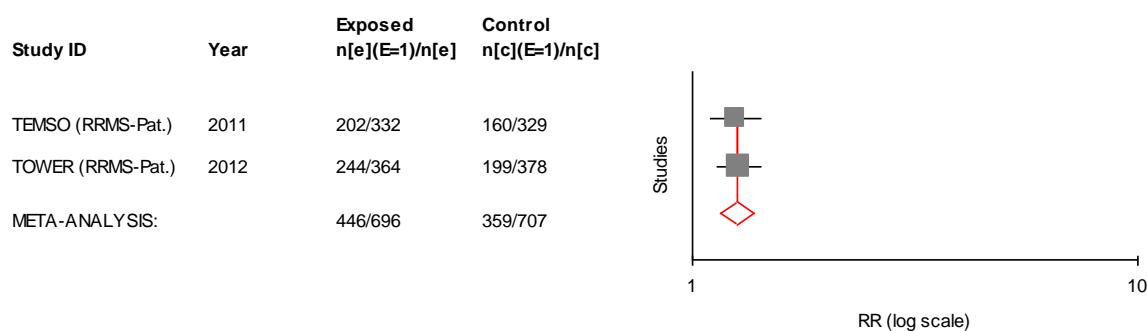


Schubfreie Patienten

Tabelle 4-43: Ergebnis der Metaanalyse von TEMSO und TOWER für „schubfreie Patienten“ – für die RRMS-Subpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	1,264 [1,154; 1,385] p < 0,0001 I ² =0%, p _Q =0,8519	1,736 [1,401; 2,151] p < 0,0001 I ² =0%, p _Q =0,6205

Abbildung 4-8: Forest-Plot der Metaanalyse (basierend auf den relativen Risiken) von TEMSO und TOWER für die „schubfreie Patienten“



4.3.1.3.1.6 Gesamtergebnis zu schubbezogenen Endpunkten

Die Wirksamkeit von 14 mg Teriflunomid gegenüber Placebo bezogen auf die Reduktion von MS-Schüben konnte sehr konsistent gezeigt werden. Ohne Ausnahme zeigt sich in allen verschiedenen Auswertungen und in den Metaanalysen von TEMSO und TOWER für schubbezogene Endpunkte eine klare, statistisch signifikante Überlegenheit von Teriflunomid gegenüber Placebo. Aus den Metaanalysen ergibt sich kein Hinweis auf Heterogenität.

Für die jährliche Schubrate, Patienten mit Schub und schubfreie Patienten konnte gezeigt werden, dass das Ergebnis der gesamten Studienpopulation auf die Zielpopulation der erwachsenen Patienten mit RRMS übertragbar ist.

Damit zeigt sich der Nutzen der Intervention im Indikationsgebiet A, Behandlung erwachsener Patienten mit schubförmig-remittierender Multipler Sklerose.

Im direkten Vergleich von 14 mg Teriflunomid mit Interferon beta-1a 44 µg, dreimal wöchentlich, subkutan (Rebif®) konnte kein Unterschied zwischen den beiden Interventionen gezeigt und damit auch kein Zusatznutzen abgeleitet werden (Ergebnisse aus TENERE).

4.3.1.3.2 Behinderungsprogression

Die Behinderungsprogression ist neben den Schüben der zweite wesentliche Morbiditäts-Endpunkt im vorliegenden Anwendungsgebiet A (Behandlung erwachsener Patienten mit schubförmig-remittierender Multipler Sklerose (RRMS)). Im Rahmen des vorliegenden Dossiers sind drei Operationalisierungen der Behinderungsprogression ausgewertet, die jeweils in einem eigenen Kapitel dargestellt sind:

- Zeit bis zur Behinderungsprogression (s. 4.3.1.3.2.1)
- Mittlere Änderung des MSFC-Z-Scores (s. 4.3.1.3.2.2)
- Mittlere Änderung des EDSS-Scores (4.3.1.3.2.3)

4.3.1.3.2.1 Zeit bis zur Behinderungsprogression (inklusive Anteil Patienten mit Behinderungsprogression zu einem bestimmten Zeitpunkt) - RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-44: Operationalisierung von“ Zeit bis zur Behinderungsprogression“ (inklusive Anteil Patienten mit Behinderungsprogression zu einem bestimmten Zeitpunkt)

Studie	Operationalisierung
TEMSO	<p>Zeit bis zur 12 Wochen anhaltenden Behinderungsprogression (12week SDP = sustained disability progression):</p> <p>Definiert als die Zeit bis es zu einer 1-Punkt Erhöhung des EDSS-Wertes im Vergleich zum Baseline-Wert gekommen ist (in den Fällen mit einem Baseline-Wert $\leq 5,5$) oder zu einer Erhöhung um mindestens 0,5 Punkte (in den Fällen mit einem Baseline-Wert $> 5,5$). Diese Erhöhung des EDSS-Scores musste für mindestens 12 Wochen bestehen bleiben (12 Wochen anhaltende Behinderungsprogression).</p> <p>Die Zeit-bis-zum-Ereignis ist definiert als die Zeit in Tagen, die von dem Tag der Randomisierung bis zum Zeitpunkt der ersten Behinderungsprogression vergeht.</p> <hr/> <p>Anteil Patienten mit Behinderungsprogression zu einem bestimmten Zeitpunkt:</p> <p>Definition der Behinderungsprogression s. o., zusätzlich zur Zeit-bis-zum-Ereignis-Variable wird auch die entsprechende Rate angegeben.</p>
TOWER	<p>Zeit bis zur 12 Wochen anhaltenden Behinderungsprogression (12week SDP = sustained disability progression):</p> <p>Definiert als die Zeit bis es zu einer 1-Punkt Erhöhung des EDSS-Scores im Vergleich zum Baseline-Wert gekommen ist (in den Fällen, bei denen der Baseline-Wert $\leq 5,5$ betrug) oder zu einer Erhöhung um mindestens 0,5 Punkte (in den Fällen, bei denen der Baseline-Wert über 5,5 lag). Diese Erhöhung des EDSS-Grades musste für mindestens 12 Wochen bestehen bleiben (12 Wochen anhaltende Behinderungsprogression).</p> <p>Die Zeit-bis-zum-Ereignis ist definiert als die Zeit in Tagen, die von dem Tag der Randomisierung bis zum Zeitpunkt der ersten Behinderungsprogression vergeht.</p> <p>In dem Fall, dass eine Progression der Erkrankung festgestellt werden konnte, wurde der Patient dazu aufgefordert, erneut der Studienteilnahme zuzustimmen.</p> <p>Bei jedem Arztbesuch wurde untersucht, ob der Patient eine anhaltende Krankheitsprogression erlitten hat.</p> <ul style="list-style-type: none"> • Eine erste anhaltende Krankheitsprogression (oder Behinderungsprogression) mit oder ohne Schub wurde definiert als eine persistierende Erhöhung von mindestens 1,0 Punkten auf der EDSS Skala bezogen auf den Baseline-Wert (bzw. 0,5 für Patienten, deren Basline EDSS-Wert über 5,5 lag), die für mindestens 12 Wochen stabil war. • Eine folgende anhaltende Behinderungsprogression war definiert als eine persistierende Erhöhung von mindestens 1,0 Punkten auf der EDSS Skala bezogen auf den Wert des letzten Arztbesuches (bzw. 0,5 für Patienten, deren letzter EDSS-Wert über 5,0 lag), die für mindestens 12 Wochen stabil war <p>Die Kombination von EDSS- und FS (FS = Funktionssystem)-Werten sind einer der ältesten und wahrscheinlich gebräuchlichsten Werkzeuge zur Erfassung des Krankheitsstatus in der MS.</p>

	<p>Basierend auf einer neurologischen Untersuchung werden die 7 Funktionssysteme (plus „andere“) eingestuft. Diese Einstufungen werden in Verbindung mit Beobachtungen und Informationen über den Gang und den Gebrauch von Gehilfen verwendet, um den EDSS-Grad festzulegen. Jeder der einzelnen FS-Werte ist eine klinische Ordinalskala im Wertebereich von 0 bis 5 oder 6. Der EDSS ist eine klinische Ordinalskala im Wertebereich von 0 (normale neurologische Funktion) bis 10 (Tod durch MS) in Stufen von 0,5.</p> <p>Der FS-Wert und der EDSS-Grad wurden durch den untersuchten Neurologen persönlich durchgeführt. Die Durchführungsdauer variierte abhängig vom Zustand des Patienten und der Fähigkeit des Untersuchers. Der gleiche Untersucher musste die Bewertung so gut wie möglich bei jedem Arztbesuch durchführen.</p> <p>Bestätigung der Behinderung:</p> <p>In den Fällen, wo ein Patient eine erste Behinderungsprogression während der Behandlungszeit erfahren hat und zum Zeitpunkt des Arztbesuchs, bei dem das Auswaschverfahren gestartet wurde, noch keine 12-Wochen-Bestätigung vorlag, wurde der Patient für eine EDSS-Bewertung in die Klinik bestellt. Wenn dann eine Behinderungsprogression bestätigt werden konnte, wurde der Patient für eine letzte Bewertung des EDSS-Grades zum Zeitpunkt 24 Wochen einbestellt.</p> <p>Im Fall einer vorzeitigen Beendigung der Studie war die Effektivitätsbewertung entsprechend am Ende des Behandlungszeitraumes durchzuführen.</p>
TENERE	<p>Die Zeit bis zur 12 Wochen anhaltenden Behinderungsprogression (inklusive Anteil Patienten mit Behinderungsprogression zu einem bestimmten Zeitpunkt) wurde in der TENERE Studie nicht als Endpunkt prädefiniert, sondern in einer Posthoc-Analyse analog zu den Studien TEMSO und TOWER ausgewertet.</p>

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-45: Bewertung des Verzerrungspotenzials für „Zeit bis zur 12 Wochen anhaltenden Behinderungsprogression“ in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunktherheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TEMSO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
TENERE	niedrig	unklar	ja	nein	nein	hoch

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wird für alle drei Studien TEMSO, TOWER und TENERE auf Studienebene als niedrig eingestuft. Das ITT-Prinzip ist adäquat umgesetzt und es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren.

Der Endpunkterheber war in TEMSO und TOWER verblindet und das Verzerrungspotential für den Endpunkt „Zeit bis zur Behinderungsprogression“ wird als niedrig eingestuft.

Bei der Auswertung zur TENERE-Studie muss darauf hingewiesen werden, dass die Behinderungsprogression kein vorab definierter Endpunkt war und in einer Post-hoc-Analyse ausgewertet worden ist [58]. Eine analog zu der Auswertung der Schübe vorgeschriebene Vorgehensweise der Erhebung, die zusätzlich zu dem nicht verblindeten behandelnden Neurologen einen auswertenden Neurologen für eine verblindete Auswertung einschaltet, findet sich für die regelmäßige Erhebung des EDSS nicht, so dass nicht mit Sicherheit davon ausgegangen werden kann, dass der Endpunkt „Zeit bis zur Behinderungsprogression“ in der TENERE-Studie verblindet erhoben worden ist. Aus diesem genannten Grund wird das Verzerrungspotential für den Endpunkt „Zeit bis zur Behinderungsprogression“ in der TENERE-Studie als hoch eingestuft.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Stellen Sie die Ergebnisse für den Endpunkt Zeit bis zur Behinderungsprogression für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

„Zeit bis zur Behinderungsprogression“ inklusive „Anteil Patienten mit Behinderungsprogression zu einem bestimmten Zeitpunkt“

Tabelle 4-46: Ergebnisse für „Zeit bis zur Behinderungsprogression“ aus RCT mit dem zu bewertenden Arzneimittel - für die gesamte Studienpopulation

Studie	Patientenzahl (ITT ^a)	Anzahl Patienten mit Behinderungsprogression (Anteil)	Anzahl zensierter Beobachtungen (Anteil)	Wahrscheinlichkeit der Behinderungsprogression 24 Wochen ^b 95 % KI []	Wahrscheinlichkeit der Behinderungsprogression 48 Wochen ^b 95 % KI []	Wahrscheinlichkeit der Behinderungsprogression 108 Wochen ^b 95 % KI []	Wahrscheinlichkeit der Behinderungsprogression 132 Wochen ^b 95 % KI []	Hazard Ratio ^c (im Vergleich zu Placebo bzw. Rebif) 95 % KI [] p-Wert ^d Reduktion des Risikos
TEMESO	T14 (n = 358)	62 (17,3 %)	296 (82,7 %)	0,062 [0,036; 0,088]	0,113 [0,079; 0,148]	0,202 [0,156; 0,247]		0,702 [0,506; 0,973] p = 0,0279 29,8 %
	Pb (n = 363)	86 (23,7 %)	277 (76,3%)	0,086 [0,057; 0,116]	0,160 [0,121; 0,200]	0,273 [0,233; 0,323]		
TOWER	T14 (n = 370)	44 (11,9%)	326 (88,1%)	0,027 [0,009; 0,044]	0,078 [0,049; 0,108]	0,158 [0,112; 0,204]	0,158 [0,112; 0,204]	0,685 [0,467; 1,004] p = 0,0442 31,5 %
	Pb (n=388)	65 (16,8%)	323 (83,2%)	0,080 [0,052; 0,107]	0,142 [0,106; 0,179]	0,197 [0,152; 0,241]	0,210 [0,159; 0,260]	
TENERE ^c	T14	10	101	0,048	0,068			0,982

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

	(n=111)	(9,0 %)	(91,0 %)	[0,007; 0,089]	[0,019; 0,117]	[0,399; 2,420]
						p = 0,9690
						1,8 %
Rb	9	95	0,063	0,099		
	(n=104)	(8,7 %)	(91,3 %)	[0,014; 0,111]	[0,037; 0,162]	

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^babgeleitet aus Kaplan-Meier-Schätzern

^cabgeleitet aus einem Cox-proportional-Hazards- Modell mit Behandlung, Baseline-EDSS-Grad ($\leq 3,5 / > 3,5$) und Region als abhängige Variablen

^dabgeleitet aus einem Log-Rank Test mit Stratifizierung nach Baseline-EDSS-Grad ($\leq 3,5 / > 3,5$) und Region

^e99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Abbildung 4-9: Kaplan-Meier-Kurven für „Zeit bis zur Behinderungsprogression“ in TEMSO (links) und TOWER (rechts)

Figure 2 - Kaplan-Meier plot of time to disability progression sustained for 12 weeks - intent-to-treat population

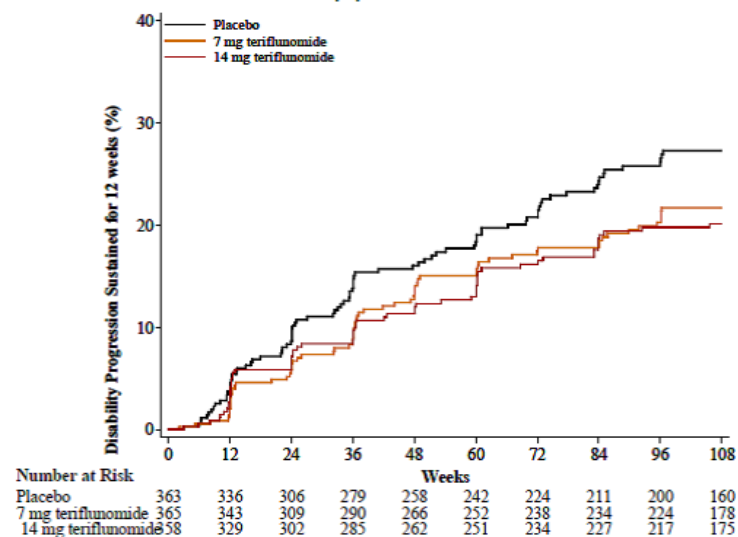


Figure 7 - Kaplan-Meier plot of time to disability progression sustained for 12 weeks - ITT population

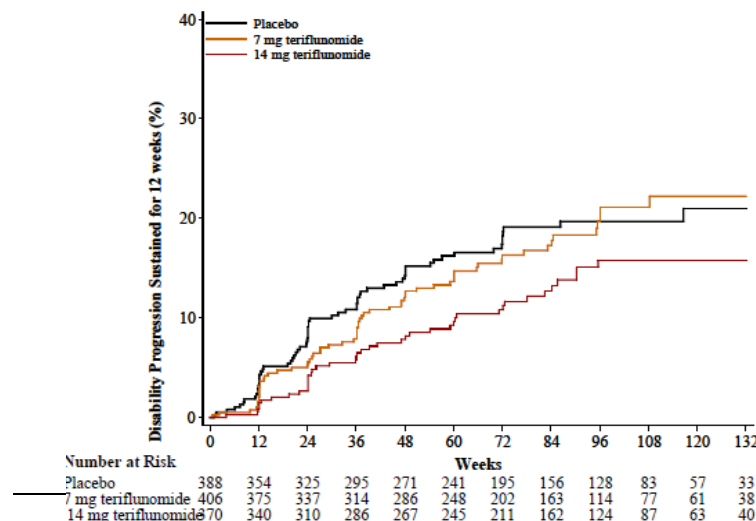


Tabelle 4-47: Ergebnisse für „Zeit bis zur Behinderungsprogression“ aus TEMSO und TOWER - für die Subpopulation RRMS-Patienten

Studie	Patientenzahl (ITT ^a)	Anzahl Patienten mit Behinderungsprogression (Anteil)	Anzahl zensierter Beobachtungen (Anteil)	Wahrscheinlichkeit der Behinderungsprogression 48 Wochen ^b 95 % KI []	Wahrscheinlichkeit der Behinderungsprogression 108 Wochen ^b 95 % KI []	Hazard Ratio ^c (im Vergleich zu Placebo) 95 % KI []
TEMSO	T14 (n=332)	59 (17,8 %)	273 (82,2 %)	0,118 [0,082; 0,155]	0,204 [0,157; 0,251]	0,721 [0,514; 1,011] p = 0,0580 ^d
	Pb (n=329)	78 (23,7 %)	251 (76,3 %)	0,166 [0,124; 0,208]	0,269 [0,218; 0,321]	
TOWER	T14 (n=364)	42 (11,5 %)	322 (88,5 %)	0,073 [0,044; 0,102]	0,154 [0,108; 0,200]	0,663 [0,449; 0,979] p = 0,0388 ^d
	Pb (n=378)	64 (16,9 %)	314 (83,1 %)	0,143 [0,106; 0,180]	0,198 [0,153; 0,243]	

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^b Kaplan-Meier-Schätzer

^c abgeleitet aus einem Cox-proportional-Hazards-Modell mit Behandlung, Baseline-EDSS-Grad ($\leq 3,5 / > 3,5$) und Region als Kovariate

^dH₀: HR = 1

Tabelle 4-48: Ergebnisse für „Patienten mit Behinderungsprogression“ – für die gesamte Studienpopulation

Studie	Patientenzahl (ITT ^a)	Anzahl Patienten mit Behinderungsprogression (Anteil)	Anzahl zensierter Beobachtungen (Anteil)	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
TEMSO	T14 (n=358)	62 (17,3 %)	296 (82,7 %)	0,731 [0,546; 0,979] p = 0,0355	0,675 [0,468; 0,972] p = 0,0347
	Pb (n=363)	86 (23,7 %)	277 (76,3 %)		
TOWER	T14 (n=370)	44 (11,9 %)	326 (88,1 %)	0,710 [0,498; 1,013] p = 0,0586	0,671 [0,444; 1,013] p = 0,0577
	Pb (n=388)	65 (16,8 %)	323 (83,2 %)		
TENERE ^c	T14 (n=111)	10 (9,0 %)	101 (91,0 %)	1,041 [0,441; 2,460] p = 0,9270	1,045 [0,407; 2,684] p = 0,9269
	Rb (n = 104)	9 (8,7 %)	95 (91,3 %)		

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bH₀: RR=1 bzw. H₀: OR=1

^c99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Tabelle 4-49: Ergebnisse für „Patienten mit Behinderungsprogression“ aus TEMSO und TOWER - für die RRMS-Subpopulation

Studie	Patienten-zahl (ITT ^a)	Anzahl Patienten mit Behinderungsprogression (Anteil)	Anzahl zensierter Beobachtungen (Anteil)	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
TEMSO	T14 (n=332)	59 (17,8 %)	273 (82,2 %)	0,750 [0,554; 1,014] p = 0,0613	0,695 [0,476; 1,016] p = 0,0604
	Pb (n=329)	78 (23,7 %)	251 (76,3 %)		
TOWER	T14 (n=364)	42 (11,5 %)	322 (88,5 %)	0,681 [0,475; 0,978] p = 0,0377	0,640 [0,421; 0,973] p = 0,0368
	Pb (n=378)	64 (16,9 %)	314 (83,1 %)		

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bH₀: RR=1 bzw. H₀: OR=1

TEMZO

Die Ereigniszeitanalyse in Bezug auf die Behinderungsprogression zeigte einen deutlichen Vorteil für Teriflunomid. Der nach der Kaplan-Meier-Methode geschätzte prozentuale Anteil von Patienten mit 12 Wochen anhaltender Behinderungsprogression nach 108 Wochen betrug in der Studie TEMZO in der Placebo-Gruppe und der 14-mg-Teriflunomid-Gruppe jeweils 27,3 % und 20,2 %. Entsprechende Berechnungen der Hazard Ratios ergaben, dass die 14-mg-Dosierung das Risiko einer 12 Wochen anhaltenden Behinderungsprogression um **29,8 %** signifikant senkte (95 % KI: [2,7 %; 49,4 %]). Der Anteil Patienten mit Behinderungsprogression konnte durch die Behandlung mit 14 mg Teriflunomid signifikant gesenkt werden (relatives Risiko von 0,731; 95 % KI: [0,546; 0,979]).

Für die Subpopulation der RRMS Patienten wurde ein Hazard Ratio von 0,721 (95 % KI:[0,514; 1,011]) ermittelt, was einer Senkung des Risikos um **27,9 %** entspricht. Auch in der Subpopulation der RRMS-Patienten war der Anteil an Patienten mit Behinderungsprogression in der 14 mg Teriflunomid-Gruppe niedriger als in der Placebo-Gruppe, das Ergebnis war mit einem relativen Risiko von 0,750 und einem zugehörigen Konfidenzintervall von 95 % KI: [0,554; 1,014]) nicht signifikant.

TOWER

Die Ereigniszeitanalyse in Bezug auf die Behinderungsprogression zeigte einen deutlichen Vorteil für Teriflunomid. Der nach der Kaplan-Meier-Methode geschätzte Anteil an Patienten mit 12 Wochen anhaltender Behinderungsprogression (12-week SDP) nach 48 Wochen betrug 14,2 % in der Placebo- und 7,8 % in der 14 mg Teriflunomid-Gruppe. Entsprechende Berechnungen der Hazard Ratios ergaben, dass die Teriflunomid-Behandlung das Risiko einer 12 Wochen anhaltenden Behinderungsprogression um **31,5 %** reduzierte (95 % KI:[-0,4 %; 53,3 %]). Der Anteil Patienten mit Behinderungsprogression konnte durch die Behandlung mit 14 mg Teriflunomid gesenkt werden (relatives Risiko von 0,710; 95 % KI: [0,498; 1,013]).

Für die Subpopulation der RRMS Patienten wurde ein Hazard Ratio von 0,663 (95 % KI: [0,449; 0,979]) ermittelt, was einer Senkung des Risikos um **33,7 %** entspricht. Der Anteil an Patienten mit Behinderungsprogression war in der 14 mg Teriflunomid Gruppe signifikant niedriger als in der Placebo-Gruppe (relatives Risiko von 0,681; 95 % KI:[0,475; 0,978]).

TENERE

In dem direkten Vergleich war die Analyse der Zeit bis zur Behinderungsprogression nicht vorgesehen. Der Endpunkt wurde post-hoc ausgewertet [59], es konnte kein signifikanter Unterschied zwischen 14 mg Teriflunomid und Rebif[®] 44 µg gezeigt werden.

Zeit bis zur Behinderungsprogression - Fazit

Die Darstellung in der Tabelle zeigt, dass die Wahrscheinlichkeit einer Behinderungsprogression für Patienten unter Placebo zu jedem Zeitpunkt höher ist als für die Patienten in

der 14 mg Teriflunomid-Gruppe. Die Effektschätzer und damit das Ausmaß mit dem die Behinderungsprogression durch die Therapie herausgezögert werden kann, sind in den Studien TEMSO und TOWER sowohl in der Gesamtpopulation als auch in der Teilpopulation der RRMS-Patienten vergleichbar. In der Subpopulation der RRMS Patienten konnte in der TEMSO-Studie zwar knapp kein statistisch signifikantes Ergebnis erreicht werden, dies kann jedoch auf die niedrigere Patientenzahl zurückzuführen sein. Durch die sehr konsistenten Ergebnisse von TEMSO und TOWER hatte auch die Meta-Analyse eine gute Homogenität (s 4.3.1.3.2.4) und eine statistisch signifikante Reduktion der Behinderungsprogression in klinisch relevantem Ausmaß. Damit ist der Nutzen von Teriflunomid im Anwendungsgebiet A ist auch für den zweiten wichtigen klinischen Endpunkt gezeigt worden.

4.3.1.3.2.2 Mittlere Änderung des MSFC-Z-Scores - RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-50: Operationalisierung von Mittlere Änderung des MSFC-Scores

Studie	Operationalisierung
TEMSO	Mittlere Änderung des MSFC-Scores:

Der MSFC (Multiple Sclerosis Functional Composite) ist ein 3-teiliges, standardisiertes, quantitatives MS Erfassungsinstrument zur Verwendung in klinischen Studien. Der MSFC umfasst die Komponenten Funktion der Beine/Gehfähigkeit (timed 25-foot walk) , Funktion der Arme und der Hände (Steckbrett-Test (9-Loch-Test)) und kognitive Funktion (PASAT = paced auditory serial addition test; Version 3).

Erhebungszeitpunkte: Baseline, Woche 24, 48, 72 und 96

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-51: Bewertung des Verzerrungspotenzials für „Mittlere Änderung des MSFC-Z-Scores“ in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TEMSO	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wird für die Studie TEMSO auf Studienebene als niedrig eingestuft. Das ITT-Prinzip ist adäquat umgesetzt und es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren.

Der Endpunkterheber war in TEMSO verblindet und das Verzerrungspotential für den Endpunkt Behinderungsprogression wird als niedrig eingestuft.

Stellen Sie die Ergebnisse für den Endpunkt „Mittlere Änderung des MSFC-Z-Scores“ für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-52: Ergebnisse für Mittlere Änderung des MSFC Z-Scores aus RCT mit dem zu bewertenden Arzneimittel - für die gesamte Studienpopulation

Studie		Change from Baseline MMRM ^b				
Patienten- zahl (ITT ^a)	Baseline- Wert Mittelwert (Standard- abweichung)	Anzahl ^b (96 Wochen)	Wert in W 96 Mittelwert (Standard- abweichung)	W 96 LS-Mean ^c (Standard- fehler)	LS-Mean- differenz im Vergleich zu Placebo (Standard- fehler) 95 % KI [] p-Wert	
Z-Score gesamt						
TEMSO	T14 (n= 358)	0,036 (0,658)	265	0,002 (0,734)	-0,082 (0,039)	0,048 (0,053) [-0,056; 0,152] p = 0,3646
	Pb (n = 363)	-0,032 (0,767)	260	-0,048 (0,906)	-0,130 (0,039)	
Leg function/ambulation Z score (Gehfähigkeit)						
TEMSO	T14 (n= 358)	0,056 (0,800)	264	-0,097 (1,444)	-0,202 (0,102)	0,139 (0,141) [-0,139; 0,417] p = 0,3253
	Pb (n = 363)	-0,037 (1,092)	259	-0,153 (1,676)	-0,341 (0,102)	
Arm/hand function Z score (Koordination der oberen Extremitäten)						
TEMSO	T14 (n= 358)	0,019 (0,983)	265	-0,082 (0,948)	-0,143 (0,038)	-0,043 (0,051) [-0,143; 0,056] p = 0,3933
	Pb (n = 363)	0,006 (0,973)	260	-0,011 (0,966)	-0,100 (0,037)	
Cognitive function Z score (Kognitive Funktion)						
TEMSO	T14 (n= 358)	0,044 (0,980)	264 (262 in MMRM)	0,182 (0,862)	0,073 (0,035)	0,095 (0,047) [0,003; 0,187] p = 0,0435
	Baseline- Werte für 356 Patienten ^d					
	Pb (n = 363)	-0,062 (1,007)	258 (257 in MMRM)	0,026 (1,080)	-0,022 (0,035)	
	Baseline- Werte for 362 Patienten ^d					

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bMMRM = Mixed-effects Model with Repeated Measures, realisiert in SAS mit Proc Mixed. Feste Effekte in dem Modell waren Behandlung, Baseline EDSS-Stratum ($\leq 3,5$ / $> 3,5$), Region, Visite, Behandlung \times Visite-

Wechselwirkung, Baseline-Wert und Behandlung \times Visite-Wechselwirkung. Für die Modellierung der within-patient Fehlerterme wurde eine nicht-strukturierte Korrelationsmatrix zu Grunde gelegt. Die Modellparameter wurden nach der restricted Maximum-Likelihood-Methode unter Verwendung des Newton-Raphson-Algorithmus geschätzt. Die Freiheitsgrade des Nenners wurden dabei nach der Approximation von Satterthwaite geschätzt.

„Anzahl“ steht in diesem Zusammenhang für die Anzahl real vorhandener Datensätze nach 96 Wochen.

^cLS-Means sind die sich aus dem MMRM-Modell ergebenden adjustierten Schätzer.

^dAngabe über die Anzahl der Patienten, für die ein Baseline-Wert vorhanden ist, ist nur angegeben, wenn die entsprechende Anzahl von der Anzahl der Patienten in der ITT-Population abweicht

TEM SO

In Woche 96 konnte im Vergleich zur Placebo-Gruppe kein statistisch signifikanter Unterschied im MSFC-Z-Score festgestellt werden. Lediglich in dem die kognitive Funktion betreffenden Subscore konnte ein signifikanter Unterschied zwischen den Behandlungsarmen festgestellt werden.

4.3.1.3.2.3 Mittlere Änderung des EDSS-Scores - RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-53: Operationalisierung von Mittlere Änderung des EDSS-Scores

Studie	Operationalisierung
TEMSO, TOWER und TENERE	<p>Die Kombination von EDSS- und FS (FS = Funktionssystem)-Werten sind einer der ältesten und wahrscheinlich gebräuchlichsten Werkzeuge zur Erfassung des Krankheitsstatus in der MS. Basierend auf einer neurologischen Untersuchung werden die 7 Funktionssysteme (plus „andere“) eingestuft. Diese Einstufungen werden in Verbindung mit Beobachtungen und Informationen über den Gang und den Gebrauch von Gehilfen verwendet, um den EDSS-Grad festzulegen. Jeder der einzelnen FS-Werte ist eine klinische Ordinalskala im Wertebereich von 0 bis 5 oder 6. Der EDSS ist eine klinische Ordinalskala im Wertebereich von 0 (normale neurologische Funktion) bis 10 (Tod durch MS) in Stufen von 0,5.</p> <p>Der FS-Wert und der EDSS-Grad wurden durch den untersuchten Neurologen persönlich durchgeführt. Die Durchführungsdauer variierte abhängig vom Zustand des Patienten und der Fähigkeit des Untersuchers. Der gleiche Untersucher musste die Bewertung so gut wie möglich bei jedem Arztbesuch durchführen.</p> <p>Die kombinierten EDSS- und FS-Werte basieren zusätzlich zu der Ambulation auf einer neurologischen Standarduntersuchung der 7 Funktionssysteme (visuell, Hirnstamm, pyramidal, zerebellar, sensorisch, Darm/Blase und cerebral) eingestuft mit einem FS-Wert zwischen 0</p>

	(normal) und 5 oder 6. Diese Einstufungen werden in Verbindung der Beobachtung des Gangs und der Verwendung unterstützender Hilfsmittel verwendet, um zur Bestimmung eines EDSS-Grades zwischen 0 (normaler neurologischer Befund) und 10 (Tod aufgrund von MS) zu gelangen.
TEM SO	Mittlere Änderung des EDSS-Scores Erhebungszeitpunkte: Baseline und während der Studie alle 12 Wochen und im Falle eines Schubes.
TOWER	Mittlere Änderung des EDSS-Scores Erhebungszeitpunkte: Baseline, Woche 12, 24, dann Woche 30 und weiter alle 12 Wochen und im Falle eines Schubes.
TENERE	Mittlere Änderung des EDSS-Scores Erhebungszeitpunkte: Baseline, Woche 12, 24, dann Woche 36 und weiter alle 12 Wochen und im Falle eines Schubes.

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-54: Bewertung des Verzerrungspotenzials für „Mittlere Änderung des EDSS-Scores“ in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TEM SO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
TENERE	niedrig	unklar	ja	nein	nein	hoch

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wird für alle drei Studien TEMSO, TOWER und TENERE auf Studienebene als niedrig eingestuft. Das ITT-Prinzip ist adäquat umgesetzt und es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren.

Der Endpunkterheber war in TEMSO und TOWER verblindet und das Verzerrungspotential für den Endpunkt Behinderungsprogression wird als niedrig eingestuft.

Bei der Auswertung zur TENERE-Studie muss darauf hingewiesen werden, dass die Behinderungsprogression kein vorab definierter Endpunkt war und in einer Posthoc-Analyse ausgewertet worden ist [58]. Eine analog zu der Auswertung der Schübe vorgeschriebene Vorgehensweise der Erhebung, die zusätzlich zu dem nicht verblindeten behandelnden Neurologen einen auswertenden Neurologen für eine verblindete Auswertung einschaltet, findet sich für die regelmäßige Erhebung des EDSS nicht, so dass nicht mit Sicherheit davon ausgegangen werden kann, dass der Endpunkt „Mittlere Änderung des EDSS-Wertes“ in der TENERE-Studie verblindet erhoben worden ist. Aus diesem genannten Grund wird das Verzerrungspotential für den Endpunkt „Mittlere Änderung des EDSS-Wertes“ in der TENERE-Studie als hoch eingestuft.

Stellen Sie die Ergebnisse für den Endpunkt „Mittlere Änderung des EDSS-Scores“ für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-55: Ergebnisse für „Mittlere Änderung des EDSS-Scores“ aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

Studie	Patientenzahl (ITT ^a)	Baseline-Wert Mittelwert (Standardabweichung)	Anzahl (108; 48 Wochen)	Wert in W 108 bzw. W 48 ^e Mittelwert (Standardabweichung)	Change from Baseline, W48 Mittelwert (Standardabweichung)	Change from Baseline, W108 Mittelwert (Standardabweichung)	Mittlerer Unterschied zwischen den Behandlungen nach 108 Wochen ^b 95 % KI [] p-Wert	Change from Baseline MMRM ^c	
								(W 108/W48) ^{c,d} LS-Mean ^d (Standardfehler)	LS-Mittelwertsdifferenz (vs. Placebo) (Standardfehler) 95 % KI [] p-Wert
TEMSO	T14 (n= 358)	2,670 (1,246)	268; 304	2,756 (1,397)	0,056 (0,736)	0,127 (0,914)	0,028 [-0,1325; 0,1885] p = 0,7320	0,226 (0,061)	-0,068 (0,081) [-0,226; 0,090] p = 0,4008
	Pb (n = 363)	2,679 (1,335)	262; 314	2,658 (1,607)	0,059 (0,914)	0,099 (0,966)		0,294 (0,061)	
TOWER	T14 (n= 370)	2,707 (1,353)	119; 290	2,572 (1,413)	-0,086 (0,790)	0,038 (0,881)	0,132 [-0,1004; 0,3644] p = 0,2643	-0,050 (0,052)	-0,139 (0,069) [-0,274; 0,004] p = 0,0429
	Pb (n = 388)	2,691 (1,366)	122; 318	2,656 (1,491)	0,033 (0,881)	-0,094 (0,950)		0,089 (0,050)	

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bt-Test für unverbundene Stichproben mit ungleichen Varianzen

^cMMRM = Mixed-effects Model with Repeated Measures, realisiert in SAS mit Proc Mixed. Feste Effekte in dem Modell waren Behandlung, Baseline EDSS-Stratum ($\leq 3,5$ / $> 3,5$), Region, Visite, Behandlung \times Visite-Wechselwirkung, Baseline-Wert und Behandlung \times Visite-Wechselwirkung. Für die Modellierung der within-patient Fehlerterme wurde eine nicht-strukturierte Korrelationsmatrix zu Grunde gelegt. Die Modellparameter wurden nach der restricted Maximum-Likelihood-Methode unter Verwendung des Newton-Raphson-Algorithmus geschätzt. Die Freiheitsgrade des Nenners wurden dabei nach der Approximation von Satterthwaite geschätzt.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

^dLS-Means sind die sich aus dem jeweiligen MMRM-Modell ergebenden Schätzer. Präsentiert sind aus der TEMSO-Studie die LS-Means zum Erhebungszeitraum 108 Wochen und aus der TOWER-Studie zum Erhebungszeitraum 48 Wochen.

^eFür die TEMSO-Studie sind die Daten zum Erhebungszeitpunkt W108 angegeben, für die TOWER-Studie entsprechend zum Zeitpunkt W48

TEMSO

Die Veränderungen des Baseline EDSS-Scores für die ITT Population wurden mit einer MMRM (mixed model for repeated measures)-Analyse untersucht. In Woche 108 konnte für keine der beiden Teriflunomid-Gruppen ein statistisch signifikanter Unterschied zu Placebo festgestellt werden.

TOWER

Die Veränderungen des Baseline EDSS-Scores für die ITT Population wurden mit einer MMRM-Analyse untersucht. Ein statistisch signifikanter Unterschied in Woche 48 gegenüber Placebo konnte für die 14 mg Teriflunomid Dosierung festgestellt werden: LS-Mittelwertsdifferenz -0,139; 95 % KI:[-0,274; 0,004]; p = 0,0429.

TENERE

In der TENERE-Studie liegen keine Daten für einen ausreichend langen Zeitraum vor. Darüber hinaus kann nicht ausgeschlossen werden, dass der EDSS-Score zu den Erhebungszeitpunkten unverblindet erhoben wurde. Aus den vorhandenen Daten lässt sich kein signifikanter Unterschied zwischen den Behandlungsgruppen feststellen. Auf eine detaillierte Darstellung wird aus den genannten Gründen verzichtet (die Daten finden sich im Studienbericht in Modul 5 des vorliegenden Dossiers [35]).

Fazit

Aus der mittleren Änderung des EDSS-Wertes kann kein klinisch relevanter Vorteil für Teriflunomid abgeleitet werden. Allerdings ist die mittlere Änderung des EDSS-Wertes als Auswertung eines patientenrelevanten Endpunkts kritisch zu sehen, da es sich um eine nicht lineare Ordinalskala handelt, d. h. die Bildung von Differenzen und Mittelwerten ist streng genommen nicht zulässig. Aus Gründen der Vollständigkeit und Transparenz ist die mittlere Änderung des EDSS an dieser Stelle dennoch aufgeführt.

4.3.1.3.2.4 Metaanalysen zur Endpunkten, die sich auf die Behinderungsprogression beziehen

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Die Studienergebnisse aus TEMSO und TOWER lassen sich bezüglich der Behinderungsprogression metaanalytisch zusammenfassen. Die Studie TENERE (anderer Komparator) lässt sich nicht im Sinne einer Metaanalyse mit TEMSO und TOWER zusammenfassen. Für die mittlere Änderung des MSFC-Z-Scores kann keine Metaanalyse durchgeführt werden, weil der MSFC-Wert von den eingeschlossenen Studien lediglich in TEMSO erhoben worden ist. In allen Operationalisierungen der Behinderungsprogression

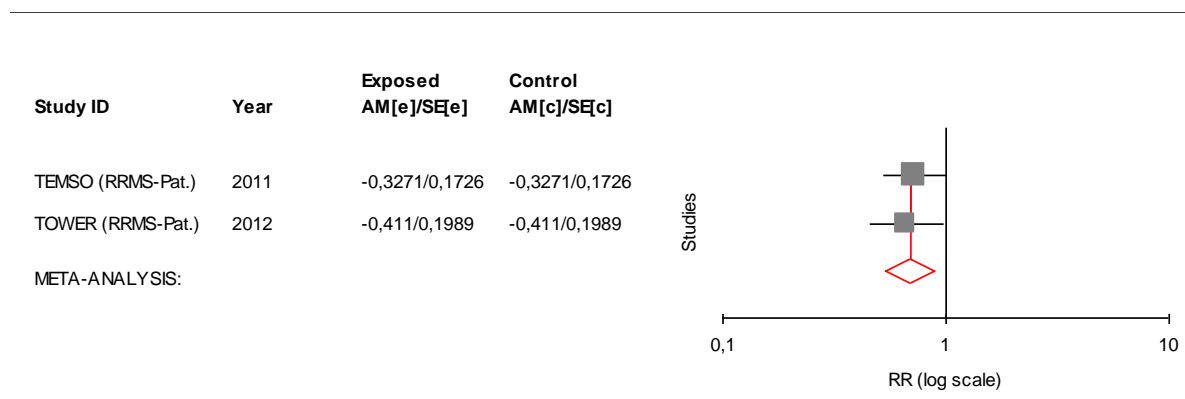
ergab sich ein homogenes Bild der Ergebnisse und ein mit den Ergebnissen der einzelnen Studien übereinstimmendes Ergebnis. Die „Zeit bis zur Behinderungsprogression“ konnte demnach gegenüber Placebo signifikant reduziert werden (Hazard Ratio von 0,696; 95 % KI: [0,539; 0,898]; $p = 0,0053$) und auch der Anteil an Patienten mit Behinderungsprogression war in der Teriflunomid 14 mg-Behandlungsgruppe entsprechend geringer (Relatives Risiko von 0,721; 95 % KI: [0,572; 0,909]; $p = 0,0056$). Die Änderung des EDSS-Wertes im Vergleich zum Baseline-Wert war in der Metaanalyse nicht signifikant.

Zeit bis zur Behinderungsprogression

Tabelle 4-56: Metaanalyse-Ergebnisse für die „Zeit bis zur Behinderungsprogression“ aus TEMSO und TOWER für die Subpopulation RRMS-Patienten

Analyse	Komparatoren	Hazard Ratio 95 % KI [] p-Wert ^a
Metaanalyse (TEMSO & TOWER)	Teriflunomid 14 mg versus Placebo	0,696 [0,539; 0,898] $p = 0,0053$ $I^2=0\%$, $p_Q=0,7501$

Abbildung 4-10: Forest-Plot der Metaanalyse von TEMSO und TOWER für die Zeit bis zur Behinderungsprogression

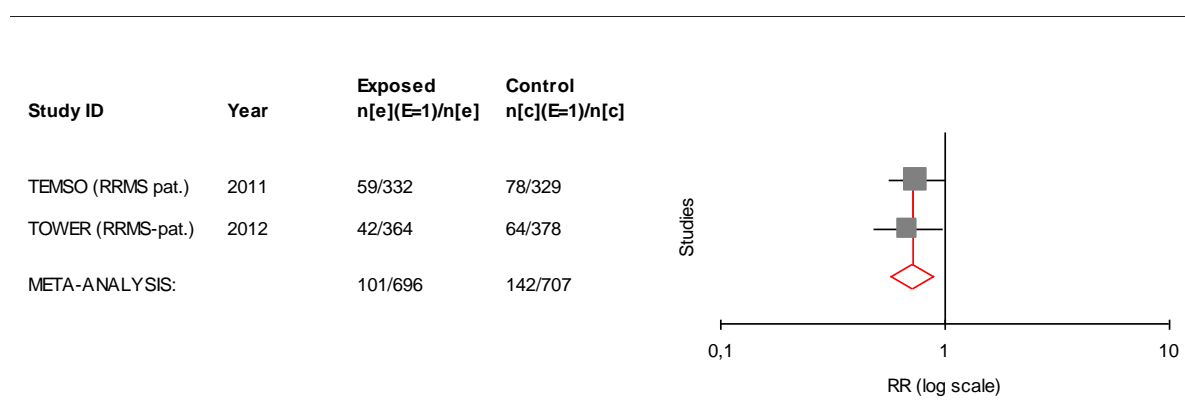


Anteil Patienten mit Behinderungsprogression

Tabelle 4-57: Ergebnis der Metaanalyse von TEMSO und TOWER für „Patienten mit Behinderungsprogression“ für die Subpopulation der RRMS-Patienten

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	0,721 [0,572; 0,909] p = 0,0056 I ² = 0%, p _Q = 0,6919	0,670 [0,506; 0,887] p = 0,0052 I ² = 0%, p _Q = 0,7730

Abbildung 4-11: Forest-Plot der Metaanalyse von TEMSO und TOWER (basierend auf den relativen Risiken) für „Patienten mit Behinderungsprogression“

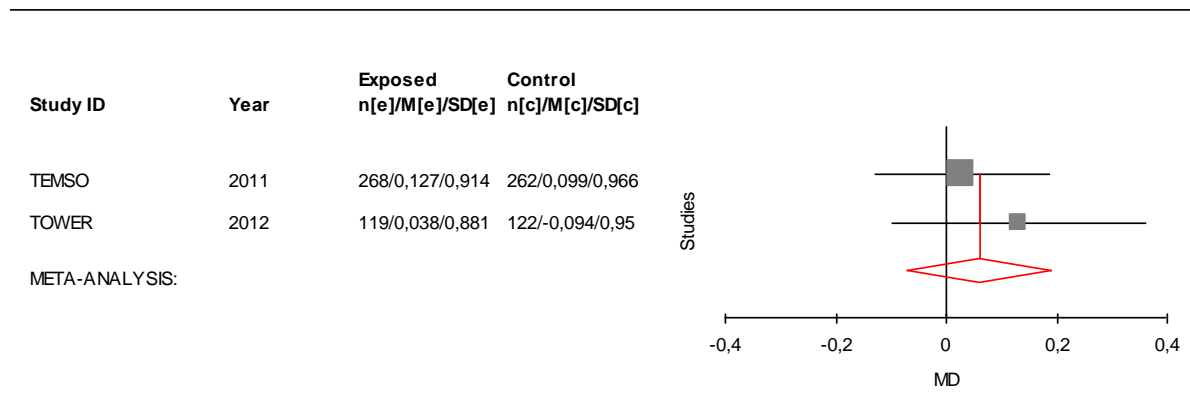


Mittlere Änderung des EDSS (2 Jahre)

Tabelle 4-58: Ergebnis der Metaanalyse von TEMSO und TOWER für „Mittlere Änderung des EDSS“ – für die gesamte Studienpopulation

Analyse	Komparatoren	Mittelwertsdifferenz 95 % KI p-Wert ^b
Metaanalyse TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	0,062 [-0,07; 0,193] p = 0,3582 I ² = 0 %, p _Q = 0,4687

Abbildung 4-12: Forest-Plot der Metaanalyse von TEMSO und TOWER (basierend auf den Mittelwertsdifferenzen) für „Mittlere Änderung des EDSS“



4.3.1.3.2.5 Zusammenfassung der Ergebnisse zu Endpunkten, die sich auf die Behinderungsprogression beziehen

Teriflunomid in der Dosierung von 14 mg zeigt gegenüber Placebo einen Vorteil bezüglich der Behinderungsprogression. In der Metaanalyse des wesentlichen Endpunktes „Zeit bis zur Behinderungsprogression“ ergab sich in der RRMS-Subpopulation mit einer Hazard Ratio von 0,696 (95 % KI: [0,539; 0,898]) eine klinisch relevante Verbesserung durch die Behandlung mit Teriflunomid, die auch statistisch signifikant war ($p = 0,0053$); für die gesamte Studienpopulation ergab sich kein von dieser Aussage abweichendes Bild, die Ergebnisse sind im Rahmen dieses Dossiers nicht dargestellt.

Bezogen auf die mittleren Änderungen der MSFC-Z- und EDSS-Scores konnte kein signifikanter Vorteil für Teriflunomid gezeigt werden.

4.3.1.3.3 Therapieversagen (Time to Failure) – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-59: Operationalisierung von „Therapieversagen (Time to Failure)“

Studie	Operationalisierung
TENERE	<p>Das Versagen der Therapie wird operationalisiert als „Time to Failure“ = „Zeit bis zum Therapieversagen“, wobei „Therapieversagen“ definiert ist als bestätigter Schub oder permanenter Therapieabbruch aus jedwedem Grund- was immer zuerst auftrat.</p> <p>Ein Schub war definiert als: Auftreten eines neuen klinischen Zeichens/Symptoms oder die klinische Verschlechterung eines vorhandenen Symptoms (eines, das mindestens 30 Tage lang stabil war), das mindestens 24 h anhielt und nicht von Fieber begleitet war. Die Schübe wurden durch den behandelnden Neurologen, basierend auf einer objektiven Bewertung (EDSS/FS-Grad) eines unabhängigen Bewerter (der untersuchende Neurologe), bestätigt, wenn eines der folgenden Kriterien erfüllt war:</p> <ul style="list-style-type: none"> • Erhöhung um 1 Punkt in mindestens 2 FS-Funktionen oder eine Erhöhung um 2 Punkte in mindestens einer FS-Funktion (ausgenommen Darm/Blase und Gehirn) • Eine Erhöhung um mindestens 0,5 Punkte im EDSS-Grad (es sei denn, der EDSS-Grad war vorher 0, dann war eine Erhöhung um 1,0 Punkte gefordert) im Vergleich zu der vorherigen stabilen Einstufung

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-60: Bewertung des Verzerrungspotenzials für „Time to Failure“ in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TENERE	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wird für die Studie TENERE auf Studienebene als niedrig eingestuft. Das ITT-Prinzip wurde adäquat umgesetzt, es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren.

Der Endpunkterheber, d. h. der auswertende Neurologe war in der TENERE-Studie verblindet.

Insgesamt wird das Verzerrungspotential für den Endpunkt „Therapieversagen“ als niedrig eingestuft.

Stellen Sie die Ergebnisse für den Endpunkt „Time to Failure“ für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-61: Ergebnisse für Time to Failure aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

Studie	Patientenzahl (ITT ^c)	Anzahl Patienten, bei denen das Ereignis eingetreten ist ^a (Anteil)	Schub (Anteil)	Permanenter Behandlungsabbruch (Anteil)	Andere Gründe für Therapieversagen ^b (Anteil)	Anzahl Patienten, die zensiert waren (Anteil)	Kaplan-Meier-Schätzer für die Wahrscheinlichkeit eines Therapieversagens zu den unterschiedlichen Zeitpunkten 95 % KI []			Hazard Ratio ^c 95 % KI [] p-Wert ^d
							24 W.	48 W.	96 W.	
Wahrscheinlichkeit für Therapieversagen „Time to Failure“										
TENERE ^h	T14 (n=111)	42 (37,8%)	26 (23,4%)	15 (13,5%)	1 (0,9%)	69 (62,2%)	0,243 [0,163; 0,323]	0,333 [0,246; 0,421]	0,411 [0,309; 0,514]	0,861 [0,564; 1,314] p = 0,5953
	Rb (n=104)	44 (42,3%)	16 (15,4%)	25 (24,0%)	3 (2,9%)	60 (57,7%)	0,298 [0,210; 0,386]	0,365 [0,273; 0,458]	0,444 [0,343; 0,544]	-
Wahrscheinlichkeit für Therapieversagen aufgrund eines Schubs^{a, g}										
	T14 (n=111)					85 (76,6 %)	0,151 [0,083; 0,219]	0,213 [0,133; 0,292]	0,292 [0,187; 0,397]	1,463 [0,784; 2,728] p = 0,2292
	Rb (n=104)					88 (84,6 %)	0,139 [0,069; 0,209]	0,164 [0,087; 0,240]	0,185 [0,100; 0,270]	-
Wahrscheinlichkeit für permanenten Behandlungsabbruch^g										
	T14 (n=111)					89 (80,2 %)	0,108 [0,050; 0,166]	0,181 [0,109; 0,252]	0,219 [0,131; 0,308]	0,568 [0,331; 0,975] p = 0,0467
	Rb (n=104)					71 (68,3 %)	0,183 [0,108; 0,257]	0,269 [0,184; 0,355]	0,334 [0,239; 0,429]	

T14 = Teriflunomid 14 mg; Rb = Rebif 44 µg

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

^aBeinhaltet auch Patienten, die nie behandelt wurden oder die falsche Behandlung erhalten haben.

^benthält Patienten, die keine Behandlung oder die falsche Behandlung erhalten haben

^cAbgeleitet aus einem Cox proportional-hazards Modell mit Behandlung, Baseline EDSS Stratum ($\leq 3,5$ / $> 3,5$) und Region als unabhängigen Variablen.

^dErmittelt mit einem Log-rank Test mit Stratifizierung nach Baseline EDSS Stratum ($\leq 3,5$ / $> 3,5$) und Region.

^es. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

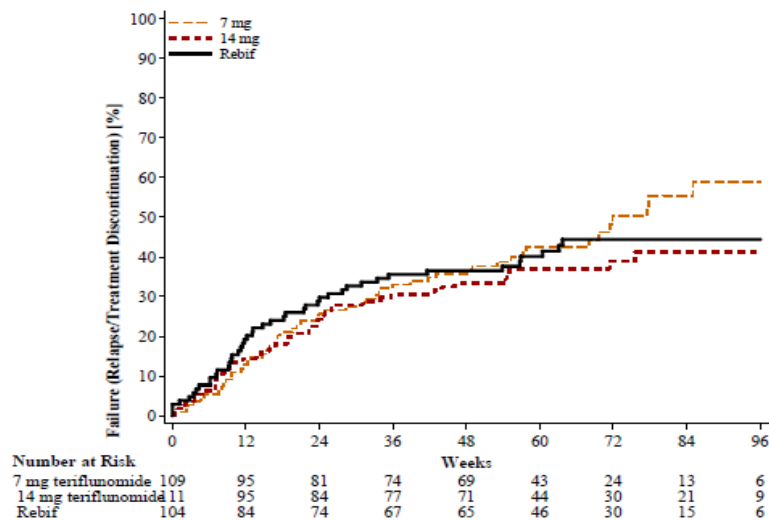
^gAnalysen sind nicht teil des Studienberichtes, Extraauswertungen s. Modul 5 [60]

^h99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Zu beachten: Die Anzahl der Patienten mit einem Schub und die Patienten mit dauerhaftem Therapieende aus anderen Gründen geben nicht additiv die Patienten mit Therapieversagen, da für das Therapieversagen das zuerst eintretende Ereignis gezählt wurde, ein Patient durchaus aber einen Schub gehabt haben kann, bevor er später aus anderen Gründen, die Therapie abgebrochen hat.

Abbildung 4-13: Kaplan-Meier-Kurve für „Zeit bis zum Therapieversagen“ in TENERE

Figure 2 - Kaplan-Meier plot of time to failure (confirmed relapse or permanent study treatment discontinuation) - ITT population



TENERE

Der nach der Kaplan-Meier-Methode geschätzte prozentuale Anteil an Patienten mit Therapieversagen (kumulative Rate) nach 48 Wochen betrug jeweils 33,3 % (95 % KI: [24,6 %; 42,1 %]) in der 14 mg Teriflunomid-Gruppe und 36,5 % (95 % KI: [27,3 %; 45,8 %]) in der Rebif[®]-Gruppe.

Nach Berechnungen der Hazard Ratios reduzierte die 14 mg Teriflunomid-Dosierung das Risiko eines Therapieversagens um 13,9 % im Vergleich zu Rebif[®] 44 µg. Es konnte jedoch keine statistisch signifikante Überlegenheit der Teriflunomid-Behandlung gegenüber Rebif[®] 44 µg festgestellt werden ($p = 0,5953$). Die Rate an dauerhaften Therapieabbrüchen war in der Rebif[®]-Gruppe höher als in der 14 mg Teriflunomid-Gruppe.

„Time to Failure“ ist ein zusammengesetzter Endpunkt. Für eine valide Auswertung und in Konformität mit dem IQWiG-Methodenpapier [17], müssen bei einem kombinierten Endpunkt die einzelnen Endpunkte, aus denen er sich zusammensetzt, auch getrennt ausgewertet werden. Die entsprechenden Auswertungen sind in Tabelle 4-61 dargestellt. In den möglichen Ereignissen, die als ein Therapieversagen gewertet werden können, zeigen sich unterschiedliche Ergebnisse. Die Wahrscheinlichkeit einen Schub zu erleiden war mit einer Hazard Ratio von 1,463 für die Patienten in der 14 mg-Teriflunomid-Gruppe höher als in der Rebif[®] 44 µg Gruppe, das Ergebnis war jedoch nicht statistisch signifikant ($p = 0,2292$). Dahingegen ist die Wahrscheinlichkeit für einen permanenten Behandlungsabbruch aus jedwedem Grund in der 14 mg Teriflunomid-Gruppe mit einer Hazard Ratio von 0,568 geringer als in der Rebif[®] 44 µg Gruppe. Die Verbesserung in diesem Teilendpunkt erreichte statistische Signifikanz. Es zeigt sich, dass die Patientenzahl, die in die TENERE-Studie eingeschlossen wurde, zu niedrig war, um die Verbesserung im kombinierten Endpunkt „Zeit bis zum Therapieversagen“ durch Teriflunomid nachzuweisen.

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Es handelt sich nur um eine einzige Studie, in der der Endpunkt „Therapieversagen“ erhoben worden ist. Eine Metaanalyse kann nicht durchgeführt werden.

4.3.1.3.4 Fatigue – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein. Tabelle 4-62: Operationalisierung von Fatigue (Fatigue Impact Scale = FIS)

Studie	Operationalisierung
<Studie 1>	
TEMSO, TOWER und TENERE	<p>Das Ausmaß der vom Patienten empfundenen Fatigue wurde mit dem Fatigue Impact Scale (FIS) erfasst. Der Fatigue Impact Scale (FIS) ist ein validierter, spezifischer, patienten-berichtetes Messinstrument, das entwickelt wurde um den Einfluss von Fatigue auf das Leben von Menschen mit MS zu bewerten. Die Anwendung kann durch den Patienten selbst, durch einen Befrager in einem Gespräch oder über das Telefon erfolgen.</p> <p>Der FIS setzt sich zusammen aus einem Gesamtwert (total score) und drei Subskalen, um den Einfluss der Fatigue auf die folgenden Bereiche zu erfassen:</p> <ul style="list-style-type: none"> - Kognitive Funktion (10 items), - Physische Funktion (10 items), - Psychosoziale Funktion (20 items). <p>Die Antworten können von 0 (kein Problem) bis 4 (extreme Probleme) reichen. Der FIS hat eine gute interne Konsistenz, wie durch das Maß Cronbach's Alpha (0,98 für die Gesamtanalyse des ganzen Fragebogens und >0,87 für alle Subskalen-Items nach Gruppen) gezeigt wurde. Die Möglichkeit, zwischen verschiedenen Patientenpopulationen und Gruppen mit unterschiedlichen MS-Schweregraden zu unterscheiden, konnte ebenfalls gezeigt werden.</p>

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-63: Bewertung des Verzerrungspotenzials für Fatigue (Fatigue Impact Scale = FIS) in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TEMZO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
TENERE	niedrig	nein	ja	nein	nein	hoch

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wurde für alle drei Studien TEMZO, TOWER und TENERE auf Studienebene als niedrig eingestuft. Es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren.

Der FIS ist ein patientenberichteter Endpunkt. Damit ist die Verblindung der Patienten für die Bewertung des Verzerrungspotentials dieses Endpunktes maßgeblich. In den beiden placebo-kontrollierten Studien TEMZO und TOWER waren die Patienten bezüglich ihrer Behandlung verblindet und das Verzerrungspotential insgesamt ist niedrig. In der TENERE-Studie waren die Patienten jedoch nicht verblindet, was möglicherweise das Ergebnis verzerrt.

Stellen Sie die Ergebnisse für den Endpunkt Fatigue für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-64: Ergebnisse für Fatigue aus RCT mit dem zu bewertenden Arzneimittel-FIS-Gesamtwert – für die gesamte Studienpopulation

Studie	Patienten- zahl (ITT ^a)	Anzahl ^b (W0)	Baseline (Woche 0) Mittelwert (Standard- abweichung)	Woche 48 Mittelwert (Standard- abweichung) Anzahl (W48)	Woche 108 Mittelwert (Standard- abweichung) Anzahl (W108)	Anzahl ^b (MMRM)	Change from Baseline ^b	Change from Baseline (MMRM) ^{b,c}	LS Mean ^d (SE)	LS Mean ^d Difference from Rebif oder Placebo (SE) 95 % KI [] p-Wert
							Mittelwert (Standard- abweichung)	(SE)		
Gesamtwert (total score)										
TEMSO	T14 (n= 358)	354	50,383 (35,952)	52,848 (37,937) 301	51,645 (38,043) 262	259	0,862 (28,729)	3,804 (1,670)	-0,497 (2,273) [-4,959; 3,965] p = 0,8271	
	Pb (n = 363)	360	53,199 (37,883)	52,393 (38,072) 310	48,610 (38,403) 254	251	-0,001 (28,328)	4,300 (1,670)		
TOWER	T14 (n= 370)	363	55,112 (38,119)	53,189 (39,613) 279	61,404 (41,373) 63	275	-0,231 (27,746)	1,915 (1,628)	-2,754 (2,187) [-7,047; 1,539] p = 0,2083	
	Pb (n = 388)	382	54,670 (37,894)	56,889 (41,867) 302	52,862 (45,223) 58	297	2,965 (29,359)	4,669 (1,576)		
TENERE ^c	T14 (n=111)	106	42,503 (37,791)	43,448 (40,920) 86	-	83	3,534 (22,291)	4,096 (3,027)	-5,003 (3,712) [-12,31; 2,307] p = 0,1789	
	Rb (n=104)	97	34,241 (32,719)	37,428 (35,775) 70	-	65	6,052 (23,473)	9,099 (3,209)		

T14 = 14 mg Teriflunomid; Rb = Rebif 44 µg; W = Woche ; SE = Standardfehler (standard error)

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

^bdie MMRM-Analyse und die Änderung in Bezug zum Baseline-Wert wurden in der TEMSO Studie mit dem Erhebungszeitpunkt Woche 108 durchgeführt, in den Studien TOWER und TENERE bezieht es sich auf den Zeitpunkt Woche 48

^cMMRM = Mixed-effects Model with Repeated Measures, realisiert in SAS mit Proc Mixed. Feste Effekte in dem Modell waren Behandlung, Baseline EDSS-Stratum ($\leq 3,5$ / $> 3,5$), Region, Visite, Behandlung \times Visite-Wechselwirkung, Baseline-Wert und Behandlung \times Visite-Wechselwirkung. Für die Modellierung der within-patient Fehlerterme wurde eine nicht-strukturierte Korrelationsmatrix zu Grunde gelegt. Die Modellparameter wurden nach der restricted Maximum-Likelihood-Methode unter Verwendung des Newton-Raphson-Algorithmus geschätzt. Die Freiheitsgrade des Nenners wurden dabei nach der Approximation von Satterthwaite geschätzt.

„Anzahl“ steht in diesem Zusammenhang für die Anzahl real vorhandener Datensätze nach 108 bzw. 48 Wochen.

^dLS-Means sind die sich aus dem MMRM-Modell ergebenden adjustierten Schätzer

^e99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Tabelle 4-65: Ergebnisse für Fatigue aus RCT mit dem zu bewertenden Arzneimittel-FIS-Subscore der kognitiven Dimension – für die gesamte Studienpopulation

Studie	Patienten- zahl (ITT ^a)	Anzahl (W0)	Baseline (Woche 0) Mittelwert (Standard- abweichung)	Woche 48 Mittelwert (Standard- abweichung) Anzahl (W48)	Woche 108 Mittelwert (Standard- abweichung) Anzahl (W108)	Anzahl ^b (MMRM)	Change from Baseline ^b	Change from Baseline (MMRM) ^{b,c}	
							Mittelwert (Standard- abweichung)	LS Mean ^d (SE)	LS Mean ^d Unterschied zu Rebif oder Placebo (SE) 95 % KI [] p-Wert
FIS-Subscore der kognitiven Dimension (Fatigue cognitive dimension score)									
TEMSO	T14 (n= 358)	354	11,669 (9,471)	12,490 (9,971) 301	12,427 (10,102) 262	259	0,579 (8,073)	1,021 (0,452)	0,041 (0,617) [-1,170; 1,251] p = 0,9474
	Pb (n = 363)	360	12,948 (10,296)	12,579 (9,686) 310	11,825 (9,938) 254	251	0,037 (7,749)	0,980 (0,452)	

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

TOWER	T14 (n= 370)	363	12,664 (10,036)	12,513 (10,095) 279	14,760 (11,062) 63	275	0,182 (7,610)	0,637 (0,447)	-0,759 (0,600) [-1,937; 0,419] p = 0,2065
	Pb (n = 388)	382	12,448 (10,071)	13,503 (11,010) 302	12,340 (12,257) 58	297	1,129 (7,984)	1,396 (0,432)	
TENERE ^c	T14 (n=111)	106	10,236 (10,207)	10,109 (10,407) 86	-	83	0,738 (6,731)	0,873 (0,838)	-1,465 (1,039) [-3,512; 0,582] p = 0,1599
	Rb (n=104)	97	7,798 (8,545)	8,891 (9,304) 70	-	65	1,546 (6,646)	2,338 (0,889)	

T14 = 14 mg Teriflunomid; Rb = Rebif 44 µg; W = Woche; SE = Standardfehler (standard error)

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien^bdie MMRM-Analyse und die Änderung in Bezug zum Baseline-Wert wurden in der TEMSO Studie mit dem Erhebungszeitpunkt Woche 108 durchgeführt, in den Studien TOWER und TENERE bezieht es sich auf den Zeitpunkt Woche 48^cMMRM = Mixed-effects Model with Repeated Measures, realisiert in SAS mit Proc Mixed. Feste Effekte in dem Modell waren Behandlung, Baseline EDSS-Stratum ($\leq 3,5$ / $> 3,5$), Region, Visite, Behandlung \times Visite-Wechselwirkung, Baseline-Wert und Behandlung \times Visite-Wechselwirkung. Für die Modellierung der within-patient Fehlerterme wurde eine nicht-strukturierte Korrelationsmatrix zu Grunde gelegt. Die Modellparameter wurden nach der restricted Maximum-Likelihood-Methode unter Verwendung des Newton-Raphson-Algorithmus geschätzt. Die Freiheitsgrade des Nenners wurden dabei nach der Approximation von Satterthwaite geschätzt.

„Anzahl“ steht in diesem Zusammenhang für die Anzahl real vorhandener Datensätze nach 108 bzw. 48 Wochen.

^dLS-Means sind die sich aus dem MMRM-Modell ergebenden adjustierten Schätzer^e99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Tabelle 4-66: Ergebnisse für Fatigue aus RCT mit dem zu bewertenden Arzneimittel-FIS-Subscore der physischen Dimension – für die gesamte Studienpopulation

Studie	Patienten- zahl (ITT ^a)	Anzahl (W0)	Baseline (Woche 0) Mittelwert (Standard- abweichung)	Woche 48 Mittelwert (Standard- abweichung) Anzahl (W48)	Woche 108 Mittelwert (Standard- abweichung) Anzahl (W108)	Anzahl ^b (MMRM)	Change from Baseline ^b	Change from Baseline (MMRM) ^{b,c}	LS Mean ^d (SE)	LS Mean ^d Unterschied zu Rebif oder Placebo (SE) 95 % KI [] p-Wert
							Mittelwert (Standard- abweichung)	(MMRM) ^{b,c}		
FIS-Subscore der physischen Dimension (Fatigue physical dimension score)										
TEMSO	T14 (n= 358)	353	14,818 (9,861)	15,098 (10,118) 301	14,803 (10,169) 262	258	0,034 (7,773)	0,771 (0,462)	0,191 (0,628) [-1,041; 1,424] p = 0,7605	
	Pb (n = 363)	360	15,270 (10,156)	14,686 (10,406) 310	13,557 (10,543) 254	251	-0,530 (8,029)	0,580 (0,462)		
TOWER	T14 (n= 370)	363	16,365 (10,738)	15,365 (10,740) 279	17,647 (11,323) 63	275	-0,578 (7,744)	0,094 (0,443)	-0,729 (0,595) [-1,896; 0,438] p = 0,2203	
	Pb (n = 388)	382	16,018 (10,786)	16,019 (11,312) 302	14,948 (11,932) 58	297	0,371 (8,185)	0,824 (0,429)		
TENERE ^c	T14 (n=111)	106	12,613 (10,161)	12,939 (11,147) 86	-	83	0,772 (6,066)	1,191 (0,869)	-0,316 (1,041) (-2,367 to 1,734) p = 0,7615	
	Rb (n=104)	97	11,061 (9,688)	10,856 (9,371) 70	-	65	0,677 (6,684)	1,507 (0,923)		

T14 = 14 mg Teriflunomid; Rb = Rebif 44 µg; W = Woche; SE = Standardfehler (standard error)

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bdie MMRM-Analyse und die Änderung in Bezug zum Baseline-Wert wurden in der TEMSO Studie mit dem Erhebungszeitpunkt Woche 108 durchgeführt, in den Studien TOWER und TENERE bezieht es sich auf den Zeitpunkt Woche 48

^cMMRM = Mixed-effects Model with Repeated Measures, realisiert in SAS mit Proc Mixed. Feste Effekte in dem Modell waren Behandlung, Baseline EDSS-Stratum ($\leq 3,5$ / $> 3,5$), Region, Visite, Behandlung \times Visite-Wechselwirkung, Baseline-Wert und Behandlung \times Visite-Wechselwirkung. Für die Modellierung der within-patient Fehlerterme wurde eine nicht-strukturierte Korrelationsmatrix zu Grunde gelegt. Die Modellparameter wurden nach der restricted Maximum-Likelihood-Methode unter Verwendung des Newton-Raphson-Algorithmus geschätzt. Die Freiheitsgrade des Nenners wurden dabei nach der Approximation von Satterthwaite geschätzt.

„Anzahl“ steht in diesem Zusammenhang für die Anzahl real vorhandener Datensätze nach 108 bzw. 48 Wochen.

^dLS-Means sind die sich aus dem MMRM-Modell ergebenden adjustierten Schätzer

^e99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Tabelle 4-67: Ergebnisse für Fatigue aus RCT mit dem zu bewertenden Arzneimittel-FIS-Subscore der psychosozialen Dimension – für die gesamte Studienpopulation

							Change from Baseline ^b	Change from Baseline (MMRM) ^{b,c}		
Studie	Patienten-zahl (ITT ^a)	Anzahl (W0)	Baseline (Woche 0) Mittelwert (Standard-abweichung)	Woche 48 Mittelwert (Standard-abweichung) Anzahl (W48)	Woche 108 Mittelwert (Standard-abweichung) Anzahl (W108)	Anzahl ^b (MMRM)	Mittelwert (Standard-abweichung)	LS Mean ^d (SE)	LS Mean ^d Unterschied zu Rebif oder Placebo (SE) 95 % KI [] p-Wert	
FIS-Subscore der psychosozialen Dimension (Fatigue psychosocial dimension score)										
TEMSO	T14 (n= 358)	354	23,872 (18,511)	25,269 (19,486) 301	24,427 (19,291) 262	259	0,287 (14,808)	2,174 (0,848)	-0,669 (1,152) [-2,931; 1,592] p = 0,5615	

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

	Pb (n = 363)	360	24,982 (18,914)	25,116 (19,343) 310	23,222 (19,262) 254	251	0,476 (14,184)	2,844 (0,847)	
TOWER	T14 (n= 370)	363	26,094 (19,285)	25,308 (20,227) 279	28,994 (20,689) 63	275	0,151 (14,166)	1,275 (0,830)	-1,284 (1,113) [-3,468; 0,901] p = 0,2491
	Pb (n = 388)	382	26,207 (19,144)	27,362 (20,939) 302	25,571 (22,665) 58	297	1,459 (14,965)	2,558 (0,803)	
TENERE ^c	T14 (n=111)	106	19,663 (19,480)	20,398 (20,429) 86	-	83	2,020 (11,694)	2,703 (1,529)	-2,813 (1,880) [-6,514; 0,888] p = 0,1357
	Rb (n=104)	97	15,387 (16,048)	17,685 (18,107) 70	-	65	3,834 (12,120)	5,515 (1,623)	

T14 = 14 mg Teriflunomid; Rb = Rebif 44 µg; W = Woche; SE = Standardfehler (standard error)

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien^bdie MMRM-Analyse und die Änderung in Bezug zum Baseline-Wert wurden in der TEMSO Studie mit dem Erhebungszeitpunkt Woche 108 durchgeführt, in den Studien TOWER und TENERE bezieht es sich auf den Zeitpunkt Woche 48^cMMRM = Mixed-effects Model with Repeated Measures, realisiert in SAS mit Proc Mixed. Feste Effekte in dem Modell waren Behandlung, Baseline EDSS-Stratum ($\leq 3,5$ / $> 3,5$), Region, Visite, Behandlung \times Visite-Wechselwirkung, Baseline-Wert und Behandlung \times Visite-Wechselwirkung. Für die Modellierung der within-patient Fehlerterme wurde eine nicht-strukturierte Korrelationsmatrix zu Grunde gelegt. Die Modellparameter wurden nach der restricted Maximum-Likelihood-Methode unter Verwendung des Newton-Raphson-Algorithmus geschätzt. Die Freiheitsgrade des Nenners wurden dabei nach der Approximation von Satterthwaite geschätzt.

„Anzahl“ steht in diesem Zusammenhang für die Anzahl real vorhandener Datensätze nach 108 and 48 Wochen.

^dLS-Means sind die sich aus dem MMRM-Modell ergebenden adjustierten Schätzer^e99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

TEMSO

In der MMRM-Analyse konnte keine statistisch signifikante Differenz (LS-Mittelwerte) der mittleren Änderung des FIS-Gesamtwertes im Vergleich zum Baseline-Wert zum Zeitpunkt Woche 108 zwischen der 14 mg-Teriflunomid-Gruppe und der Placebo-Gruppe festgestellt werden.

Zur besseren Vergleichbarkeit mit den Ergebnissen aus der TOWER und TENERE-Studie sind in den Tabelle 4-64 bis Tabelle 4-67 zusätzlich die Werte für die Woche 48 angegeben.

TOWER

Für den Großteil der erhobenen FIS-Gesamt- und Subscores zeigen die Daten einen numerischen Vorteil für die Behandlung mit Teriflunomid. Für den Erhebungszeitpunkt 48 Wochen ist das Ergebnis für den Gesamtwert nicht statistisch signifikant. Der Hinweis auf einen Vorteil zeigt sich auch in den Werten der drei Subskalen (kognitive, physische und psychosoziale Dimension), ist aber auch hier nicht signifikant.

Zusätzliche Analysen mit einem ANCOVA-Modell bestätigten die Ergebnisse der MMRM-Analyse.

Zur besseren Vergleichbarkeit mit den Ergebnissen aus der TEMSO-Studie, sind in Tabelle 4-64 bis Tabelle 4-67 zusätzlich die erhobenen Werte für die Woche 108 angegeben.

TENERE

Nach 48 Wochen betragen die mittleren Änderungen des FIS-Gesamtwertes 3,534 in der 14 mg Teriflunomid-Gruppe und 6,052 in der Rebif[®]-Gruppe. In der MMRM-Analyse ergab sich in den durch das Modell adjustierten Mittelwerten (LS-Means) entsprechend ein Unterschied von -5,003 (95% KI: [-12,31; 2,307]), der nicht statistisch signifikant ist ($p = 0,1789$). Der Hinweis auf einen Vorteil zeigt sich auch in den Werten der drei Subskalen (kognitive, physische und psychosoziale Dimension), ist aber auch hier nicht signifikant.

Zusätzliche Analysen mit einem ANCOVA-Modell bestätigten die Ergebnisse der MMRM-Analyse.

Fatigue - Fazit

Für den Endpunkt Fatigue konnte keine statistisch signifikante Überlegenheit von Teriflunomid 14 mg gegenüber Placebo oder Rebif[®] 44 µg festgestellt werden. Bezüglich des Gesamtscores zeigt sich in allen drei Studien zwar ein Hinweis für einen Vorteil für Teriflunomid, der jedoch keine statistische Signifikanz erreichte. In den Studien TOWER und TENERE ist dieser Vorteil auch in allen erhobenen Subscores zu sehen. Bei Auswertung der Daten mit einem ANCOVA-Modell erreicht der Vorteil in einigen Teilauswertungen statistische Signifikanz. Die vorliegenden Daten deuten die Möglichkeit eines Vorteils für Teriflunomid sowohl gegenüber Placebo als auch gegenüber Rebif[®] 44 µg an, er wird durch

die Studiendaten jedoch nicht belegt. Eine Aussage zum Einfluss auf die Fatigue über einen indirekten Vergleich wäre sinnvoll, ist jedoch nicht möglich, da die Fatigue in entsprechenden Interferon-beta-Studien nicht mit dem validierten Messinstrument FIS erfasst worden ist (s. Modul 4.3.2).

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Eine Metaanalyse zu den berichteten LS-Means aus der TEMSO- und der TOWER-Studie wird nicht durchgeführt, da sich die Analysen auf unterschiedliche Erhebungszeitpunkte beziehen.

Die Ergebnisse der TENERE-Studie können nicht mit den Ergebnissen aus TEMSO und TOWER zusammengefasst werden, weil sich die Komparatoren der Studien unterscheiden.

4.3.1.3.5 Krankheitslast (Burden of Disease (MRT)) – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-68: Operationalisierung von Krankheitslast (Burden of Disease - MRT)

Studie	Operationalisierung
<Studie 1>	
TEM SO	<p>Die Krankheitslast (Burden of Disease) wird als die Zunahme des Gesamtvolumens an Läsionen operationalisiert, d. h. Gesamtvolumen hyperintenser T2-Läsionen plus Gesamtvolumen hypointenser T1-Läsionen.</p> <p>Zur unterstützenden Aussage wurden weitere MRT-bezogene Endpunkte ausgewertet.</p> <p>Magnetresonanztomographie</p> <p>Die MRT (= Magnetresonanztomographie, entspricht MRI = Magnetic Resonance Imaging) wurde mit Standard-Bildaufnahmeparametern durchgeführt. Diese schlossen ein:</p> <ul style="list-style-type: none"> • Transversales Spinecho (axial dual echo spin density) und T2-gewichtete Sequenzen (T2-weighted sequences) • Sagittale und transversale Dark-Fluid-Bildgebung (sagittal and axial fluid-attenuated inversion recovery (FLAIR)), • volumetrische 3D-SPGR (3-dimensional spoiled gradient recalled), • Transversal T1-gewichtete Aufnahme vor und nach der Gabe von Gadolinium (an axial T1-weighted acquisition before and after the administration of Gd) <p>Die Daten, die aus dem multimodalen Bilddatenmaterial extrahiert werden konnten, umfassten die absolute Anzahl und Volumen Gadolinium-anreichernder Läsionen, Volumen an T2-Läsionen, Volumen an T1-hypointensen Läsionen und parenchymaler Anteil des Gehirns.</p> <p>Die Bilddaten wurden an MRT-Einrichtungen aufgenommen, die mit den jeweiligen Kliniken verbunden waren , und wurden dann zum Texas University MRI Analysis Center in Houston geschickt, wo sowohl die Daten weiterverarbeitet wurden, als auch die Gewebe- und Läsionen-Segmentierung und Datenextraktion vorgenommen wurden.</p>

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-69: Bewertung des Verzerrungspotenzials für Krankheitslast (Burden of Disease (MRT)) in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
TEM SO	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wurde für die Studie TEMSO auf Studienebene als gering eingestuft. Es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren. Die aufgenommenen MRT-Daten wurden in einer zentralen Einrichtung analysiert, so dass die Gefahr einer subjektiven Auswertung minimal gehalten wurde. Der Endpunkt wurde verblindet erhoben. Damit ist das Verzerrungspotential für diesen Endpunkt niedrig.

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Stellen Sie die Ergebnisse für den Endpunkt Krankheitslast für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

4.3.1.3.5.1 Krankheitslast (Burden of Disease)

Tabelle 4-70: Ergebnisse für Krankheitslast (Burden of Disease) aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

Krankheitslast (Burden of Disease ^a)										
Studie	Patienten- zahl (ITT ^b)	Baseline: Gesamt- volumen [ml] (Läsionen) Mittelwert (SD)	Gesamt- volumen ml an Läsionen (W108) Mittelwert (SD)	Change from Baseline [ml] (W108) Mittelwert (SD)	Change from Baseline [%] (W108) Mittelwert (SD)	Baseline für Kubikwurzel- transfor- mierte Volumen- daten Mittelwert (SD)	Woche 108: Kubikwurzel- transfor- mierte Volumen- daten Mittelwert (SD)	Change from Baseline Mittelwert (SD) Anzahl ^c	Change from Baseline MMRM (W108) ^c	
									LS- Mean ^d (SE)	LS- Mean ^d - Differenz (SE) 95% KI[] p-Wert
TEMSEO	T14 (n= 358) 355 in W0	18,076 (17,488)	18,172 (16,378)	0,723 (7,591)	16,549 (62,264)	2,359 (0,825)	2,383 (0,801)	0,045 (0,300) n = 260	0,043 (0,018)	-0,089 (0,025) [-0,137; - 0,041] p = 0,0003
	Pb (n = 363) 358 in W0	19,337 (18,939)	20,728 (19,858)	2,208 (7,002)	31,676 (83,769)	2,383 (0,892)	2,465 (0,873)	0,111 (0,307) n = 256	0,132 (0,018)	

T14 = 14 mg Teriflunomid; Rb = Rebif 44 µg; W = Woche; SE = Standardfehler (standard error); SD = Standardabweichung (standard deviation)

^aBurden of disease ist definiert als das Gesamtvolumen an hyperintensen T2-Läsionen und hypointensen T1-Läsionen, gemessen durch MRT-Scan

^bs. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^cMMRM = Mixed-effects Model with Repeated Measures, realisiert in SAS mit Proc Mixed. Feste Effekte in dem Modell waren Behandlung, Baseline EDSS-Stratum ($\leq 3,5$ / $> 3,5$), Region, Visite, Behandlung \times Visite-Wechselwirkung, Baseline-Wert und Behandlung \times Visite-Wechselwirkung. Für die Modellierung der within-

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

patient Fehlerterme wurde eine nicht-strukturierte Korrelationsmatrix zu Grunde gelegt. Die Modellparameter wurden nach der restricted Maximum-Likelihood-Methode unter Verwendung des Newton-Raphson-Algorithmus geschätzt. Die Freiheitsgrade des Nenners wurden dabei nach der Approximation von Satterthwaite geschätzt.

„Anzahl“ steht in diesem Zusammenhang für die Anzahl real vorhandener Datensätze nach 108 Wochen.

^dLS-Means sind die sich aus dem MMRM-Modell ergebenden adjustierten Schätzer

TEMSO

Zu allen Erhebungszeitpunkten (Wochen 24, 48, 72 und 108 [32]) zeigte sich unter Teriflunomid eine geringere Krankheitslast als unter Placebo. In Woche 108 war das Volumen an abnormalem Gewebe in der 14 mg- Teriflunomid-Gruppe gegenüber der Placebo-Gruppe statistisch signifikant reduziert (-0,089; 95 % KI.[-0,137; -0,041]).

4.3.1.3.5.2 Übersicht weiterer MRT-Parameter

Tabelle 4-71: Ergebnisse für ausgewählte weitere MRT-Parameter aus RCT mit dem zu bewertenden Arzneimittel - für die gesamte Studienpopulation

Gadolinium-anreichernde T1-Läsionen									
Studie	Patientenzahl (ITT ^c)	Baseline Patienten mit ≥1 Gadolinium-anreichernden T1-Läsionen	Post-baseline Anzahl und Anteil Patienten ohne Gd-anreichernde Läsionen Anzahl	Post-baseline Patienten mit ≥1 Gadolinium-anreichernder T1-Läsion Anzahl	Absolute Zahl Gadolinium-anreichernder T1-Läsionen	Absolute Zahl der Scans	Nicht adjustierte Rate Gd-anreichernde T1-Läsionen/Scan ^a	Adjustierte Rate Gd-anreichernde T1-Läsionen/Scan ^b 95 % KI []	Relatives Risiko 95 % KI [] p-Wert
TEMSO	T14 (n= 358) 355 in W0	125 (35,2 %)	218 (64,1 %) 340	122 (35,9 %) 340	478	1176	0,406	0,261 [0,167; 0,407]	0,196 [0,120; 0,321] p < 0,0001

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Pb	137	135	211 (61,0 %)	1810	1191	1,52	1,331
(n = 363)	(38,2 %)	(39,0 %)	346				[1,059; 1,673]
358 in W0		346					
Hypointense post-gadolinium T1-Läsionen							
Studie	Patienten- zahl (ITT ^c)	Baseline für Kubikwurzel- transformierte Volumendaten Mittelwert (Standard- abweichung)	Change from Baseline (W108) Mittelwert (Standard- abweichung)	Kubikwurzel- transformierte Volumendaten in W108 (Standard- abweichung)	MMRM ^d Anzahl	Change from Baseline MMRM (W108) ^d	
						LS-Mean ^e (Standardfehler)	LS-Mean-Differenz ^e (Standard-Fehler) 95% KI [] p-Wert
TEM SO	T14	1,251 (0,493)	0,062 (0,142)	1,321 (0,497)	260	0,066 (0,009)	-0,030 (0,013)
	(n= 358)			261			[-0,055; -0,006]
	Pb	1,278 (0,541)	0,088 (0,153)	1,341 (0,528)	256	0,096 (0,009)	p = 0,0161
	(n = 363)			258			
358 in W0							
Hyperintense T2-gewichtete Läsionen							
Studie	Patienten- zahl (ITT ^c)	Baseline Kubikwurzel- transformierte Volumendaten Mittelwert (Standard- abweichung)	Change from Baseline (W108) Mittelwert (Standard- abweichung)	Kubikwurzel- transformierte Volumendaten in W108 (Standard- abweichung)	MMRM Anzahl ^d	Change from Baseline MMRM (W108) ^d	
						LS-Mean ^e (Standardfehler)	LS-Mean-Differenz ^e (Standard- Fehler) 95% KI[] p-Wert
TEM SO	T14	2,222 (0,782)	0,030 (0,305)	2,226 (0,753)	260	0,026 (0,018)	-0,089 (0,025)
	(n= 358)			261			[-0,139; -0,040]
	355 in W0						p = 0,0004

 Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

Pb (n = 363) 358 in W0	2,240 (0,840)	0,096 (0,315)	2,312 (0,823)	256	0,116 (0,018)
			258		

Gd = Gadolinium

^aentspricht der absoluten Zahl Gd-anreichernder T1-Läsionen dividiert durch die absolute Anzahl der Scans

^bPoisson-Modell mit der absoluten Anzahl Gadolinium-anreichernder T1-Läsionen als abhängiger Variable und Behandlung, EDSS-Baseline Strata, Region und Baseline-Anzahl Gadolinium-anreichernder T1-Läsionen als unabhängige Variablen, sowie die log-transformierte Anzahl der Scans als Offset-Variable

^cs. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^dMMRM = Mixed-effects Model with Repeated Measures, realisiert in SAS mit Proc Mixed. Feste Effekte in dem Modell waren Behandlung, Baseline EDSS-Stratum ($\leq 3,5$ / $> 3,5$), Region, Visite, Behandlung \times Visite-Wechselwirkung, Baseline-Wert und Behandlung \times Visite-Wechselwirkung. Für die Modellierung der within-patient Fehlerterme wurde eine nicht-strukturierte Korrelationsmatrix zu Grunde gelegt. Die Modellparameter wurden nach der restricted Maximum-Likelihood-Methode unter Verwendung des Newton-Raphson-Algorithmus geschätzt. Die Freiheitsgrade des Nenners wurden dabei nach der Approximation von Satterthwaite geschätzt.

„Anzahl“ steht in diesem Zusammenhang für die Anzahl real vorhandener Datensätze nach 108 Wochen.

^eLS-Means sind die sich aus dem MMRM-Modell ergebenden adjustierten Schätzer

TEM SO

In Tabelle 4-71 sind die Ergebnisse zu T1- und T2- gewichteten Läsionen dargestellt, weitere Auswertungen (z. B. Atrophie, Volumen an weißer und grauer Substanz, Z4 score und mittlere Anzahl Gd-anreichernder Läsionen) finden sich in den Anhängen des Studienberichts [32]. In Woche 108 war das Volumen an abnormalem Gewebe bezogen auf jeweils Gadolinium-anreichernde Läsionen, hypointense T1-Läsionen und hyperintense T2-Läsionen in der 14 mg Teriflunomid-Gruppe gegenüber der Placebo-Gruppe statistisch signifikant reduziert.

Zusammenfassung

Teriflunomid hatte im Vergleich zu Placebo einen positiven Einfluss auf diverse MRT-Parameter. Zu diesen MRT-Parametern gehört neben der Krankheitslast auch die Anzahl Gadolinium-anreichernder T1-Läsionen. Sowohl die Anzahl an Patienten, die frei von Gd-

anreichernden Läsionen waren als auch die adjustierte Rate (Läsionen pro Scan) war in der 14 mg Teriflunomid-Gruppe signifikant niedriger als in der Placebo-Gruppe. Die relative Risikoreduktion betrug 80,4 % (relatives Risiko 0,196; 95 % KI: [0,120; 0,321]). Da die Gadolinium-anreichernden Läsionen einen Hinweis auf akute entzündliche Prozesse im Gehirn geben (mit einhergehender erhöhter Durchlässigkeit der Blut-Hirn-Schranke), steht dieses Ergebnis in gutem Einklang mit der anti-inflammatorischen Wirkung von Teriflunomid (s. Modul 2).

4.3.1.3.5.3 Aktive Läsionen („unique active lesions“)

Tabelle 4-72: Ergebnisse für aktive Läsionen („unique active lesions“) aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

aktive Läsionen („unique active lesions“)								
Studie	Patienten-zahl (ITT ^a) Anzahl in der Auswertung	Post-baseline Anzahl und Anteil Patienten ohne unique active lesions	Post-baseline Patienten mit ≥1 unique active lesions	Absolute Zahl unique active lesions	Absolute Zahl der Scans	Nicht adjustierte Rate unique active lesions / Scan ^b	Adjustierte Rate unique active lesions / Scan ^c 95 % KI []	Relatives Risiko 95 % KI [] p-Wert
TEM SO	T14 (n= 358) 340	127 (37,4 %)	213 (62,6 %)	1199	1178	1,018	0,754 [0,576; 0,987]	0,306 [0,228; 0,411] p < 0,0001
	Pb (n = 363) 346	73 (21,1 %)	273 (78,9 %)	3360	1194	2,814	2,463 [2,102; 2,886]	

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bGesamtzahl an unique active lesions, die während der Studie aufgetreten sind geteilt durch die absolute Anzahl an Scans, die im Rahmen der Studie durchgeführt worden sind

^cPoisson-Modell mit der Gesamtzahl an „unique active lesions“ als Response-Variable und Behandlung, EDSS-Baseline-Wert, Region und Baseline Anzahl an „unique active lesions“ als Kovariate sowie der log-transformierten Anzahl der Scans als Offset-Variable.

TEM SO

Die Anzahl der aktiven Läsionen pro Scan ist mit Bezug auf die CHMP-guideline für die Bewertung von Interferon-haltige Biosimilars [61] gesondert aufgeführt. In dieser Leitlinie wird für den Wirksamkeitsnachweis die Wahl von MRT-Parametern als primärer Endpunkt empfohlen, wobei die „combined unique active lesions“ als der sensitivste Parameter gewertet wird. Mit einem relativen Risiko von 0,306 (95 % KI: [0,228; 0,411]) konnte die Rate an aktiven Läsionen im Vergleich zu Placebo signifikant gesenkt werden.

Krankheitslast (Burden of Disease (MRT)) - Fazit

Die verschiedenen MRT-Parameter haben einen hohen Stellenwert in Phase II- Studien, was so auch in der entsprechenden EMA-Leitlinie anerkannt wird [1, 2], und in der Diagnostik der MS [15, 62]. Für die Patienten ist es von großem Interesse und daher auch relevant, ob und in welchem Ausmaß die Entzündungsaktivität in ihrem ZNS durch eine Therapie gesenkt werden kann. Über die Resultate, die in bildgebenden Verfahren gewonnen werden, können darüber hinaus erste Rückschlüsse auf den Wirkmechanismus einer Therapie gezogen werden und es ist anzunehmen, dass die Bedeutung der MRT-Parameter in zukünftigen Studien zunehmen wird [63]. Daher ist ausgesprochen sinnvoll, zusätzlich zu den klinischen Endpunkten Schübe und Behinderungsprogression auch die MRT-Parameter übersichtlich darzustellen. Hier konnte ein deutlicher Vorteil für Teriflunomid im Vergleich zu Placebo gezeigt werden. Dieses Ergebnis ist im guten Einklang mit der antiinflammatorischen Aktivität von Teriflunomid und dem positiven Einfluss auf die klinischen Endpunkte, der für den Wirkstoff gezeigt werden konnte (s. Kapitel 4.3.1.3.1 und 4.3.1.3.2).

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Lediglich in der TEMSO-Studie sind MRT-Parameter erfasst worden. Eine Metaanalyse ist aus diesem Grund nicht durchführbar.

4.3.1.3.6 Patientenzufriedenheit – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-73: Operationalisierung von Patientenzufriedenheit

Studie	Operationalisierung
<Studie 1>	
TENERE	<p>Der Endpunkt Patientenzufriedenheit wurde mit Hilfe des Messinstrumentes „Treatment Satisfaction Questionnaire for Medication“ (TSQM) erfasst. Der TSQM wurde in der Version 1.4 verwendet.</p> <p>Der TSQM ist ein vom Patienten selbst angewendetes Instrument zur Messung der maßgeblichen Dimensionen der Patientenzufriedenheit hinsichtlich ihrer Medikation. Der Fragebogen wurde entwickelt, um ein breites Spektrum an Behandlungsintentionen (d. h. kurativ, präventiv, symptomlindernd) und verschiedene Applikationsformen (d. h. Injektionen, orale Verabreichung, topische Applikation, Inhalation) abbilden zu können. Es gibt Hinweise, dass der TSQM ein gutes Instrument ist um die Therapieadhärenz der Patienten vorauszusagen- das betrifft verschiedene Patientenpopulationen und Arten der Medikation.</p> <p>Der TSQM 1.4 besteht aus 14 Fragen. Eine Frage zielt auf die allgemeine Zufriedenheit mit der Therapie, die weiteren Fragen zielen auf die Dimensionen:</p> <ul style="list-style-type: none"> • Wirksamkeit (3 Fragen) • Verträglichkeit (5 Fragen) • Anwenderfreundlichkeit (5 Fragen) <p>Der TSQM lässt sich in ca. 5 Minuten ausfüllen.</p> <p>Die einzelnen Fragen werden unterschiedlich entweder in Kategorien von 1 bis 5, 1 bis 7 oder mit ja oder nein beantwortet. Insgesamt ergeben sich Skalenwerte von 0 bis 100.</p>

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-74: Bewertung des Verzerrungspotenzials für Patientenzufriedenheit in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TENERE	niedrig	nein	ja	nein	nein	hoch

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wurde für die Studie TENERE auf Studienebene als niedrig eingestuft. Es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren.

Der TSQM ist ein Fragebogen, der von dem Patienten selbst ausgefüllt wurde. Damit ist der Patient als Endpunkterheber zu werten. In der TENERE-Studie waren die Patienten jedoch nicht verblindet, was möglicherweise das Ergebnis verzerrt. Andererseits ist es auch nicht möglich, das subjektive Empfinden der Patienten zur Zufriedenheit mit der Behandlung verblindet zu erheben, wenn eine parenterale Darreichungsform mit einer oralen Darreichungsform verglichen wird. Obwohl also aufgrund der mangelnden Verblindung das Verzerrungspotential möglicherweise höher ist als bei anderen Endpunkten, ist es die beste Form der Datenerhebung, die an dieser Stelle möglich ist.

Stellen Sie die Ergebnisse für den Endpunkt Patientenzufriedenheit für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-75: Ergebnisse für Patientenzufriedenheit aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

Studie	Patienten-zahl (ITT ^a)	Anzahl ^b	Wert in Woche 48 Mittelwert (Standardabweichung)	LS Mean ^c (mittlerer Fehler)	Differenz der beiden Interventionen MMRM (W48) ^b
					LS Mean ^c Differenz (mittlerer Fehler) 95 % KI [] p-Wert
Wirksamkeit					
TENERE ^d	T14 (n=111)	84	67,16 (22,66)	63,133 (2,748)	3,835 (3,512) [-3,082; 10,75] p = 0,2759
	Rb (n=104)	69	63,45 (20,09)	59,298 (2,971)	
Verträglichkeit					
TENERE ^d	T14 (n=111)	84	95,09 (13,27)	93,149 (2,339)	21,77 (2,688) [16,48; 27,07] p < 0,0001
	Rb (n=104)	69	75,27 (25,28)	71,376 (2,495)	
Anwenderfreundlichkeit					
TENERE ^d	T14 (n=111)	85	90,78 (11,13)	89,851 (1,983)	27,96 (2,338) [23,35; 32,56] p < 0,0001
	Rb (n=104)	70	63,17 (19,06)	61,895 (2,111)	
Allgemeine Zufriedenheit					
TENERE ^d	T14 (n=111)	84	72,79 (17,86)	68,818 (2,782)	7,843 (3,244) [1,458; 14,23] p = 0,0162
	Rb (n=104)	70	65,38 (22,81)	60,975 (2,942)	

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bMMRM = Mixed-effects Model with Repeated Measures, realisiert in SAS mit Proc Mixed. Feste Effekte in dem Modell waren Behandlung, Baseline EDSS-Stratum ($\leq 3,5$ / $> 3,5$), Region, Visite, Behandlung \times Visite-

Wechselwirkung, Baseline-Wert und Behandlung \times Visite-Wechselwirkung. Für die Modellierung der within-patient Fehlerterme wurde eine nicht-strukturierte Korrelationsmatrix zu Grunde gelegt. Die Modellparameter wurden nach der restricted Maximum-Likelihood-Methode unter Verwendung des Newton-Raphson-Algorithmus geschätzt. Die Freiheitsgrade des Nenners wurden dabei nach der Approximation von Satterthwaite geschätzt.

„Anzahl“ steht in diesem Zusammenhang für die Anzahl real vorhandener Datensätze nach 48 Wochen.

^cLS-Means sind die sich aus dem MMRM-Modell ergebenden Schätzer.

^d99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

TENERE

Allgemeine Zufriedenheit

Die Patienten waren mit der Behandlung in der 14 mg Teriflunomid-Gruppe zufriedener als in der Rebif[®]-Gruppe. Die Patienten erreichten in der 14 mg Teriflunomid Gruppe einen signifikant höheren Score in Bezug auf allgemeine Zufriedenheit als in der Rebif[®]-Gruppe. In der MMRM-Analyse betrug die LS-Mittelwertsdifferenz im Vergleich zur Rebif[®]-Gruppe in der 14 mg Teriflunomid-Gruppe 7,843 (95 % KI: [1,458; 14,23]). Der Unterschied ist nicht nur statistisch signifikant, sondern auch von klinischer Relevanz.

Eine neuere Untersuchung, die das Ausmaß der Patientenzufriedenheit (gemessen unter anderem mit TSQM) bei Patienten mit chronischen Erkrankungen (z. B. Typ-II-Diabetes oder Rheuma) mit der Therapieadhärenz in Beziehung setzt [64], zeigt die klinische Relevanz des erreichten Unterschiedes. In der genannten Untersuchung korreliert ein mittlerer TSQM-Wert in der allgemeinen Zufriedenheit von 67,72 ($\pm 20,03$) mit nur minimalen Problemen in der Adhärenz, ein entsprechender Wert von 63,27 ($\pm 20,94$) jedoch bereits mit einer schlechten Therapieadhärenz. Auch wenn sich die Absolutwerte dieser Untersuchung nicht direkt auf das Studienergebnis von TENERE übertragen lassen, kann dementsprechend der in TENERE erreichte Unterschied von 7,843 Punkten gerade in Bezug auf eine Dauertherapie ausgesprochen bedeutsam sein.

Zufriedenheit in Bezug auf die Wirksamkeit

Nach 48 Wochen betrug der mittlere Score im Bereich Wirksamkeit 67,16 in der 14 mg Teriflunomid-Gruppe und 63,45 in der Rebif[®]-Gruppe. In der MMRM-Analyse betrug die LS-Mittelwertsdifferenz im Vergleich zur Rebif[®]-Gruppe in der 14 mg Teriflunomid-Gruppe 3,835 (95% KI: [-3,082; 10,75]) mit einem zugehörigen p-Wert von 0,2759 (nicht signifikant).

Zufriedenheit in Bezug auf die Verträglichkeit

Nach 48 Wochen betrug der mittlere Score im Bereich Verträglichkeit 95,09 in der 14 mg Teriflunomid-Gruppe und 75,27 in der Rebif[®]-Gruppe. In der MMRM-Analyse betrug die LS-Mittelwertsdifferenz im Vergleich zur Rebif[®]-Gruppe in der 14 mg Teriflunomid-Gruppe

21,77 (95% KI: [16,48; 27,07]) mit einem korrespondierenden p-Wert <0,0001. Dies zeigt eine signifikant größere Zufriedenheit der Patienten mit der Verträglichkeit von Teriflunomid in einem relevanten Ausmaß.

Der TSQM-Score bewegt sich in einem Bereich zwischen 0 und 100, eine Differenz von über 20 kann als beträchtlicher Unterschied gewertet werden. Für eine Einschätzung der Relevanz des Unterschiedes wird auf die Validierung des TSQM verwiesen [5]. Bei der Auswertung der Patientenzufriedenheit einer oralen Medikation war ein mittlerer Wert in Bezug auf die Verträglichkeit von 75,8 ($\pm 19,9$) in dieser Auswertung mit der Einschätzung eines schlechten Gesundheitszustandes korreliert, ein entsprechender Wert von 91,8 ($\pm 17,5$) mit der Wahrnehmung eines exzellenten Gesundheitszustandes. Diese Werte lassen sich nicht als Absolutwerte auf das Patientenkollektiv der TENERE-Studie übertragen, vermitteln jedoch sehr eindrucksvoll die klinische Relevanz des erreichten Unterschiedes.

Zufriedenheit in Bezug auf die Anwenderfreundlichkeit

Nach 48 Wochen betrug der mittlere Score im Bereich Anwenderfreundlichkeit 90,78 in der 14 mg Teriflunomid-Gruppe und 63,17 in der Rebif[®]-Gruppe. In der MMRM-Analyse betrug die LS-Mittelwertsdifferenz im Vergleich zur Rebif[®]-Gruppe in der 14 mg Teriflunomid-Gruppe 27,96 (95% KI: [23,35; 32,56]) mit einem korrespondierenden p-Wert <0,0001. Dies zeigt eine signifikant größere Zufriedenheit der Patienten mit der Anwenderfreundlichkeit von Teriflunomid in einem relevanten Ausmaß.

Der TSQM-Score bewegt sich in einem Bereich zwischen 0 und 100, eine Differenz von über 20 kann als beträchtlicher Unterschied gewertet werden. Die überzeugende Zufriedenheit der Patienten mit der Anwenderfreundlichkeit lässt sich mit hoher Wahrscheinlichkeit auf die Darreichungsform (Filmpille) zurückführen.

Patientenzufriedenheit - Fazit

Insgesamt waren die Patienten mit der Teriflunomid-Behandlung zufriedener als mit der Behandlung mit der zweckmäßigen Vergleichstherapie Interferon beta-1a, 44 µg, subkutan (Rebif[®]). Besonders deutlich war die größere Zufriedenheit in den Bereichen Verträglichkeit und Anwenderfreundlichkeit, im Bereich Wirksamkeit zeigte sich kein signifikanter Vorteil. Dieses Ergebnis ist konsistent mit dem Bild, das sich aus der Auswertung sämtlicher Endpunkte für den Vergleich des zu bewertenden Arzneimittels mit seiner zweckmäßigen Vergleichstherapie abzeichnet: Eine vergleichbare Wirksamkeit der beiden Wirkstoffe und ein anderes Nebenwirkungsprofil kombiniert mit einer oraler Darreichungsform, was insgesamt zu einer größeren Zufriedenheit bei den Teriflunomid-Patienten führt. Die signifikant größere Zufriedenheit der Patienten mit der Therapie kann zu einer besseren Therapieadhärenz führen, wie unter dem Punkt „Allgemeine Zufriedenheit“ ausgeführt. Dementsprechend wird bezogen auf die Patientenzufriedenheit ein beträchtlicher Zusatznutzen für 14 mg Teriflunomid abgeleitet.

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Die Patientenzufriedenheit ist lediglich in der TENERE Studie erhoben worden. Eine Metaanalyse kann nicht durchgeführt werden.

4.3.1.3.7 Lebensqualität (SF-36)– RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-76: Operationalisierung von SF-36 (Lebensqualität)

Studie	Operationalisierung
<Studie 1>	
TEM SO	Lebensqualität erfasst mit SF-36 Der SF-36 (=Short Form 36) ist ein krankheitsübergeifendes Messinstrument zur Erfassung der gesundheitsbezogenen Lebensqualität. Der Fragebogen umfasst 36 Fragen, womit 8

	<p>Dimensionen erfasst werden, wobei jede Skala einen Wertebereich von 0 bis 100 umfasst. Zwei zusammenfassende Scores bewerten jeweils 4 Maße körperlicher Gesundheit (körperliche Funktionsfähigkeit, körperliche Rollenfunktion, körperlicher Schmerz und allgemeiner Gesundheitszustand) und geistiger Gesundheit (Vitalität, soziale Funktionsfähigkeit, emotionale Rollenfunktion, und psychische Gesundheit). Zusätzlich dazu wird mit einer Frage eine berichtete Änderung des Gesundheitszustandes erfasst.</p> <p>Der SF-36 ist das am weitesten eingesetzte Maß gesundheitsbezogener Lebensqualität in RRMS und kann in der Regel in 5 bis 10 Minuten ausgefüllt werden.</p> <p>Der SF-36 wurde zu folgenden Zeitpunkten ausgefüllt: Screening, Wochen 20, 42, 66, 90 und zur Abschlussuntersuchung (Woche 108).</p>
TOWER	<p>Lebensqualität erfasst mit SF-36 s. TEMSO</p> <p>Der SF-36 wurde zu folgenden Zeitpunkten ausgefüllt: Screening, Woche 12, Woche 24, anschließend alle 24 Wochen und zum Zeitpunkt der Abschlussuntersuchung.</p>

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-77: Bewertung des Verzerrungspotenzials für SF-36 (Lebensqualität) in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TEMSO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotenzial wurde für beide Studien TEMSO und TOWER auf Studienebene als niedrig eingestuft. Es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotenzial erhöhende Einflussfaktoren.

Der SF-36 ist ein generisches Messinstrumente (Fragebogen) zur Erhebung der gesundheitsbezogenen Lebensqualität, der von dem Patienten selbst ausgefüllt wird, es handelt sich um einen patientenberichteten Endpunkt (PRO = patient reported outcome). Damit ist der Patient als Endpunkterheber zu werten. In den beiden placebokontrollierten Studien TEMSO und TOWER waren die Patienten bezüglich ihrer Behandlung verblindet und das Verzerrungspotential insgesamt ist niedrig.

Stellen Sie die Ergebnisse für den Endpunkt „Lebensqualität erfasst mit SF-36“ für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-78: Ergebnisse für Lebensqualität SF-36 aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

Studie	Patienten- zahl (ITT ^a)	Baseline-Wert Mittelwert (Standard- abweichung)	Anzahl (Woche 108 bzw. Woche 48) ^b	Wert in W 108 bzw. 48 ^b Mittelwert (Standard- abweichung)	Change from Baseline MMRM ^c	
					(W 108/ W 48) ^b LS-Mean ^d (Standard- fehler)	LS-Mean- Differenz ^d im Vergleich zu Placebo (Standard- fehler) 95 % KI [] p-Wert
Gesamtwert in der physischen Dimension (Physical Health Summary Score)						
TEMSO	T14 (n= 358) 342 in W0	44,296 (9,415)	258 (246 in MMRM)	44,638 (9,394)	-0,855 (0,476)	0,387 (0,645) [-0,879;1,653] p = 0,549
	Pb (n = 363) 347 in W0	44,160 (9,546)	246 (237 in MMRM)	45,040 (10,488)	-1,242 (0,478)	
TOWER	T14 (n= 370) 352 in W0	43,396 (9,521)	265 (256 in MMRM)	44,394 (9,662)	-0,105 (0,418)	0,977 (0,561) [-0,123; 2,077] p = 0,0817
	Pb (n = 388) 372 in W0	43,612 (9,750)	287 (277 in MMRM)	43,503 (10,358)	-1,082 (0,405)	
Gesamtwert in der psychischen Dimension (Mental Health Summary Score)						
TEMSO	T14 (n= 358) 342 in W0	44.104 (11.487)	258 (246 in MMRM)	44,282 (11,598)	-0,756 (0,603)	0,655 (0,821) [-0,957;2,266] p = 0,4255
	Pb (n = 363) 347 in W0	43,428 (11,541)	246 (237 in MMRM)	43,567 (11,485)	-1,410 (0,608)	

TOWER	T14 (n= 370) 352 in W0	43,049 (11,337)	265 (256 in MMRM)	42,896 (12,131)	-1,434 (0,606)	1,479 (0,816) [-0,122; 3,080] p = 0,0702
	Pb (n = 388) 372 in W0	42,345 (11,785)	287 (277 in MMRM)	40,895 (12,236)	-2,913 (0,586)	

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bdie MMRM-Analyse der Änderung in Bezug zum Baseline-Wert wurde in der TEMSO Studie mit dem Erhebungszeitpunkt Woche 108 durchgeführt, in der Studie TOWER bezieht es sich auf den Zeitpunkt Woche 48

^cMMRM = Mixed-effects Model with Repeated Measures, realisiert in SAS mit Proc Mixed. Feste Effekte in dem Modell waren Behandlung, Baseline EDSS-Stratum ($\leq 3,5$ / $> 3,5$), Region, Visite, Behandlung \times Visite-Wechselwirkung, Baseline-Wert und Behandlung \times Visite-Wechselwirkung. Für die Modellierung der within-patient Fehlerterme wurde eine nicht-strukturierte Korrelationsmatrix zu Grunde gelegt. Die Modellparameter wurden nach der restricted Maximum-Likelihood-Methode unter Verwendung des Newton-Raphson-Algorithmus geschätzt. Die Freiheitsgrade des Nenners wurden dabei nach der Approximation von Satterthwaite geschätzt.

„Anzahl“ steht in diesem Zusammenhang für die Anzahl real vorhandener Datensätze nach 108 bzw. 48 Wochen.

^dLS-Means sind die sich aus dem MMRM-Modell ergebenden adjustierten Schätzer

TEMSO

Mit dem SF-36 werden die zwei Bereiche der physischen und der psychischen Gesundheit getrennt voneinander in zwei zusammenfassenden Werten abgebildet. In keinem der beiden Bereiche konnte ein statistisch signifikanter Unterschied in den verschiedenen Behandlungsgruppen nachgewiesen werden. Sowohl bei dem „Gesamtwert physischer Dimension“ als auch bei dem „Gesamtwert psychischer Dimension“ handelt es sich um zusammenfassende Bewertungen der physischen und psychischen Gesundheit, die auch einzeln in aufgeteilten Dimensionen erfasst sind. Für Woche 108 wurden in einer analogen MMRM-Analyse die Dimensionen körperliche Funktionsfähigkeit, körperliche Rollenfunktion, körperlicher Schmerz, allgemeiner Gesundheitszustand, Vitalität, soziale Funktionsfähigkeit, emotionale Rollenfunktion und psychische Gesundheit analysiert. In keinem dieser Bereiche konnte eine signifikante Verbesserung gegenüber Placebo festgestellt werden, die entsprechenden Auswertungen finden sich in Appendix 14.2.6.7 des klinischen Studienberichtes [32].

TOWER

Mit dem SF-36 werden die zwei Bereiche der physischen und der psychischen Gesundheit getrennt voneinander in zwei zusammenfassenden Werten abgebildet. Auch in der TOWER-Studie konnte in keinem der beiden Bereiche eine statistisch signifikante Überlegenheit von 14 mg Teriflunomid im Vergleich zu Placebo gezeigt werden. Sowohl bei dem „Gesamtwert physischer Dimension“ als auch bei dem „Gesamtwert psychischer Dimension“ handelt es

sich um zusammenfassende Bewertungen der physischen und psychischen Gesundheit, die auch einzeln in aufgeteilten Dimensionen erfasst sind. Für Woche 108 wurden in einer analogen MMRM-Analyse die Dimensionen erfasst. Für viele der SF-36 Einzelwerte (körperliche Funktionsfähigkeit, körperliche Rollenfunktion, körperlicher Schmerz, allgemeiner Gesundheitszustand, Vitalität, soziale Funktionsfähigkeit, emotionale Rollenfunktion und psychische Gesundheit) ließ sich eine tendenzielle Überlegenheit der 14 mg Teriflunomid-Behandlungsgruppe gegenüber Placebo feststellen, ein signifikantes Ergebnis wurde in den Bereichen körperlicher Schmerz und psychische Gesundheit erreicht, die Auswertungen finden sich in Appendix 16.2.6.5 des entsprechenden Studienberichtes [31].

4.3.1.3.8 Lebensqualität (EQ-5D) – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-79: Operationalisierung von Lebensqualität

Studie	Operationalisierung
<Studie 1>	

TEMSO	<p>Lebensqualität erfasst mit EQ-5D</p> <p>Der EQ-5D (EuroQoL-5D) ist ein standardisierter Fragebogen zur Erfassung der gesundheitsbezogenen Lebensqualität, der von der EuroQoL-Gruppe entwickelt worden ist, um ein einfaches, generisches Gesundheitsmessinstrument für die klinische und ökonomische Bewertung zur Verfügung zu haben.</p> <p>Der Fragebogen besteht aus dem deskriptiven System des EQ-5D und der VAS (visuelle Analogskala, visual analogue scale). Das deskriptive System umfasst 5 Dimensionen: Mobilität, Selbstversorgung, alltägliche Aktivitäten, Schmerz/Unwohlsein und Angst/Depressionen. Das 5-Dimensionen-3-Level-System wurde übertragen in einen einzigen Lebensqualitätsindex (single index utility score) (durch ein Regressionsmodell), der sich in Bereichen von -0,594 (korrespondiert mit Level 3 (schwere Probleme) in den Bereichen Mobilität, Selbstversorgung, alltägliche Aktivitäten, Schmerz/Unwohlsein und Angst/Depressionen) und 1,0 (korrespondiert mit voller Gesundheit (Level 1, keine Probleme) bewegen kann. Die VAS bildet die Selbsteinschätzung der Befragten auf einer vertikalen visuellen Analogskala ab; die Ankerpunkte sind dabei 100 (besten vorstellbarer Gesundheitszustand) und 0 (schlechtesten vorstellbarer Gesundheitszustand). Damit ist der VAS ein quantitatives Maß für gesundheitsbezogene Endpunkte, eingeschätzt durch die einzelnen Befragten.</p> <p>EQ-5D Score wurde als Baseline-Wert erfasst (Woche 0), alle 12 Wochen während der Studie und im Falle eines Schubes.</p>
--------------	---

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-80: Bewertung des Verzerrungspotenzials für EQ-5D (Lebensqualität) in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TEMSO	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wurde für die Studie TEMSO auf Studienebene als niedrig eingestuft. Es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren.

Der EQ-5D ist ein generisches Messinstrumente (Fragebogen) zur Erhebung der gesundheitsbezogenen Lebensqualität, der von dem Patienten selbst ausgefüllt wird, es handelt sich um einen patientenberichteten Endpunkt (PRO = patient reported outcome). Damit ist der Patient als Endpunkterheber zu werten. In den beiden placebokontrollierten Studien TEMSO und TOWER waren die Patienten bezüglich ihrer Behandlung verblindet und das Verzerrungspotential insgesamt ist niedrig.

Stellen Sie die Ergebnisse für den Endpunkt „Lebensqualität erfasst mit EQ-5D“ für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-81: Ergebnisse für Lebensqualität EQ-5D aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

Studie	Patienten- zahl (ITT ^a)	Baseline- Wert	Anzahl ^b (Woche 108)	Wert in W 108	Change from Baseline MMRM (W 108) ^b	
		Mittelwert (Standard- abweichung)		Mittelwert (Standard- abweichung)	LS-Mean ^c (Standard- fehler)	LS-Mean- Differenz ^c im Vergleich zu Placebo (Standard- fehler) 95 % KI [] p-Wert
Change from Baseline im Lebensqualitätsindex (single index utility score)						
TEMSO	T14 (n= 358)	0,732 (0,223)	262 (258 in MMRM)	0,730 (0,223)	-0,035 (0,012)	0,004 (0,016) [-0,028; 0,036] p = 0,8138
	Pb (n = 363)	0,713 (0,225)		0,739 (0,226)		
Change from Baseline im Wert auf der visuellen Analogskala (EQ-5D, visual analogue scale)						
TEMSO	T14 (n= 358)	71,798 (18,080)	260 (249 in MMRM)	69,646 (20,655)	-5,613 (1,264)	-2,064 (1,735) [-5,469; 1,340] p = 0,2343
	Pb (n = 363)	70,819 (18,077)		72,607 (27,012)		

^as. Definition der ITT-Population unter 4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien

^bMMRM = Mixed-effects Model with Repeated Measures, realisiert in SAS mit Proc Mixed. Feste Effekte in dem Modell waren Behandlung, Baseline EDSS-Stratum ($\leq 3,5$ / $> 3,5$), Region, Visite, Behandlung \times Visite-Wechselwirkung, Baseline-Wert und Behandlung \times Visite-Wechselwirkung. Für die Modellierung der within-patient Fehlerterme wurde eine nicht-strukturierte Korrelationsmatrix zu Grunde gelegt. Die Modellparameter wurden nach der restricted Maximum-Likelihood-Methode unter Verwendung des Newton-Raphson-

Algorithmus geschätzt. Die Freiheitsgrade des Nenners wurden dabei nach der Approximation von Satterthwaite geschätzt.^cLS-Means sind die sich aus dem MMRM-Modell ergebenden adjustierten Schätzer.

„Anzahl“ steht in diesem Zusammenhang für die Anzahl real vorhandener Datensätze nach 108 Wochen.

TEMSO

Der Lebensqualitätsindex ist die Übertragung der einzelnen Werte der Patienten in den 5 Dimensionen Mobilität, Selbstversorgung, gewöhnliche Aktivitäten, Schmerz/Unwohlsein und Angst/Depressionen in einen einzigen Wert (s. Tabelle 4-76). Weder im Lebensqualitätsindex noch bei der Bewertung der Patienten auf der visuellen Analogskala konnten zwischen den Behandlungsgruppen nach 108 Wochen statistisch signifikante Unterschiede gezeigt werden. Die Einzelwerte finden sich in Appendix 14.2.6.8 des entsprechenden Studienberichtes [32].

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Eine Metaanalyse zu den berichteten LS-Means des SF-36 aus der TEMSO- und der TOWER-Studie wird nicht durchgeführt, da sich die Analysen auf zu unterschiedliche Zeitpunkte beziehen (48 bzw. 108 Wochen).

Der Fragebogen EQ-5D wurde nur in der TEMSO-Studie zur Erfassung der Lebensqualität eingesetzt, eine Metaanalyse ist nicht möglich.

4.3.1.3.9 Mortalität – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe

- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-82: Operationalisierung von „Während der Studie verstorbene Patienten“ (Mortalität)

Studie	Operationalisierung
TEMSO	Mortalität wurde bestimmt als die Anzahl an Todesfällen die während der Behandlung aufgetreten sind
TOWER	Mortalität wurde bestimmt als die Anzahl an Todesfällen die während der Behandlung aufgetreten sind
TENERE	Mortalität wurde bestimmt als die Anzahl an Todesfällen die während der Behandlung aufgetreten sind

Für die Darstellung der Mortalität wird- abweichend von den entsprechenden Darstellungen der Effektivitäts-Endpunkte- nicht die ITT-Population, sondern die Safety-Population Tabelle 4-21 zur Auswertung herangezogen.

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-83: Bewertung des Verzerrungspotenzials für Mortalität in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TEMPO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
TENERE	niedrig	nein	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wurde für die Studien TEMPO, TOWER und TENERE auf Studienebene als niedrig eingestuft. Es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren.

Es gibt keine endpunktspezifischen Faktoren, die das Verzerrungspotential erhöhend beeinflussen. Trotz der Tatsache, dass die Patienten in der TENERE-Studie nicht verblindet waren, ist dies nicht als ein das Verzerrungspotential erhöhender Faktor für den Endpunkt „Während der Studie verstorbene Patienten (Mortalität)“ zu werten.

Stellen Sie die Ergebnisse für den Endpunkt Mortalität für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-84: Ergebnisse für Mortalität aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

Studie	Patientenzahl (Safety-Population) ^a	Anzahl Todesfälle (prozentualer Anteil)	Relative Risk (RR) ^b p-Wert ^b 95% KI []	Odds Ratio (OR) ^b p-Wert ^b 95% KI []
TEMSO	T14 (n=358)	0		
	Pb (n=360)	0		
TOWER	T14 (n=371)	2 (0,5%)	2,076 p = 0,5503 [0,189; 22,79]	2,081 p = 0,5502 [0,188; 23,05]
	Pb (n=385)	1 (0,3%)		
TENERE ^c	T14 (n=110)	0		
	Rb (n=101)	0		

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bp-Werte und 95 % KI asymptotisch berechnet, für relative Risiken unter Verwendung der Delta-Methode

^c99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

TEMSO

Im Rahmen der TEMSO-Studie sind keine Todesfälle aufgetreten.

TOWER

Während der TOWER-Studie sind insgesamt 4 Todesfälle aufgetreten. Ein Patient der Placebo-Gruppe starb an einer Atemwegsinfektion, in der 14 mg Teriflunomid-Gruppe verstarb ein Patient durch Suizid und einer als Folge einer durch einen Gram-negativen Erreger verursachten Sepsis. In der 7 mg Teriflunomid-Gruppe verstarb ein Patient an den Folgen eines Motorrad-Unfalls. Es besteht kein Hinweis darauf, dass einer dieser Todesfälle mit der Medikation in Zusammenhang steht.

TENERE

Im Rahmen der TENERE-Studie sind keine Todesfälle aufgetreten.

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die

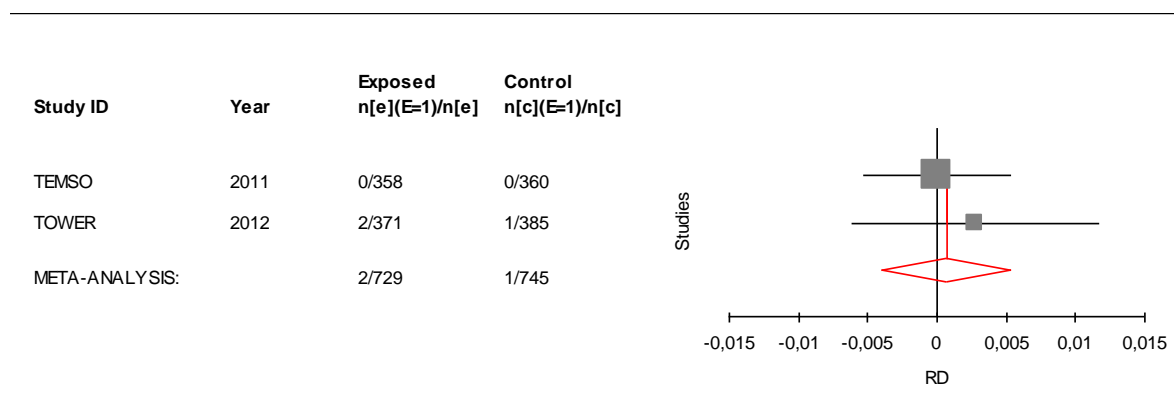
Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Todesfälle sind lediglich in einer Studie aufgetreten (TOWER). Eine metaanalytische Zusammenfassung findet sich als Bestandteil des indirekten Vergleichs in Modul 4.3.2.

Tabelle 4-85: Metaanalyse-Ergebnisse für Mortalität aus TEMSO und TOWER – für die gesamte Studienpopulation

Analyse	Komparatoren	Risiko-Differenz (RD) 95 % KI p(H ₀ : RD=0)
Metaanalyse 1 (TEMSO & TOWER)	Teriflunomid 14 mg versus Placebo	0,0007 [-0,0039; 0,0054] p = 0,7539 I ² =0%, p _Q =0,6032

Abbildung 4-14: Forest-Plot der Metaanalyse von TEMSO und TOWER für Mortalität



4.3.1.3.10 Unerwünschte Ereignisse (Gesamtraten AEs, SAEs, Behandlungsabbrüche) – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe

- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-86: Operationalisierung von Gesamtraten unerwünschter Ereignisse (AEs, SAEs, Behandlungsabbrüche)

Studie	Operationalisierung
TEMSO	<ul style="list-style-type: none"> Anzahl an Patienten mit unerwünschten Ereignissen während der Studienzeit Anzahl an Patienten mit schwerwiegenden unerwünschten Ereignissen während der Studienzeit Anzahl an Patienten, die die Behandlung aufgrund von unerwünschten Ereignissen abgebrochen haben
TOWER	<ul style="list-style-type: none"> Anzahl an Patienten mit unerwünschten Ereignissen während der Studienzeit Anzahl an Patienten mit schwerwiegenden unerwünschten Ereignissen während der Studienzeit Anzahl an Patienten, die die Behandlung aufgrund von unerwünschten Ereignissen abgebrochen haben
TENERE	<ul style="list-style-type: none"> Anzahl an Patienten mit unerwünschten Ereignissen während der Studienzeit Anzahl an Patienten mit schwerwiegenden unerwünschten Ereignissen während der Studienzeit Anzahl an Patienten, die die Behandlung aufgrund von unerwünschten Ereignissen abgebrochen haben
Für alle Studien geltende Abkürzungen und Definitionen	
Adverse event (AE)	<ul style="list-style-type: none"> Ein unerwünschtes Ereignis (AE = adverse event) wurde definiert als jedes unerwünschte medizinische Ereignis bei einem behandelten Patienten, wobei das Ereignis nicht notwendigerweise durch die Behandlung verursacht worden sein muss. Effektivitäts-Endpunkte, so wie sie im Rahmen der Studie definiert worden sind, wurden nicht als unerwünschte Ereignisse eingestuft. Somit wurden MS-Schübe als Effektivitätsendpunkte und nicht als unerwünschte oder schwerwiegende unerwünschte Ereignisse (SAE = serious adverse events) gewertet. Mit einem Schub zusammenhängende Symptome oder Ereignisse, die initial für einen Schub gehalten wurden, aber nicht als solcher bestätigt werden konnten, wurden jedoch als AE/SAE gewertet.
Treatment emergent adverse events (TEAE)	<ul style="list-style-type: none"> TEAE steht für „Treatment emergent adverse events“ und ist definiert als unerwünschte Ereignisse, die während der Behandlung aufgetreten sind, es wird kein kausaler Zusammenhang mit der Intervention impliziert. Alle AE sind als TEAE angegeben. <p>Die Behandlungsdauer ist definiert als Zeitpunkt von der erstmaligen Einnahme der Studienmedikation bis zu 6 Wochen (TOWER) bzw. 16 Wochen (TEMSO und TENERE) nach Einnahme der letzten Dosis der Studienmedikation.</p>
Schwerwiegendes unerwünschtes Ereignis (SAE = serious adverse event)	<p>Ein schwerwiegendes unerwünschtes Ereignis (SAE = serious adverse event) war jedes Ereignis, das</p> <ul style="list-style-type: none"> tödlich ist, lebensbedrohend ist (zu beachten: „lebensbedrohend“ bezieht sich auf ein Ereignis, in dem der Patient zur Zeit des Ereignisses vom Tod bedroht war, nicht aber auf ein Ereignis, das hypothetisch den Tod verursacht haben könnte, wenn es schwerer gewesen wäre), eine stationäre Behandlung oder deren Verlängerung erforderlich macht, zu bleibender oder schwerwiegender Behinderung oder Invalidität führt, eine kongenitale Anomalie oder einen Geburtsfehler zur Folge hat oder ein medizinisch relevantes Ereignis darstellt.

AESI	AESI = adverse events of special interest, ein unerwünschtes Ereignis ist dann von besonderem Interesse, wenn ein kausaler Zusammenhang mit einer Intervention vermutet werden kann oder wenn das Ereignis mit dem Krankheitsbild oder dem Wirkmechanismus (z. B. Immunmodulation) in Verbindung gebracht werden kann (im jeweiligen Studienprotokoll definiert)
Vordefinierte AEs und abnorme Laborwerte	<p>Für die folgenden Ereignisse, definiert als vordefinierte AEs und abnorme Laborwerte, folgte der Untersucher dem beschleunigten Meldeverfahren, um das Monitoring des Sponsors und die unabhängige DMC (Data Monitoring Committee)-Bewertung zu ermöglichen - selbst dann wenn die Kriterien für „schwerwiegend“ nicht erfüllt wurden:</p> <ul style="list-style-type: none"> • Bestätigte SGPT/ALT-Erhöhung über das 2fache der Obergrenze des Normwertes (ULN = upper limit of normal) • Irgendein Ereignis einer Erhöhung auf einen Wert über das 3fache des ULN der SGPT/ALT (gemeldet als schwerwiegendes unerwünschtes Ereignis nach dem Kriterium der medizinischen Wichtigkeit) • Irgendein Ereignis einer Erhöhung auf einen Wert über das 8fache des ULN der SGPT/ALT (gemeldet als schwerwiegendes unerwünschtes Ereignis nach dem Kriterium der medizinischen Wichtigkeit) • Depressive Störung, die durch den Untersucher als klinisch relevant eingestuft wurde • Irgendein Auftreten einer Neutrophilenzahl von weniger als 1000 Zellen/μL • Durch einen wiederholten Test (so früh wie möglich) bestätigter Wert der Serum Amylase oder Lipase, der das 2fache des ULN überstiegen hat • Ultraschall-Untersuchungsbericht mit pankreatischen Abnormalitäten • Pulmonale Symptome sowie Beginn oder Verschlimmerung eines Hustens oder Dyspnoe, die eine Lungenuntersuchung notwendig machten und mit einer pulmonalen Toxizität konsistent waren • Durch Untersuchung der Nervenleitung bestätigte periphere Neuropathie
Schwangerschaft	<ul style="list-style-type: none"> • Schwangerschaft einer weiblichen Patientin wurde als AE gewertet. Als SAE wurde sie nur eingestuft, wenn die SAE-Kriterien erfüllt worden sind. Im Falle einer Schwangerschaft wurde die Studienmedikation abgesetzt.

Für die Darstellung der unerwünschten Ereignisse wird- abweichend von den entsprechenden Darstellungen der Effektivitäts-Endpunkte- nicht die ITT-Population, sondern die Safety-Population zur Auswertung herangezogen, s. Tabelle 4-87.

Die Safety-Population ist in den Studien wie folgt definiert:

Tabelle 4-87: Definitionen der Safety-Population

Studie	Population	Definition	Auswertung
TEM SO	Safety-Population	<p>Die Safety-Population ist definiert als alle randomisierten Patienten, die der Studienmedikation ausgesetzt waren, ungeachtet der verabreichten Menge. Sollte ein Patient eine von der Randomisierung abweichende Behandlung bekommen, so erfolgte die Auswertung nach der tatsächlich erfolgten Behandlung.</p> <p>Randomisierte Patienten, bei denen eine tatsächliche Einnahme der Studienmedikation nicht</p>	<p>Mortalität und Verträglichkeitsendpunkte:</p> <ul style="list-style-type: none"> • Gesamtrate unerwünschter Ereignisse • Gesamtrate schwerwiegender unerwünschter

		<p>nachgewiesen werden konnte, wurden in die Safety-Population eingeschlossen.</p> <p>Patienten, die behandelt, aber nicht randomisiert wurden, sind nicht Teil der Safety-Population.</p> <p>Patienten, die im Laufe der Studie verschiedene Studienmedikationen erhalten haben, sind der niedrigeren erhaltenen Dosis zugeordnet.</p>	<p>Ereignisse</p> <ul style="list-style-type: none"> • Unerwünschte Ereignisse, die zum Behandlungsabbruch führten • Unerwünschte Ereignisse von besonderem Interesse
TOWER	Safety-Population	<p>Die Safety-Population war eine Teilmenge aller randomisierten Patienten, die mindestens eine Dosis an Studienmedikation erhalten haben. Sollte ein Patient eine von der Randomisierung abweichende Behandlung bekommen, so erfolgte die Auswertung nach der tatsächlich erfolgten Behandlung.</p> <p>Randomisierte Patienten, bei denen eine tatsächliche Einnahme der Studienmedikation nicht nachgewiesen werden konnte, wurden in die Safety-Population eingeschlossen.</p> <p>Patienten, die behandelt, aber nicht randomisiert wurden, sind nicht Teil der Safety-Population.</p> <p>Patienten, die im Laufe der Studie verschiedene Studienmedikationen erhalten haben, sind der niedrigeren erhaltenen Dosis zugeordnet.</p>	<p>Mortalität und Verträglichkeitsendpunkte:</p> <ul style="list-style-type: none"> • Gesamtrate unerwünschter Ereignisse • Gesamtrate schwerwiegender unerwünschter Ereignisse • Unerwünschte Ereignisse, die zum Behandlungsabbruch führten • Unerwünschte Ereignisse von besonderem Interesse
TENERE	Safety-Population	<p>Die Safety-Population ist definiert als alle randomisierten Patienten, die der Studienmedikation ausgesetzt waren, ungeachtet der verabreichten Menge.</p> <p>Randomisierte Patienten, bei denen eine tatsächliche Einnahme der Studienmedikation nicht nachgewiesen werden konnte, wurden in die Safety-Population eingeschlossen.</p>	<p>Mortalität und Verträglichkeitsendpunkte:</p> <ul style="list-style-type: none"> • Gesamtrate unerwünschter Ereignisse • Gesamtrate schwerwiegender unerwünschter Ereignisse • Unerwünschte Ereignisse, die zum Behandlungsabbruch führten • Unerwünschte Ereignisse von besonderem Interesse

Die Auswertung der unerwünschten Ereignisse basieren auf den relativen Risiken. Zusätzlich sind die Ereignisraten auch anhand der Odds Ratios ausgewertet und in den Tabellen angegeben, eine ausführliche Diskussion der Ergebnisse basierend auf den Odds Ratios wäre jedoch nur in Fällen angegeben, in denen diese zu einem von der auf relativen Risiken basierenden Analyse abweichenden Ergebnis führt.

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-88: Bewertung des Verzerrungspotenzials für Gesamtraten unerwünschter Ereignisse (AEs, SAEs, Behandlungsabbrüche) in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TEMPO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
TENERE	niedrig	nein	ja	nein	nein	niedrig ^a

^aIn der TENERE-Studie sind die Patienten und der behandelnde Neurologe bezüglich der Medikation nicht verblindet. Jedoch werden die Gesamtrate unerwünschter Ereignisse, Gesamtrate schwerwiegender unerwünschter Ereignisse und die meisten spezifischen unerwünschten Ereignisse entweder klar definiert und mit einer objektivierbaren Komponente eingestuft oder aber es handelt sich um die Erfassung von Laborparametern, die in einem zentralen Labor erfasst und nicht durch eine subjektive Einschätzung der Patienten beeinflusst werden können.

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wurde für die Studien TEMPO, TOWER und TENERE auf Studienebene als niedrig eingestuft. Es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren.

In den Studien TEMPO und TOWER gibt es keine Endpunkt-spezifischen Faktoren, die das Verzerrungspotential erhöhend beeinflussen, die Endpunkterheber waren verblindet und das ITT-Prinzip wurde adäquat umgesetzt (Analyse der Safety-Population).

Die Durchführung der Studie und auch die Analyse der unerwünschten Ereignisse erfolgten nach GCP-Standards. Die Analyse der Safety-Population entspricht der methodengerechten Analyse sicherheitsrelevanter Ereignisse. Die Ereignisse sind nach MedDRA[®] (Medical Dictionary for Regulatory Activities) 13.0 oder höher kodiert.

In der TENERE-Studie waren die Patienten bezüglich ihrer Medikation nicht verblindet. Dies kann, je nach Ausprägung der unerwünschten Ereignisse, das Ergebnis relevant verzerren. Der Großteil der unerwünschten Ereignisse ist eindeutig definiert und beinhaltet eine

objektivierbare Komponente (s. Fußnote Tabelle 4-88). Dadurch sind die Ergebnisse im Regelfall nicht verzerrt, eine mögliche Verzerrung bei einzelnen unerwünschten Ereignissen betrifft nicht die Gesamtraten unerwünschter Ereignisse, schwerwiegender unerwünschter Ereignisse und unerwünschter Ereignisse, die zum Behandlungsabbruch führten, sondern lediglich einzelne unerwünschte Ereignisse von besonderem Interesse und wird dort diskutiert.

Stellen Sie die Ergebnisse für den Endpunkt unerwünschte Ereignisse für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-89: Betrachteter Studienzeitraum, in dem die unerwünschten Ereignisse beobachtet werden konnten

Studie	Patientenzahl (Safety-Population) ^a	Kumulative Dauer der Behandlung (Patientenjahre)	Behandlungsdauer (Tage)
			Arithmetischer Mittelwert (Standardabweichung)
TEMSO	T14 (n=358)	614,21	626,65 (242,49)
	Pb (n=360)	622,13	631,21 (228,12)
TOWER	T14 (n=371)	576,18	567,25 (298,33)
	Pb (n=385)	601,64	570,78 (274,53)
TENERE	T14 (n=110)	130,83	434,43 (181,42)
	Rb (n=101)	112,04	405,18 (190,84)

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

4.3.1.3.10.1 Gesamtrate unerwünschter Ereignisse

Tabelle 4-90: Ergebnisse für die Gesamtrate unerwünschter Ereignisse aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

Gesamtrate unerwünschter Ereignisse					
Studie	Patientenzahl (Safety-Population) ^c	Anzahl Patienten mit dem entsprechenden Ereignis	Prozentualer Anteil Patienten mit dem entsprechenden Ereignis	Relative Risk (RR) ^a	Odds Ratio (OR) ^a
				p-Wert ^b 95% KI []	p-Wert ^b 95% KI []
TEMSO	T14 (n=358)	325	90,8 %	1,038 p = 0,1581 [0,986; 1,092]	1,407 p = 0,1591 [0,875; 2,263]
	Pb (n=360)	315	87,5 %		
TOWER	T14 (n=371)	320	86,3%	1,038 p = 0,2312	1,275 p = 0,2323

				[0,977; 1,103]	[0,856; 1,898]
	Pb (n=385)	320	83,1%		
TENERE ^d	T14 (n=110)	102	92,7 %	0,966 p = 0,2946 [0,904; 1,031]	0,526 p = 0,3064 [0,153; 1,802]
	Rb (n=101)	97	96,0 %		

^ap-Werte und 95 % Konfidenzintervalle wurden asymptotisch berechnet, für relative Risiken unter Verwendung der Delta-Methode

^bH₀: RR = 1 bzw. OR = 1

^cs. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^d99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

4.3.1.3.10.2 Gesamtrate schwerwiegender unerwünschter Ereignisse

Tabelle 4-91: Ergebnisse für die Gesamtrate schwerwiegender unerwünschter Ereignisse aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

Gesamtrate schwerwiegender unerwünschter Ereignisse					
	Patienten- zahl (Safety- Population) ^c	Anzahl Patienten mit dem entsprechenden Ereignis	Prozentualer Anteil Patienten mit dem entsprechenden Ereignis	Relative Risk (RR) ^a p-Wert ^b 95% KI []	Odds Ratio (OR) ^a p-Wert ^b 95% KI []
TEMSO	T14 (n=358)	57	15,9 %	1,246 p = 0,2309 [0,870; 1,786]	1,293 p = 0,2303 [0,850; 1,966]
	Pb (n=360)	46	12,8%		
TOWER	T14 (n=371)	44	11,9%	0,972 p = 0,8832 [0,661; 1,429]	0,968 p = 0,8831 [0,624; 1,500]
	Pb (n=385)	47	12,2%		
TENERE ^d	T14 (n=110)	6	5,5 %	0,787 p = 0,6568 [0,274; 2,264]	0,775 p = 0,6567 [0,251; 2,388]
	Rb (n=101)	7	6,9 %		

^ap-Werte und 95 % Konfidenzintervalle wurden asymptotisch berechnet, für relative Risiken unter Verwendung der Delta-Methode

^bH₀: RR = 1 bzw. OR = 1

^cs. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^d99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

4.3.1.3.10.3 Gesamtrate unerwünschter Ereignisse, die zum Behandlungsabbruch führten

Tabelle 4-92: Ergebnisse für die Gesamtrate unerwünschter Ereignisse, die zum Behandlungsabbruch führten, aus RCT mit dem zu bewertenden Arzneimittel – für die gesamte Studienpopulation

Unerwünschte Ereignisse, die zum Behandlungsabbruch führten					
	Patienten- zahl (Safety- Population) ^c	Anzahl Patienten mit dem entsprechenden Ereignis	Prozentualer Anteil Patienten mit mindestens einem entsprechenden Ereignis	Relative Risk (RR) ^a p-Wert ^b 95% KI []	Odds Ratio (OR) ^a p-Wert ^b 95% KI []
TEM SO	T14 (n=358)	39	10,9 %	1,352 p = 0,1963 [0,856; 2,137]	1,395 p = 0,1956 [0,842; 2,311]
	Pb (n=360)	29	8,1 %		
TOWER	T14 (n=371)	58	15,6%	2,508 p < 0,0001 [1,593; 3,948]	2,787 p < 0,0001 [1,692; 4,592]
	Pb (n=385)	24	6,2%		
TENERE ^d	T14 (n=110)	12	10,9 %	0,501 p = 0,0369 [0,262; 0,959]	0,440 p = 0,0349 [0,205; 0,943]
	Rb (n=101)	22	21,8 %		

^ap-Werte und 95 % Konfidenzintervalle wurden asymptotisch berechnet, für relative Risiken unter Verwendung der Delta-Methode

^bH₀: RR = 1 bzw. OR = 1

^cs. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^d99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

4.3.1.3.10.4 Diskussion der Gesamtraten unerwünschter Ereignisse

TEMSO

Die Gesamtrate an Patienten mit unerwünschten Ereignissen, die im Behandlungszeitraum aufgetreten sind, war mit 87,5 % in der Placebo-Gruppe und 90,8 % in der 14 mg Teriflunomid-Gruppe in den Behandlungsgruppen ähnlich, das entsprechende relative Risiko betrug 1,038 (95 % KI: [0,986; 1,092]). Ähnlich waren in beiden Gruppen auch die Raten schwerwiegende unerwünschte Ereignisse (relatives Risiko von 1,246 (95 % KI: [0,870; 1,786])), sowie unerwünschte Ereignisse, die zu einem Behandlungsabbruch führten (relatives Risiko von 1,352 (95 % KI: [0,856; 2,137])).

TOWER

Die Gesamtrate an Patienten mit unerwünschten Ereignissen (relatives Risiko von 1,038 (95 % KI: [0,977; 1,103]) oder schwerwiegenden unerwünschten Ereignissen (relatives Risiko von 0,972 (95 % KI: [0,661; 1,429])) unterschied sich wenig zwischen den Behandlungsgruppen. Der Anteil an Patienten mit unerwünschten Ereignissen, die zu einem Behandlungsabbruch führten, war unter Teriflunomid-Behandlung signifikant höher als in der Placebo-Gruppe. Das relative Risiko betrug 2,508 (95 % KI: [1,593; 3,948]), d. h. die Anzahl Patienten, die aufgrund von unerwünschten Ereignissen die Behandlung abbrachen, war in der Teriflunomid-Gruppe gegenüber Placebo auf das Zweieinhalbfache erhöht. Die häufigsten Gründe für einen Behandlungsabbruch aufgrund von unerwünschten Ereignissen waren Neutropenie, eine Erhöhung der ALT-Werte, Alopezie und gastrointestinale Beschwerden. Auf dieser Grundlage ist es für eine adäquate Bewertung des Verträglichkeitsprofils angezeigt, diese Ereignisse als von besonderem Interesse (AESI) auszuwerten.

TENERE

Insgesamt war der Anteil der Patienten mit mindestens einem TEAE in den unterschiedlichen Behandlungsarmen ähnlich (92,7 % und 96,0 % jeweils in der 14 mg Teriflunomid-Gruppe und der Rebif-Gruppe, was einem relativen Risiko von 0,966 (95 % KI: [0,904; 1,031]) entspricht). Auch die Rate schwerwiegender unerwünschter Ereignisse unterschied sich in der 14 mg Teriflunomid-Gruppe und der Rebif[®] 44 µg-Gruppe wenig (relatives Risiko von 0,787 (95 % KI: [0,274; 2,264])). Jedoch gab es in der Rebif[®] 44 µg-Gruppe signifikant mehr Patienten, die die Studienmedikation abbrachen, als in der 14 mg Teriflunomid-Gruppe, das relative Risiko betrug 0,501 (95 % KI: [0,262; 0,959]), d. h. das Risiko aufgrund von Nebenwirkungen die Therapie abzubrechen war unter Rebif[®] 44 µg doppelt so hoch wie unter 14 mg Teriflunomid. MS als chronische Erkrankung erfordert eine langfristige Therapie. Das reduzierte Risiko für Therapieabbrüche aufgrund von Nebenwirkungen ist ein patientenrelevanter Vorteil von 14 mg Teriflunomid gegenüber der etablierten Therapie.

Möglicherweise ist das Ergebnis durch die mangelnde Verblindung der Patienten verzerrt. Dabei ist unklar, zugunsten welcher Therapieoption eine mögliche Verzerrung auftreten könnte. Denkbar ist, dass Patienten unter Teriflunomid mit unerwünschten Ereignissen eher auf eine etablierte Therapie wechseln wollten, eine Enttäuschung darüber, nicht die neue, oral

einzunehmende Therapieoption zu erhalten könnte aber auch zu einem schnelleren Abbruch der Rebif® 44 µg-Therapie geführt haben. Insgesamt ist das Ergebnis als nicht oder niedrig verzerrt einzustufen. Für eine Erhöhung der Aussagesicherheit soll ein indirekter Vergleich durchgeführt werden, der sich lediglich auf doppelblinde Studien stützt.

4.3.1.3.10.5 Metaanalysen zu den Gesamtraten unerwünschter Ereignisse

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Die Studie TENERE lässt sich nicht mit den Studien TEMSO und TOWER zusammenfassen, während die letzteren beiden placebokontrollierte Studien sind, handelt es sich bei TENERE um eine aktivkontrollierte Studie, eine Metaanalyse ist somit nicht möglich.

Die Ergebnisse der Studien TEMSO und TOWER lassen sich in Bezug auf die Gesamtraten unerwünschter Ereignisse, schwerwiegender unerwünschter Ereignisse und unerwünschter Ereignisse, die zum Behandlungsabbruch führten im Sinne einer Metaanalyse zusammenfassen. Das entstehende Bild entspricht den auf den Ergebnissen der Einzelstudien basierenden Erwartungen: man findet keinen statistisch signifikanten Unterschied zwischen den Studienarmen und ein homogenes Ergebnis bezüglich der Gesamtraten unerwünschter Ereignisse und schwerwiegender unerwünschter Ereignisse. In Bezug auf die unerwünschten Ereignisse, die zum Behandlungsabbruch führten sind die Ergebnisse heterogen. In den beiden Studien hatten in der TOWER-Studie im 14 mg Teriflunomid-Arm signifikant mehr Patienten die Therapie aufgrund von unerwünschten Ereignissen abgebrochen als im Placebo-Arm, wogegen in der TEMSO-Studie in Bezug auf den gleichen Endpunkt kein Unterschied zwischen den Behandlungsgruppen nachzuweisen war. Eine schlüssige Erklärung für diesen Befund gibt es nicht.

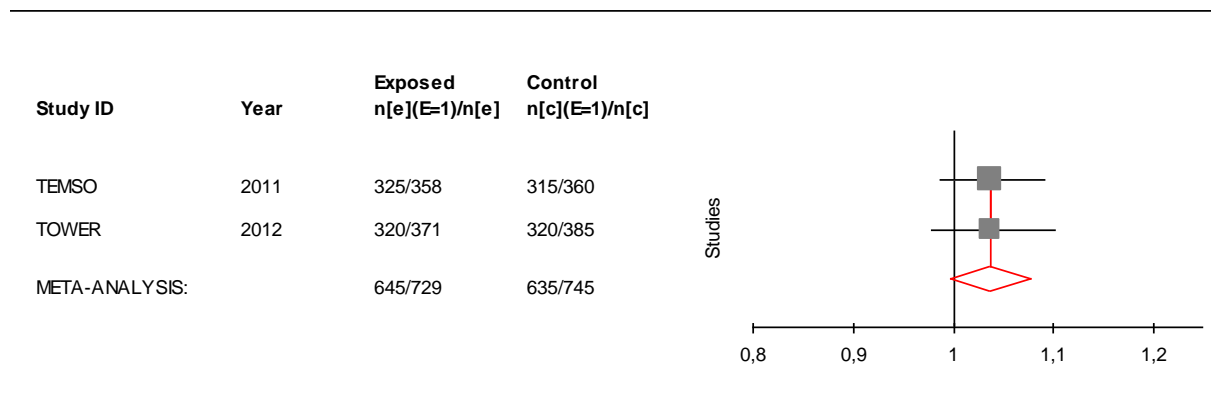
Gesamtrate unerwünschter Ereignisse

Tabelle 4-93: Metaanalyse-Ergebnisse für Gesamtrate unerwünschter Ereignisse aus TEMSO und TOWER – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^a	Odds Ratio 95 % KI p-Wert ^a
Metaanalyse von TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	1,038 [0,998; 1,079] p = 0,0642 I ² =0%, p _Q =0,9957	1,328 [0,978; 1,801] p = 0,0688 I ² =0%, p _Q =0,7546

^aH₀: RR=1 bzw. OR=1

Abbildung 4-15: Forest-Plot der Metaanalyse von TEMSO und TOWER für Gesamtrate unerwünschter Ereignisse (basierend auf relativen Risiken)



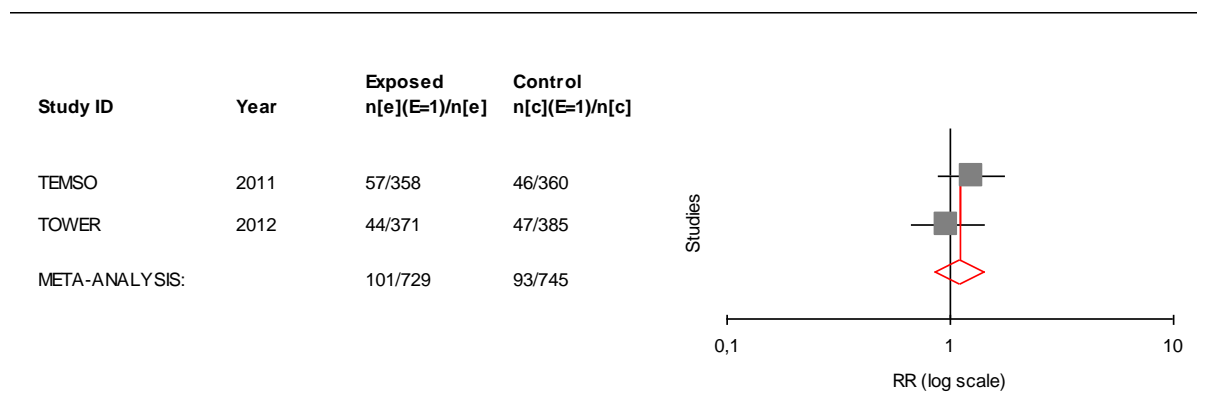
Gesamtrate schwerwiegender unerwünschter Ereignisse

Tabelle 4-94: Metaanalyse-Ergebnisse für Gesamtrate schwerwiegender unerwünschter Ereignisse aus TEMSO und TOWER – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^a	Odds Ratio 95 % KI p-Wert ^a
Metaanalyse von TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	1,110 [0,853; 1,444] p = 0,4379 I ² =0%, p _Q =0,3550	1,126 [0,831; 1,524] p = 0,4442 I ² =0%, p _Q =0,3495

^aH₀: RR=1 bzw. OR=1

Abbildung 4-16: Forest-Plot der Metaanalyse von TEMSO und TOWER für Gesamtrate schwerwiegender unerwünschter Ereignisse (basierend auf relativen Risiken)



Gesamtrate unerwünschter Ereignisse, die zum Behandlungsabbruch führten

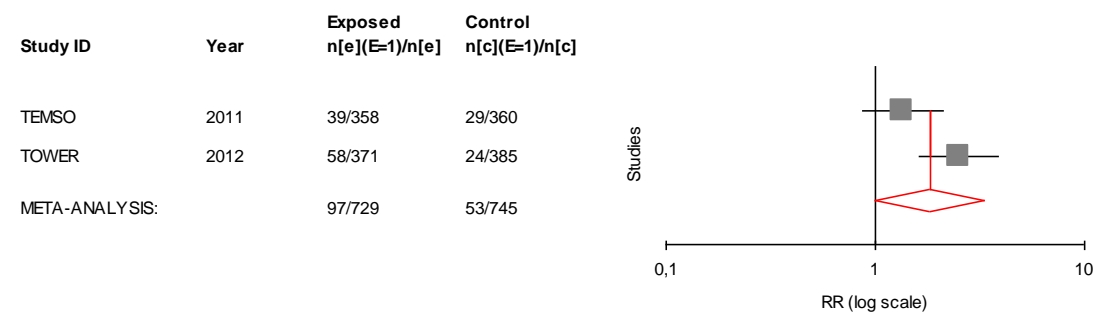
Tabelle 4-95: Metaanalyse-Ergebnisse für unerwünschte Ereignisse, die zum Behandlungsabbruch führten aus TEMSO und TOWER – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^a	Odds Ratio 95 % KI p-Wert ^a
Metaanalyse von TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	DL ^a : 1,843 [1,006; 3,376] p = 0,0477 I ² =72%, p _Q =0,0604	DL ^a : 1,974 [1,002; 3,889] p = 0,0493 I ² =73%, p _Q =0,0561

^aH₀: RR=1 bzw. OR=1

^aaufgrund der Heterogenität ist I²>0, so dass die Metaanalyse mit Studie als zufälligem Effekt (nach DerSimonian und Laird (=DL)) andere Ergebnisse liefert als das Modell mit Studie als festem Effekt. Verwendet und interpretiert wurden in diesen Fällen ausschließlich die Ergebnisse nach DerSimonian und Laird

Abbildung 4-17: Forest-Plot der Metaanalyse von TEMSO und TOWER für unerwünschte Ereignisse, die zum Behandlungsabbruch führten (basierend auf relativen Risiken)



4.3.1.3.11 Unerwünschte Ereignisse (Unerwünschte Ereignisse von besonderem Interesse) – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer Metaanalyse zusammengefasst und beschrieben werden.

Die tabellarische Darstellung der Ergebnisse für den jeweiligen Endpunkt soll mindestens die folgenden Angaben enthalten:

- Ergebnisse der ITT-Analyse
- Zahl der Patienten, die in die Analyse eingegangen sind
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe

- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler

Sofern die vorliegenden Studien für eine Metaanalyse geeignet sind, sollen die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammengefasst und als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Werden Studien in der Metaanalyse nicht berücksichtigt, ist das zu begründen.

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-96: Operationalisierung von unerwünschten Ereignissen von besonderem Interesse

Studie	Operationalisierung
Allgemeine Angaben	<p>Allgemeine Anmerkung: Unerwünschte Ereignisse, einschließlich abnormer Laborwerte, wurden nach dem MedDRA[®] Kodiersystem (Version 13.0 oder höher) kodiert. Verwendete MedDRA[®]-Versionen: TEMSO Version 13.0; TOWER: Version 15.0; TENERE: Version 14.0</p> <hr/> <p>TEMSO: Termine für „adverse event reporting“ waren – abgesehen vom Datum der Randomisierung – angesetzt in den Wochen 4, 8, 12, 16, 20, 24 und 30 und dann alle 6 Wochen bis zum Studienende.</p> <p>TOWER und TENERE: Entsprechende Reportings erfolgten in Woche 2 und Woche 6 und dann alle 6 Wochen bis Woche 24, dann in Woche 36 und ab diesem Termin alle 12 Wochen bis zum Studienende.</p>
TEMSO	<p>Folgende unerwünschte Ereignisse waren im Studienprotokoll als AESI (adverse event of special interest) prädefiniert:</p> <ul style="list-style-type: none"> • Leberfunktionsstörungen • Erkrankungen des Knochenmarks (jegliche Neutrophilenzahl von unter 1 Giga/L) • Erkrankungen des Pankreas (bestätigter Serumspiegel der Amylase oder Lipase auf über das Zweifache des oberen Normwertes, abdominaler Ultraschall oder CT- oder MRT-Scans mit dem Befund einer pankreatischen Auffälligkeit) • Malignität • Überempfindlichkeit und damit möglicherweise zusammenhängende Störungen • Alopezie • Lungenfunktionsstörungen und Ereignisse, die potentiell mit einer interstitiellen Lungenerkrankung in Verbindung stehen • Bluthochdruck • Periphere Neuropathie
TOWER	<p>Folgende unerwünschte Ereignisse waren im Studienprotokoll als AESI (adverse event of special interest) prädefiniert:</p>

	<ul style="list-style-type: none"> • Gastrointestinale Störungen (Übelkeit und Durchfall) • Leberfunktionsstörungen • Erkrankungen des Pankreas • Erkrankungen des Knochenmarks • Infektionen und Infestationen • Überempfindlichkeit/Hauterkrankungen • Malignität • Bluthochdruck • Herzrhythmusstörungen • Lungenfunktionsstörungen und Ereignisse, die potentiell mit einer interstitiellen Lungenerkrankung in Verbindung stehen • Thrombose und Embolie • Hämorrhagie • Periphere Neuropathie und potentiell damit zusammenhängende Ereignisse • Konvulsionen • Alopezie
TENERE	<p>Folgende unerwünschte Ereignisse waren im Studienprotokoll als AESI (adverse event of special interest) prädefiniert:</p> <ul style="list-style-type: none"> • Gastrointestinale Störungen (Übelkeit und Durchfall) • Leberfunktionsstörungen • Erkrankungen des Pankreas • Immunsuppression (Erkrankungen des Knochenmarks) • Infektionen und Infestationen • Malignität • Bluthochdruck • Herzrhythmusstörungen • Lungenfunktionsstörungen und Ereignisse, die potentiell mit einer interstitiellen Lungenerkrankung in Verbindung stehen • Hämorrhagie • Periphere Neuropathie und potentiell damit zusammenhängende Ereignisse • Konvulsionen • Depression und Suizid • Alopezie
AESI im vorliegenden Nutzendossier (für alle eingeschlossenen Studien berichtet)	<p>Unerwünschte Ereignisse von besonderem Interesse oder auch AESI (= adverse events of special interest) im Sinne des vorliegenden Dossiers sind solche, die bei der Behandlung mit der Intervention (Teriflunomid) oder der Vergleichstherapie (Beta-Interferon) zu erwarten sind und bei diesem Wirkstoff möglicherweise ein besonderes Sicherheitsrisiko darstellen. Die Auswahl der Ereignisse begründet sich auf Informationen der jeweiligen Fachinformationen. Vorzugsweise werden solche unerwünschten Ereignisse als von besonderem Interesse gewertet, die in den Studien als AESI prädefiniert waren.</p> <p>Erkrankungen des Pankreas, Malignität, Herzrhythmusstörungen, Hämorrhagie, Konvulsionen, Depressionen und Lungenerkrankungen sind in ein oder mehreren Studien prädefinierte AESI. Die Auswertungen der Studien in Bezug auf diese unerwünschten Ereignisse führen jedoch nicht zu deren Einstufung als besonderes Sicherheitsrisiko für 14 mg Teriflunomid oder die zweckmäßige Vergleichstherapie Rebif® 44 µg. Daher sei für diese unerwünschten Ereignisse auf die entsprechenden Studienberichte verwiesen [31, 32, 35].</p>

Das vorliegende Modul 4.3.1 stellt die Ergebnisse von RCT mit dem zu bewertenden Arzneimittel dar. Somit sind besondere Sicherheitsrisiken der zweckmäßigen Vergleichstherapie bei Beschränkung auf im Studienprotokoll prädefinierte Risiken noch nicht adäquat berücksichtigt. Zusätzlich werden deshalb Reaktionen an der Injektionsstelle, Grippeähnliche Symptome und Kopfschmerzen in den Katalog der dargestellten unerwünschten Ereignisse von besonderem Interesse aufgenommen.

Als unerwünschte Ereignisse von besonderem Interesse werden für das vorliegende Dossier demnach definiert:

- Reaktionen an der Injektionsstelle
- Grippeähnliche Symptome
- Alopezie
- Leberfunktionsstörungen (sowie Erhöhung des Alanin-Aminotransferase-(ALT)-Wertes)
 - Eine Erhöhung der ALT-Werte ist definiert als ein Wert der das 2fache des ULN (=upper normal level) übersteigt, wobei der Normbereich als ein Wert zwischen 6 und 43 IU/L angesehen wird
- Gastrointestinale Beschwerden
- Überempfindlichkeit/Hauterkrankungen
- Neutropenie
- Infektionen und Infestationen
- Bluthochdruck
- Periphere Neuropathie (inklusive Parästhesie)
- Kopfschmerzen

Die Kodierung erfolgt nach der MedDRA[®]-Terminologie, in dem die Terme in hierarchisch angeordnete Gruppen eingeordnet sind. Die Ereignisse sind in die folgenden 5 Kategorien eingeteilt:

- System Organ Class (SOC)
- High Level Group Term (HLGT)
- High Level Term (HLT)
- Preferred Term (PT)
- Lowest Level Term (LLT)

Im vorliegenden Dossier sind die unerwünschten Ereignisse je nach Art des Ereignisses als die Häufigkeit des Auftretens in den Kategorien SOC, HLGT, HLT oder PT berichtet. In jedem Fall wird für jedes Ereignis nicht nur die Häufigkeit des Ereignisses in seiner spezifischen Untergruppe, sondern auch die Anzahl der Ereignisse in der zugehörigen primären SOC angegeben. Durch dieses Vorgehen wird gewährleistet, dass alle relevanten Ereignisse erfasst sind, die Auswertung sich aber auf das spezifische, definierte Ereignis bezieht.

Tabelle 4-97: Ausgewertete AESI aufgeschlüsselt nach SOC, HLGT, HLT und PT

Primary SOC	HLGT	HLT	PT	Übersetzung (unterste angegebene Hierarchiestufe)
General disorders and administration site conditions	General system disorders NEC	General signs and symptoms NEC	Influenza like illness	Grippeähnliche Symptome
	Administration site reactions	Injection site reactions		Reaktionen an der Injektionsstelle
Skin and subcutaneous tissue disorders	Skin appendage conditions	Alopecias		Alopezien
Investigations	Hepatobiliary investigations	Liver function analyses	Alanine aminotransferase increased	Erhöhung des ALT-Wertes
Gastrointestinal disorders	Gastrointestinal motility and defaecation conditions	Diarrhoea (excl. infective)	Diarrhoea	Durchfall
	Gastrointestinal signs and symptoms	Nausea and vomiting symptoms	Nausea	Übelkeit
Blood and lymphatic system disorders	White blood cell disorders	Neutropenias		Neutropenien
Infections and Infestations				Infektionen und Infestationen
Vascular disorders	Vascular hypertensive disorders	Vascular hypertensive disorders NEC	Hypertension	Bluthochdruck
Nervous system disorders	Headaches			Kopfschmerzen
	Peripheral Neuropathies			Periphere Neuropathien

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-98: Bewertung des Verzerrungspotenzials für „Unerwünschte Ereignisse von besonderem Interesse“ in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TEMSO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
TENERE	niedrig	nein	ja	nein	nein	niedrig ^a

^aIn der TENERE-Studie sind die Patienten und der behandelnde Neurologe bezüglich der Medikation nicht verblindet. Jedoch werden die Gesamtrate unerwünschter Ereignisse, Gesamtrate schwerwiegender unerwünschter Ereignisse und die meisten spezifischen unerwünschten Ereignisse entweder klar definiert und mit einer objektivierbaren Komponente eingestuft oder aber es handelt sich um die Erfassung von Laborparametern, die in einem zentralen Labor erfasst und nicht durch eine subjektive Einschätzung der Patienten beeinflusst werden können. Bei unerwünschten Ereignissen, die diese Kriterien nicht erfüllen oder die als möglicherweise verzerrt eingestuft werden, weil sie nicht in den zugehörigen Studienprotokollen vorab definiert waren, ist dieses erhöhte Verzerrungspotential bei den entsprechenden Auswertungen diskutiert.

Begründen Sie für jede Studie die abschließende Einschätzung.

Das Verzerrungspotential wird für die Studien TEMSO, TOWER und TENERE auf Studienebene als niedrig eingestuft. Es gibt keinerlei Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Einflussfaktoren.

In den Studien TEMSO und TOWER gibt es keine Endpunkt-spezifischen Faktoren, die das Verzerrungspotential erhöhend beeinflussen, die Endpunkterheber waren verblindet und das ITT-Prinzip wurde adäquat umgesetzt (Analyse der Safety-Population (Definition s. 4.3.1.3.9)).

Die Durchführung der Studie und auch die Analyse der unerwünschten Ereignisse erfolgten nach GCP-Standards. Die Analyse der Safety-Population entspricht der methodengerechten Analyse sicherheitsrelevanter Ereignisse. Die Ereignisse sind nach MedDRA[®] 13.0 oder höher kodiert (s. Operationalisierung).

In der TENERE-Studie waren die Patienten bezüglich ihrer Medikation nicht verblindet. Dies kann, je nach Ausprägung der unerwünschten Ereignisse, das Ergebnis relevant verzerren. Der Großteil der unerwünschten Ereignisse ist jedoch klar definiert und beinhaltet eine objektivierbare Komponente (s. Fußnote Tabelle 4-98). Dadurch sind die Ergebnisse im Regelfall nicht verzerrt, eine mögliche Verzerrung bei einzelnen unerwünschten Ereignissen wird bei den entsprechenden Auswertungen diskutiert.

Stellen Sie die Ergebnisse für den Endpunkt Unerwünschte Ereignisse von besonderem Interesse für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Alle Auswertungen zu unerwünschten Ereignissen setzten die absoluten Häufigkeiten der entsprechenden Ereignisse in der Behandlungsgruppe (14 mg Teriflunomid-Gruppe) in relativen Bezug zu den Häufigkeiten in der jeweiligen Vergleichsgruppe (Placebo- oder Rebif®44 µg-Gruppe), so dass die Studiendauer von vernachlässigbarem Einfluss ist und im Weiteren nicht diskutiert wird.

4.3.1.3.11.1 Reaktionen an der Injektionsstelle

Reaktionen an der Injektionsstelle können bei der Verabreichung von Teriflunomid (Filmtablette) per definitionem nicht auftreten. TEMSO und TOWER waren doppelblinde Studien, in denen die entsprechenden Placebo-Tabletten optisch nicht von den Teriflunomid-Tabletten unterscheidbar waren. Sowohl im Behandlungs- als auch im Placebo-Arm wurde keine Injektion verabreicht, so dass es als unerwünschte Ereignisse keine Reaktionen an der Injektionsstelle geben kann.

Tabelle 4-99: Reaktionen an der Injektionsstelle – Ergebnisse für die gesamte Studienpopulation

Studie	Patientenzahl (Safety-Population) ^a	Primary SOC: General disorders and administration site conditions Anzahl (prozentualer Anteil)	HLGT: Administration site reactions Anzahl (prozentualer Anteil)	HLT: Injection site reactions 95% KI []
TEMSO	T14 (n=358)	90 (25,1 %)	1 (0,3 %)	1 (0,3 %)
	Pb (n=360)	92 (25,6 %)	1 (0,3 %)	0 (0 %)
TOWER	T14 (n=371)	72 (19,4%)	0 (0 %)	0 (0 %)
	Pb (n=385)	81 (21,0%)	0 (0 %)	0 (0 %)
TENERE ^c	T14 (n=110)	16 (14,5 %)	0 (0 %)	0 (0 %)
	Rb (n=101)	65 (64,4 %)	22 (21,8 %)	Inzidenz mit 95% KI: 0,218 [0,142; 0,311] ^b

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bberechnet nach Pearson und Clopper

^c99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Reaktionen an der Injektionsstelle sind eine wesentliche Nebenwirkung der parenteral applizierten Beta-Interferone (auch Rebif[®] 44 µg) [34], die in der Fachinformation mit einer Häufigkeit von ungefähr 30 % angegeben sind. Dies zeigt sich auch in der TENERE-Studie: Reaktionen an der Injektionsstelle traten bei 21,8 % der Patienten auf, die Rebif[®] 44 µg erhalten haben. Ein relativer Vergleich der Häufigkeit des Ereignisses im Rebif[®] 44 µg -Arm mit der entsprechenden Häufigkeit in den Teriflunomid-Armen ist nicht sinnvoll, da der Applikationsweg der beiden Interventionen unterschiedlich ist und bei der Behandlung mit Teriflunomid dieses unerwünschte Ereignis nicht eintreten kann. Darüber hinaus ist davon auszugehen, dass Reaktionen an der Injektionsstelle in einem geringeren Ausmaß auch bei einer Placebo-Injektion auftreten könnten, was den Vergleich zusätzlich erschwert.

Dennoch sind Reaktionen an der Injektionsstelle als unerwünschtes Ereignis an dieser Stelle beschreibend aufgeführt, um deutlich zu machen, dass die bekannte Häufigkeit dieser Nebenwirkung auch in der Teriflunomid-Studie TENERE bestätigt wurde. Ein wahrgenommener Mangel an Effektivität und unerwünschte Ereignisse sind die häufigsten Gründe für MS-Patienten, die Therapie mit krankheitsmodifizierenden Wirkstoffen abzubrechen und neben grippeähnlichen Symptomen sind Reaktionen an der Injektionsstelle ein Hauptgrund für mangelnde Therapieadhärenz [20, 65]. In der TENERE-Studie konnten Reaktionen an der Injektionsstelle als „sehr häufiges“ unerwünschtes Ereignis bei der Therapie mit Interferon beta-1a bestätigt werden. Die Tatsache, dass diese Nebenwirkung unter Teriflunomid-Therapie nicht auftreten kann, zeigt den beträchtlichen Zusatznutzen, den Teriflunomid aufgrund seiner oralen Darreichungsform hat.

Im Beratungsgespräch informierte der G-BA darüber, dass „Reaktionen an der Injektionsstelle“ als zusätzlicher Endpunkt, der den Vorteil oraler versus parenteraler Applikation abbilden kann, anerkannt wird [11].

4.3.1.3.11.2 Grippeähnliche Symptome

Wie unter der Diskussion zu Reaktionen an der Injektionsstelle bereits aufgeführt, sind grippeähnliche Symptome eine weitere häufige Nebenwirkung von Interferon beta-1a, die auch einen erheblichen Einfluss auf Therapieadhärenz und Compliance haben können. Grippeähnliche Symptome treten laut Fachinformation von Rebif[®] 44 µg bei etwa 70 % der Patienten auf [34]. Rebif[®] 44 µg ist die zweckmäßige Vergleichstherapie zur Nutzenbewertung von Teriflunomid und dementsprechend ist die Auswertung grippeähnlicher Symptome für das vorliegende Nutzendossier ausgesprochen relevant.

Grippeähnliche Symptome sind relativ unspezifisch und können auch unabhängig von einer Interferon-beta-Applikation auftreten. Aus diesem Grund gibt es bezüglich dieses Ereignisses keine Studienarme, in denen kein Ereignis („Doppel-Null-Studien“ oder „Einfach-Null-Studien“, s. 4.2.5.3) aufgetreten ist und dieser Endpunkt kann auf Basis der relativen Risiken ausgewertet werden.

Tabelle 4-100: Grippeähnliche Symptome – Ergebnisse für die gesamte Studienpopulation

Studie	Patientenzahl (Safety-Population) ^a	Primary SOC: General disorders and administration site conditions Anzahl (prozentualer Anteil)	HLGT: General system disorders NEC ^c Anzahl (prozentualer Anteil)	HLT: General signs and symptoms NEC ^c Anzahl (prozentualer Anteil)	PT: Influenza like illness Anzahl (prozentualer Anteil)	Relative Risk (RR) 95% KI [] p-Wert ^b	Odds Ratio (OR) 95% KI [] p-Wert ^b
TEMSO	T14 (n=358)	90 (25,1 %)	83 (23,2 %)	9 (2,5 %)	8 (2,2 %)	0,894 [0,349; 2,291] p = 0,8152	0,891 [0,340; 2,337] p = 0,8152
	Pb (n=360)	92 (25,6 %)	83 (23,1 %)	11 (3,1 %)	9 (2,5 %)		
TOWER	T14 (n=371)	72 (19,4%)	69 (18,6%)	10 (2,7 %)	8 (2,2%)	0,922 [0,360; 2,365] p = 0,8665	0,921 [0,351; 2,412] p = 0,8665
	Pb (n=385)	81 (21,0%)	69 (17,9%)	11 (2,9 %)	9 (2,3%)		
TENERE ^f	T14 (n=110)	16 (14,5 %)	14 (12,7 %)	3 (2,7 %)	3 (2,7 %)	0,051 ^d [0,017; 0,158] p < 0,0001	0,024 ^e [0,007; 0,082] p < 0,0001
	Rb (n=101)	65 (64,4 %)	61 (60,4 %)	54 (53,5 %)	54 (53,5 %)		

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bH₀: RR = 1 bzw. H₀: OR = 1, 95 % KI und p-Werte asymptotisch berechnet für relative Risiken nach der Delta-Methode

^cNEC = not elsewhere classified

^dkombinatorisch exakt berechneter Wert: 0,051, 95 % KI: [0,012; 0,139], p < 0,0001

^ekombinatorisch exakt berechneter Wert: 0,024, 95 % KI: [0,005; 0,083]; p < 0,0001

^f99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

TEMSO und TOWER

In den Studien TEMSO und TOWER traten Grippeähnliche Symptome mit ähnlicher Ereignishäufigkeit (2,2 % – 2,5 %) in den unterschiedlichen Behandlungsgruppen auf, es konnte mit relativen Risiken in der 14 mg Teriflunomid-Gruppe von 0,894 (95 % KI: [0,349; 2,291]) in der TEMSO-Studie und 0,922 (95 % KI: [0,360; 2,365]) in der TOWER-Studie kein statistisch signifikanter Unterschied zu Placebo festgestellt werden. Das Auftreten von grippeähnlichen Symptomen ist unter Teriflunomid-Behandlung nicht häufiger als unter Placebo.

TENERE

In der TENERE-Studie zeigt sich hinsichtlich grippeähnlicher Symptome ein deutlicher Vorteil für Teriflunomid. Unter der Behandlung mit 14 mg Teriflunomid trat das Ereignis mit 2,7 % in ähnlicher Häufigkeit auf wie in den placebokontrollierten Studien TEMSO und TOWER. Die vergleichbare absolute Häufigkeit entspricht trotz der unterschiedlichen Studiendauer (s. Tabelle 4-89) dem Ergebnis der placebokontrollierten Studien.

Die Behandlung mit Rebif[®] 44 µg führt dagegen in über der Hälfte der Fälle (53,5 %) zu dem Auftreten grippeähnlicher Symptome. Dieses Ergebnis entspricht einem relativen Risiko von 0,051, der Unterschied ist statistisch signifikant ($p < 0,0001$). Das Risiko, an grippeähnlichen Symptomen zu leiden, ist unter der Behandlung mit 14 mg Teriflunomid um 95 % gesenkt.

Medizinisch werden grippeähnliche Symptome, die auf eine Therapie mit Interferon beta-1a zurückzuführen sind, als nicht schwerwiegende Nebenwirkung eingestuft, auch wenn sie für die Patienten sehr unangenehm sind und bei einer Dauermedikation die Lebensqualität der Betroffenen erheblich einschränken können. Das Ausmaß der Vermeidung von grippeähnlichen Symptomen, das in dem relativen Risiko von 0,051 und dem zugehörigen 95 %-Konfidenzintervall von [0,017; 0,158] deutlich wird, ist beträchtlich und für die Dauertherapie und die Lebensqualität der MS-Patienten ausgesprochen bedeutsam.

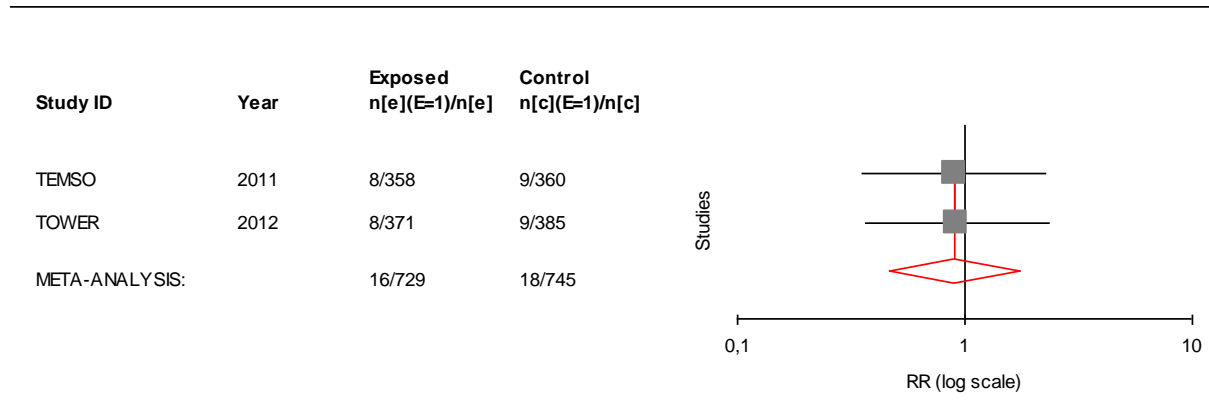
Fazit: Mit der Vermeidung grippeähnlicher Symptome in dem gezeigten Ausmaß ist für das zu bewertende Arzneimittel (Teriflunomid (AUBAGIO[®])) ein beträchtlicher Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie (Interferon beta-1a, 44 µg, subkutan (Rebif[®])) gegeben.

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Die Studie TENERE lässt sich nicht mit den Studien TEMSO und TOWER zusammenfassen, während die letzteren beiden placebokontrollierte Studien sind, handelt es sich bei TENERE um eine aktivkontrollierte Studie, eine Metaanalyse ist nicht möglich.

Eine Metaanalyse der Ergebnisse zu grippeähnlichen Symptomen der Studien TEMSO und TOWER ist Bestandteil des indirekten Vergleichs und findet sich dort (s. Modul 4.3.2).

Abbildung 4-18: Forest-Plot der Metaanalyse von TEMSO und TOWER für grippeähnliche Symptome (basierend auf relativen Risiken)



4.3.1.3.11.3 Neutropenie

Für die Bewertung der sicheren Anwendung von Teriflunomid sind Erkrankungen des Knochenmarks als AESI in den Studienprotokollen prädefiniert, da es sich um eine selektiv immunsuppressiv wirkende Substanz handelt. Als Erkrankung des Knochenmarks werden spezifische Veränderungen des Blutbildes gewertet. Im Studienprotokoll der Studie TEMSO ist darunter eine Neutrophilenzahl von unter 1000 /L definiert. Die Häufigkeit von Lymphopenie, Thrombozytopenie, und einer pathologisch verringerten Anzahl der Blutplättchen war in den Studien gering, daher wird in diesem Dossier lediglich die Neutropenie ausgewertet (für die Häufigkeiten der anderen Veränderungen des Blutbildes sei an dieser Stelle auf die entsprechenden Studienberichte verwiesen [31, 32, 35]). Im Rahmen der Studie wurden alle 2 Wochen bis zur Woche 30, dann in Woche 36 und 48 und im Anschluss alle 12 Wochen Laboruntersuchungen durchgeführt, die u. a. die Bestimmung der Anzahl folgender Zellen beinhaltete:

- Leukozyten
- Neutrophile
- Lymphozyten
- Monozyten
- Eosinophile
- Basophile
- Thrombozyten

Zusätzlich zu dem entsprechenden high level term (HLT) Neutropenie wird auch die entsprechende primäre Organklasse (Blood and Lymphatic system disorders) in Tabelle 4-101 dargestellt.

Tabelle 4-101: Neutropenie – Ergebnisse für die gesamte Studienpopulation

Studie	Patienten- zahl (Safety- Population) ^a	Primary SOC: Blood and Lymphatic system disorders	HLGT: White blood cell disorders	HLT: Neutro- penias	Relative Risk (RR) 95% KI [] p(H ₀ : RR = 1)	Odds Ratio (OR) 95% KI [] p(H ₀ : OR = 1)
TEMESO	T14 (n=358)	43 (12,0 %)	23 (6,4 %)	16 (4,5 %)	8,045 [1,863; 34,73] p = 0,0052	8,374 [1,911; 36,69] p = 0,0048
	Pb (n=360)	35 (9,7 %)	11 (3,1 %)	2 (0,6 %)		
TOWER	T14 (n=371)	54 (14,6 %)	41 (11,1 %)	35 (9,4%)	3,302 [1,703; 6,403] p = 0,0004	3,542 [1,771; 7,085] p = 0,0004
	Pb (n=385)	24 (6,2 %)	14 (3,6 %)	11 (2,9%)		
TENERE ^b	T14 (n=110)	7 (6,4 %)	6 (5,5 %)	4 (3,6 %)	0,612 [0,178; 2,107] p = 0,4364	0,597 [0,164; 2,182] p = 0,4357
	Rb (n=101)	10 (9,9 %)	8 (7,9 %)	6 (5,9 %)		

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^b99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

TEMESO

In der TEMESO-Studie ergab sich in der 14 mg Teriflunomid-Gruppe ebenfalls ein signifikant höheres Risiko für eine Neutropenie als in der Placebo-Gruppe (relatives Risiko von 8,045; 95 % KI: [1,863; 34,73]). Die Höhe des relativen Risikos ist auch auf die geringe Ereignishäufigkeit (lediglich 2 Patienten) in der Placebo-Gruppe zurückzuführen, wobei das Ergebnis mit einer entsprechenden Unsicherheit verknüpft ist, was sich in der Breite des Konfidenzintervalls widerspiegelt.

TOWER

In der TOWER-Studie ergab sich für die 14 mg Teriflunomid-Gruppe ein signifikant höheres Risiko für eine Neutropenie als in der Placebo-Gruppe (relatives Risiko von 3,302; 95 % KI: [1,703; 6,403]).

Sowohl in der TEMSO als auch in der TOWER-Studie ergab sich in den Teriflunomid-Gruppen ein signifikant höheres Risiko für eine Neutropenie als in der Placebo-Gruppe. Dieses erhöhte Risiko einer auftretenden Neutropenie ist der Grund dafür, dass bei der Behandlung mit Teriflunomid regelmäßige Blutbildkontrollen durchgeführt werden sollen (s. Modul 3.4 Anforderungen an eine qualitätsgesicherte Anwendung). Das Ausmaß des höheren Risikos lässt sich schwer abschätzen, da sich die Ergebnisse in TEMSO und TOWER stark unterscheiden. In der TOWER-Studie sind Neutropenien insgesamt häufiger aufgetreten, das 95 % Konfidenzintervall ist deutlich schmaler und der hier ermittelte Unterschied zwischen den Gruppen ist wahrscheinlich näher an dem wahren Unterschied als der in TEMSO ermittelte Wert. Es bleibt jedoch unklar, warum in der TOWER-Studie verstärkt Neutropenien aufgetreten sind.

TENERE

Im direkten Vergleich von Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie Rebif® 44 µg zeigt sich kein signifikanter Unterschied zwischen den Interventionen. Mit einem relativen Risiko von 0,612 (95 % KI: [0,178; 2,107]) zeigt sich zwar tendenziell ein leichter Vorteil für 14 mg Teriflunomid, der Unterschied war jedoch nicht signifikant.

Neutropenie ist keine subjektiv durch den Patienten berichtete Nebenwirkung, sondern wird zu vorab definierten Zeitpunkten durch eine unabhängige Laboranalyse erfasst. Das Verzerrungspotential ist dementsprechend trotz mangelnder Verblindung von Patient und behandelndem Arzt als niedrig einzustufen.

Fazit

Durch die Therapie mit Teriflunomid ist das Risiko einer Neutropenie erhöht. Das Risiko ist jedoch nicht höher als bei Therapie mit Rebif® 44 µg. Eine Unterstützung der Daten durch einen indirekten Vergleich wäre für die Interpretation dieses unerwünschten Ereignisses sinnvoll, ist jedoch nicht möglich, da der HLT Neutropenie in der Publikation der placebokontrollierte Rebif-Studie PRISMS nicht ausgewertet ist. Im indirekten Vergleich werden deswegen Lymphopenie und Leukopenie als pathologische Beeinträchtigungen des blutbildenden Systems ausgewertet (s. Modul 4.3.2).

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Die Studie TENERE lässt sich nicht mit den Studien TEMSO und TOWER zusammenfassen, während die letzteren beiden placebokontrollierte Studien sind, handelt es sich bei TENERE um eine aktivkontrollierte Studie, eine Metaanalyse ist nicht möglich.

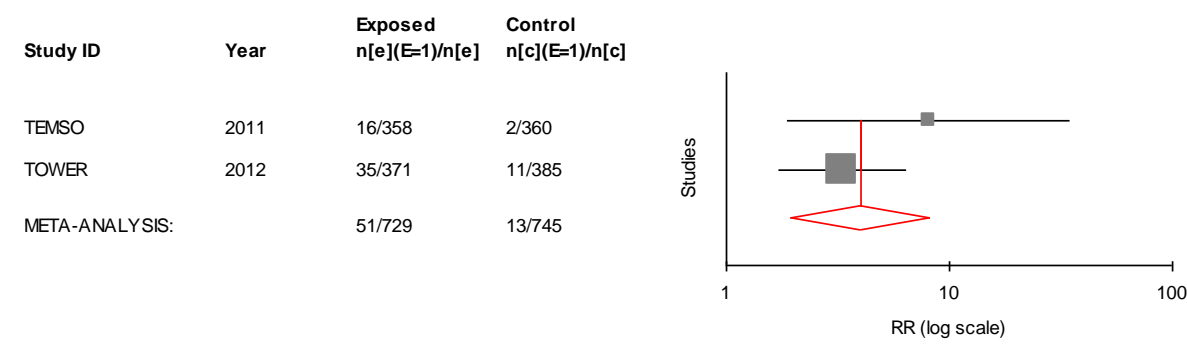
Die Metaanalyse der Ergebnisse zur Neutropenie der Studien TEMSO und TOWER zeigt leichte Heterogenität von wahrscheinlich unbedeutendem Ausmaß und zusätzlich deutliche gleichgerichtete Effekte.

Tabelle 4-102: Metaanalyse-Ergebnisse für Neutropenie aus TEMSO und TOWER – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse von TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	DL ^a : 4,020 [1,949; 8,291] p = 0,0002 I ² =15,4 %, p _Q =0,2770	DL ^a : 4,210 [2,142; 8,275] p<0,0001 I ² =6,4 %, p _Q =0,3014

^aaufgrund der Heterogenität ist I²>0, so dass die Metaanalyse mit Studie als zufälligem Effekt (nach DerSimonian und Laird (=DL)) andere Ergebnisse liefert als das Modell mit Studie als festem Effekt. Verwendet und interpretiert wurden in diesen Fällen ausschließlich die Ergebnisse nach DerSimonian und Laird

Abbildung 4-19: Forest-Plot der Metaanalyse von TEMSO und TOWER für Neutropenie (basierend auf relativen Risiken)



4.3.1.3.11.4 Alopecie

Teriflunomid ist der Hauptmetabolit von Leflunomid. Verstärkter Haarausfall ist eine häufige Nebenwirkung von Leflunomid (Arava[®]) [66]. Aus diesem Grund wurde ein verstärkter Haarausfall in den Teriflunomid-Studien als AESI prädefiniert und nach dem MedDRA[®] HLT (high level term) „Alopecias“ ausgewertet.

Tabelle 4-103: Alopecie – Ergebnisse für die gesamte Studienpopulation

Studie	Patientenzahl (Safety-Population) ^a	Primary SOC: Skin and subcutaneous tissue disorders Anzahl (prozentualer Anteil)	HLGT: Skin Appendage conditions Anzahl (prozentualer Anteil)	HLT: Alopecias Anzahl (prozentualer Anteil)	Relative Risk (RR) 95% KI [] p-Wert ^b	Odds Ratio (OR) 95% KI [] p-Wert ^b
--------	---	--	--	---	--	---

Anteil)						
TEMSO	T14	125	60	47	3,939	4,383
	(n=358)	(34,9 %)	(16,8 %)	(13,1%)	[2,126; 7,298]	[2,283; 8,413]
					p < 0,0001	p < 0,0001
	Pb	66	23	12		
	(n=360)	(18,3 %)	(6,4 %)	(3,3%)		
TOWER	T14	104	59	52	2,998	3,324
	(n=371)	(28,0 %)	(15,9 %)	(14,0%)	[1,788; 5,026]	[1,905; 5,799]
					p < 0,0001	p < 0,0001
	Pb	69	25	18		
	(n=385)	(17,9%)	(6,5 %)	(4,7 %)		
TENERE ^c	T14	41	28	22	20,2 ^c	25,0 ^d
	(n=110)	(37,3 %)	(25,5 %)	(20,0 %)	[2,773; 147,1]	[3,302; 189,3]
					p = 0,0030	p = 0,0018
	Rb	18	3	1		
	(n=101)	(17,8%)	(3,0 %)	(1,0 %)		

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bH₀: RR = 1 bzw. H₀: OR = 1

^ckombinatorisch exakt berechneter Wert: 20,2; 95 % KI: [3,835; 572,2], p < 0,0001

^dkombinatorisch exakt berechneter Wert: 25,0; 95 % KI: [3,841; 1042], p < 0,0001

^e99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

TEMSO

Die Wahrscheinlichkeit eines unerwünschten Ereignisses, dass unter den HLT „Alopecias“ fällt ist für 14 mg Teriflunomid im Vergleich zu Placebo auf das Vierfache erhöht (relatives Risiko von 3,939; 95 % KI: [2,126; 7,298]). Das Ergebnis ist statistisch signifikant. Die Berichte über eine aufgetretene Alopecie beschrieben typischerweise einen milden oder moderaten Verlauf, was für den Patienten die Wahrnehmung einer Verminderung der Haardichte bedeutet; in der 14 mg Teriflunomid-Gruppe ist ein Fall einer schweren Alopecie aufgetreten.

TOWER

Die Wahrscheinlichkeit einer Alopecie ist für 14 mg Teriflunomid im Vergleich zu Placebo auf das Dreifache erhöht (relatives Risiko von 2,998; 95 % KI: [1,788; 5,026]). Das Ergebnis ist statistisch signifikant. Die meisten Fälle waren auch in dieser Studie als milde Formen einer Alopecie eingestuft, zwei Fälle einer schweren Alopecie sind in der 14 mg Teriflunomid-Gruppe aufgetreten.

Die Ergebnisse aus den beiden unabhängigen placebokontrollierten Studien stimmen gut miteinander überein. In keiner der beiden Studien ist ein kompletter Haarverlust aufgetreten und keine der entsprechenden Ereignisse wurde als schwerwiegendes unerwünschtes Ereignis eingestuft. Der überwiegende Teil der Fälle nahm einen milden bis moderaten Verlauf. Alopezie trat in der Regel innerhalb der ersten 6 Monate der Therapie auf, mit einer durchschnittlichen Dauer von ungefähr 4 Monaten. Die aufgetretene Alopezie bildete sich ohne eine korrigierende Therapie wieder zurück (auch bei Fortführung der Behandlung mit Teriflunomid) [33, 67].

TENERE

In der TENERE-Studie traten unerwünschte Ereignisse, die unter den HLT „Alopecias“ fallen unter 14 mg Teriflunomid deutlich häufiger auf als in der Vergleichsgruppe (relatives Risiko von 20,2; 95 % KI: [2,773; 147,1]).

Das extrem breite 95%-Konfidenzintervall von 2,773 bis 147,1 macht die große Unsicherheit, die mit dem Wert (relatives Risiko von 20,2) verknüpft ist, deutlich, es lässt sich jedoch in beiden Dosierungen ein signifikant höheres Risiko für Teriflunomid feststellen als für die zweckmäßige Vergleichstherapie.

Fazit:

Unter der Therapie mit Teriflunomid kann es zu Haarausfall (Alopezie) im Sinne einer Lichtung des Haares oder auch verminderter Haardichte kommen. Die meisten Fälle wurden als diffus oder generalisiert auf der Kopfhaut beschrieben (es wurde kein vollständiger Haarausfall berichtet). Diese Ereignisse traten zumeist innerhalb der ersten 6 Monate der Behandlung auf und bildeten sich bei dem Großteil der mit Teriflunomid 14 mg behandelten Patienten zurück. Zu einem Behandlungsabbruch aufgrund von Alopezie kam es nur in Einzelfällen [33].

Die unter dem HLT Alopecias kodierten Effekte auf das Haarvolumen sind unerwünschte Ereignisse, die sich mit der Behandlung mit Teriflunomid assoziieren lassen, sie nehmen in der Regel jedoch einen leichten und zeitlich begrenzten Verlauf. Aus der TENERE-Studie lässt sich die Größe des relativen Risikos für dieses Ereignis nicht präzise abschätzen. Basierend auf den Werten aus den Studien TEMSO und TOWER wird eine Risikoerhöhung auf das Drei- bis Vierfache angenommen.

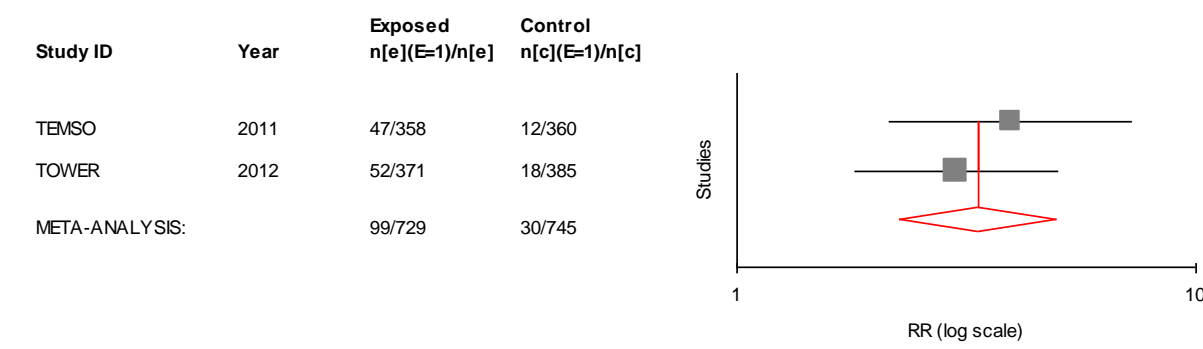
Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Die Studie TENERE lässt sich nicht mit den Studien TEMSO und TOWER zusammenfassen, während die letzteren beiden placebokontrollierte Studien sind, handelt es sich bei TENERE um eine aktivkontrollierte Studie, eine Metaanalyse ist nicht möglich.

Tabelle 4-104: Metaanalyse-Ergebnisse für verminderte Haardichte (HLT Alopecias) aus TEMSO und TOWER – Ergebnisse für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse von TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	3,355 [2,258; 4,985] p<0,0001 I ² =0%, p _Q =0,5062	3,735 [2,446; 5,703] p<0,0001 I ² =0%, p _Q =0,5272

Abbildung 4-20: Forest-Plot der Metaanalyse von TEMSO und TOWER für verminderte Haardichte (HLT Alopecias) (basierend auf relativen Risiken)



4.3.1.3.11.5 Überempfindlichkeit/Hautreaktionen

Überempfindlichkeit gegenüber der verabreichten Studienmedikation wird – abweichend von der Vorgehensweise bei anderen AESI – als Gesamtrate über alle SOC's (any class) angegeben, da die typischen Symptome einer Überempfindlichkeit, wie z. B. Asthma, Dyspnoe, Husten, Blutdruckabfall, Pruritus und Hautausschlag, verschiedenen SOC's zugeordnet sind.

Tabelle 4-105: Überempfindlichkeit/Hautreaktionen – Ergebnisse für die gesamte Studienpopulation

Studie	Patienten-zahl (Safety-Population) ^a	Gesamtrate über alle SOC's	Relative Risk (RR) 95% KI [] p-Wert ^b	Odds Ratio (OR) 95% KI [] p-Wert ^b
TEMSO	T14 (n=358)	73 (20,4 %)	1,335 [0,971; 1,834] p = 0,0751	1,420 [0,966; 2,088] p = 0,0743
	Pb (n=360)	55 (15,3 %)		
TOWER	T14	60	1,092	1,110

	(n=371)	(16,2 %)	[0,783; 1,524] p = 0,6034	[0,748; 1,647] p = 0,6034
	Pb	57		
	(n=385)	(14,8 %)		
TENERE ^c	T14	14	1,071	1,082
	(n=110)	(12,7 %)	[0,520; 2,206] p = 0,8519	[0,475; 2,464] p = 0,8519
	Rb	12		
	(n=101)	(11,9 %)		

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bH₀: RR = 1 bzw. H₀: OR = 1

^c99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

TEM SO

Der Anteil an Patienten mit Überempfindlichkeits- und Hautreaktionen war in der 14 mg Teriflunomid-Gruppe höher als in der Placebo-Gruppe. Der Unterschied zwischen den Gruppen war statistisch nicht signifikant, eine leichte Tendenz für einen Nachteil von Teriflunomid lässt sich jedoch erkennen. Ein Patient in der 14 mg Teriflunomid-Gruppe brach die Studienmedikation aufgrund von Pruritus ab.

TOWER

Die Inzidenz an Überempfindlichkeits- und Hautreaktionen war in den Behandlungsgruppen sehr ähnlich, es konnte kein Unterschied zwischen den Behandlungsgruppen gezeigt und auch keine deutliche Tendenz abgeleitet werden. Ein Patient in der Placebo-Gruppe und 3 Patienten in der 14 mg Teriflunomid-Gruppe brachen die Studienmedikation aufgrund von Überempfindlichkeits- und Hautreaktionen ab.

TENERE

Die Inzidenz an Überempfindlichkeits- und Hautreaktionen war in den Behandlungsgruppen sehr ähnlich, es konnte kein Unterschied zwischen den Behandlungsgruppen gezeigt und auch keine deutliche Tendenz abgeleitet werden.

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

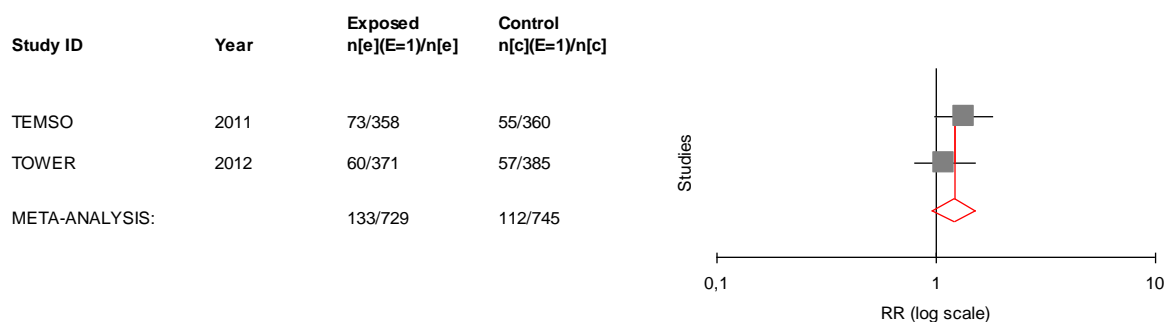
Die Studie TENERE lässt sich nicht mit den Studien TEMSO und TOWER zusammenfassen, während die letzteren beiden placebokontrollierte Studien sind, handelt es sich bei TENERE um eine aktivkontrollierte Studie, eine Metaanalyse ist nicht möglich.

Tabelle 4-106: Metaanalyse-Ergebnisse für Überempfindlichkeits- und Hautreaktionen aus TEMSO und TOWER – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^a	Odds Ratio 95 % KI p-Wert ^a
Metaanalyse von TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	1,213 [0,964; 1,527] p = 0,0997 I ² =0%, p _Q =0,3939	1,259 [0,956; 1,659] p = 0,1011 I ² =0%, p _Q =0,3810

^aH₀: RR = 1 bzw. H₀: OR = 1

Abbildung 4-21: Forest-Plot der Metaanalyse von TEMSO und TOWER für Überempfindlichkeits- und Hautreaktionen (basierend auf relativen Risiken)



4.3.1.3.11.6 Hepatische unerwünschte Ereignisse (inklusive Erhöhung der ALT-Werte)

Leberfunktionsstörungen sind in allen Teriflunomid-Studien als unerwünschtes Ereignis von besonderem Interesse definiert. Diese Einstufung für die Teriflunomid-Studien basierte auf bestehenden Erfahrungen mit der Muttersubstanz Leflunomid, für die eine Erhöhung der Leberwerte als häufiges Ereignis eingestuft wurde [66]. In der Fachinformation der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg sind schwerwiegende Transaminasenerhöhungen ebenfalls als häufiges Ereignis angegeben [34], so dass Leberfunktionsstörungen sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie als spezielles Sicherheitsrisiko anzusehen sind, obwohl es in präklinischen Studien zu Teriflunomid keine Signale für hepatische Toxizität gab.

Zur Überprüfung der Leberfunktion wurden in Abständen von 2 Wochen bis zur Woche 30, dann in Woche 36 und 48 und im Anschluss alle 12 Wochen Laboruntersuchungen durchgeführt, die unter anderem die Erfassung der folgenden Laborparameter beinhaltete:

- ALT (Alanin-Aminotransferase (ALAT, ALT), auch Glutamat-Pyruvat-Transaminase (GPT) genannt)
- AST (Aspartat-Aminotransferase (AST, ASAT, AAT), auch Glutamat-Oxalacetat-Transaminase (GOT) genannt)
- GGT (Gamma-Glutamyl-Transferase (GGT, Gamma-GT))
- Gesamtbilirubin
- Direktes/indirektes Bilirubin

Tabelle 4-107: Hepatische unerwünschte Ereignisse – Ergebnisse für die gesamte Studienpopulation

Mit der Leberfunktion assoziierte Ereignisse						
Studie	Patienten- zahl (Safety- Population) ^a	Patienten mit irgend- einem hepatischen TEAE	Patienten mit hepatische m TESAE	Patienten mit hepatischem TEAE, das zum Behandlungs- abbruch führte	EAIR ^b	Exposure adjusted Relative Risk [95 % KI] vs. Placebo bzw. Rebif
TEM SO	T14 (n=358)	69 (19,3 %)	9 (2,5 %)	13 (3,6 %)	12,5	1,72 [1,18; 2,54] p = 0,0056 ^c
	Pb (n=360)	43 (11,9 %)	9 (2,5 %)	15 (4,2 %)	7,2	
TOWER	T14 (n=371)	81 (21,8 %)	6 (1,6 %)	18 (4,9 %)	15,5	1,54 [1,10; 2,17] p = 0,0127 ^c
	Pb (n=385)	58 (15,1 %)	11 (2,9 %)	12 (3,1 %)	10,0	
TENERE ^d	T14 (n=110)	14 (12,7 %)	1 (0,9 %)	4 (3,6 %)	11,0	0,25 [0,13; 0,45] p < 0,0001 ^c
	Rb (n=101)	40 (39,6 %)	1 (1,0 %)	10 (9,9 %)	44,3	

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bEAIR: Exposure-adjusted incidence rate, erwartete Anzahl an Patienten mit mindestens einem mit der Leberfunktion assoziierten Ereignis pro 100 Patienten, die ein Jahr unter Behandlung stehen (52 Wochen)

^cH₀: exposure adjusted RR = 1

^d99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Tabelle 4-108: Patienten mit hepatischen unerwünschten Ereignissen – Ergebnisse für die gesamte Studienpopulation

Patienten mit irgendeinem hepatischen TEAE					
Studie	Patientenzahl (Safety-Population) ^a	Anzahl Patienten mit mindestens einem entsprechenden Ereignis	Prozentualer Anteil Patienten mit mindestens einem entsprechenden Ereignis	Relative Risk (RR) 95% KI [] p-Wert ^b	Odds Ratio (OR) 95% KI [] p-Wert ^b
TEMSO	T14 (n=358)	69	19,3 %	1,614 [1,135; 2,293] p = 0,0076	1,760 [1,165; 2,660] P = 0,0073
	Pb (n=360)	43	11,9 %		
TOWER	T14 (n=371)	81	21,8 %	1,449 [1,068; 1,967] p = 0,0173	1,575 [1,085; 2,285] p = 0,0168
	Pb (n=385)	58	15,1 %		
TENERE ^c	T14 (n=110)	14	12,7 %	0,321 [0,186; 0,554] p < 0,0001	0,222 [0,112; 0,443] p < 0,0001
	Rb (n=101)	40	39,6 %		

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bH₀: RR = 1 bzw. H₀: OR = 1

^c99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

TEMSO

Insgesamt waren mit der Leberfunktion assoziierte unerwünschte Ereignisse in der 14 mg Teriflunomid-Gruppe häufiger als in der Placebo-Gruppe (relatives Risiko von 1,72, 95 % KI: [1,18; 2,54]). Dabei entfiel der größte Anteil auf den HLT „hepatobiliary investigations“ (v. a. Erhöhung der ALT-Werte). Der Anteil an Patienten in der SOC „hepatobiliary disorders“ war in der 14 mg Teriflunomid-Gruppe nicht erhöht. Die Anzahl Patienten mit schweren unerwünschten Ereignissen, die mit der Leberfunktion assoziiert waren und entsprechende Ereignisse, die zum Behandlungsabbruch führten, waren in den Behandlungsgruppen vergleichbar.

TOWER

Der Anteil an Patienten mit unerwünschten Ereignissen, die mit der Leberfunktion assoziiert waren, war in der 14 mg Teriflunomid-Gruppe höher als in der Placebo-Gruppe (relatives

Risiko von 1,54; 95 % KI: [1,10; 2,17]). Der höchste Anteil entfiel – ähnlich wie in der TEMSO-Studie – auf die SOC „investigations“ (v. a. Erhöhung der ALT-Werte). Die Anzahl Patienten mit schweren unerwünschten Ereignissen, die mit der Leberfunktion assoziiert waren und entsprechende Ereignisse, die zum Behandlungsabbruch führten, waren in den Behandlungsgruppen vergleichbar.

TENERE

Die Häufigkeit der mit der Leberfunktion assoziierten Ereignisse war insgesamt in der Rebif® 44 µg -Gruppe höher als in der 14 mg Teriflunomid-Gruppe (relatives Risiko von 0,25; 95 % KI: [0,13; 0,45]). Dabei entfiel der größte Anteil auf die Erhöhung der ALT-Werte.

Fazit

Das Risiko eines hepatischen unerwünschten Ereignisses war für Patienten unter 14 mg Teriflunomid-Behandlung gegenüber Placebo um das 1,7fache bzw. 1,5fache signifikant erhöht. Im Vergleich mit der zweckmäßigen Vergleichstherapie in der TENERE-Studie traten unter 14 mg Teriflunomid-Behandlung jedoch nur ein Viertel der unter Rebif® 44 µg beobachteten Ereignisse auf. Beide Wirkstoffe können die Leber schädigen, das Risiko ist für die Patienten unter Teriflunomid jedoch signifikant niedriger.

Da der Großteil der TEAE, die mit der Leberfunktion assoziiert waren, eine Erhöhung der ALT-Werte darstellte, sind die Häufigkeiten dieses PT in Tabelle 4-109 noch einmal übersichtlich dargestellt. Eine Erhöhung der Konzentration der Alanin-Aminotransferase im Blut kann ein Zeichen für eine heptozelluläre Schädigung sein, deswegen ist die Beobachtung der ALT-Werte ein Signal, dass eine leberschädigende Wirkung von Teriflunomid nicht ausgeschlossen werden kann. Dies ist in der Fachinformation und dem Risk-Management-Plan für Teriflunomid entsprechend berücksichtigt. Im Vergleich mit Rebif® ergibt sich für diesen PT alleine das gleiche Bild wie im Vergleich der mit der Leberfunktion assoziierten Ereignisse allgemein: Das Risiko ist für die mit 14 mg Teriflunomid behandelten Patienten höher als für unbehandelte Patienten (Placebo-Gruppe), jedoch deutlich niedriger als für Patienten, die mit Rebif® 44 µg behandelt werden.

Tabelle 4-109: Erhöhung der Alanin-Aminotransferase (ALT)-Werte – Ergebnisse für die gesamte Studienpopulation

Studie	Patientenzahl (Safety-Population) ^a	Primary SOC: Investigations Anzahl (prozentualer Anteil)	HLGT: Hepatobiliary investigations Anzahl (prozentualer Anteil)	HLT: Liver function analyses Anzahl (prozentualer Anteil)	PT: Alanine aminotransferase increased Anzahl (prozentualer Anteil)	Relative Risk (RR) 95% KI [] p-Wert ^b	Odds Ratio (OR) 95% KI [] p-Wert ^b
TEMSO	T14 (n=358)	102 (28,5 %)	64 (17,9 %)	64 (17,9 %)	51 (14,2 %)	2,137 [1,346; 3,394]	2,326 [1,398; 3,870]

					p = 0,0013	p = 0,0012	
	Pb	88 (n=360)	33 (9,2 %)	33 (9,2 %)	24 (6,7 %)		
TOWER	T14	111 (n=371)	67 (18,1%)	66 (17,8%)	52 (14,0 %)	1,686 [1,112; 2,558] p = 0,0140	1,798 [1,129; 2,865] p = 0,0135
	Pb	78 (n=385)	41 (10,6%)	40 (10,4%)	32 (8,3 %)		
TENERE ^c	T14	26 (n=110)	13 (11,8 %)	13 (11,8 %)	11 (10,0 %)	0,326 [0,173; 0,613] p = 0,0005	0,251 [0,118; 0,533] p = 0,0003
	Rb	41 (n=101)	37 (36,6 %)	36 (35,6 %)	31 (30,7 %)		

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bH₀: RR = 1 bzw. H₀: OR = 1

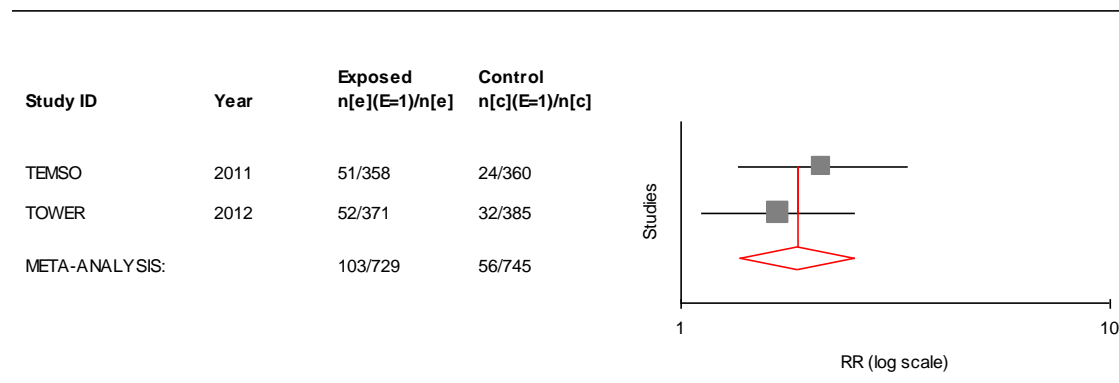
^c99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Die Studie TENERE lässt sich nicht mit den Studien TEMSO und TOWER zusammenfassen, während die letzteren beiden placebokontrollierte Studien sind, handelt es sich bei TENERE um eine aktivkontrollierte Studie, eine Metaanalyse ist nicht möglich.

Eine Metaanalyse der Erhöhung der ALT-Werte der Studien TEMSO und TOWER ist Bestandteil des entsprechenden indirekten Vergleiches, die Ergebnisse finden sich in Modul 4.3.2.

Abbildung 4-22: Forest-Plot der Metaanalyse von TEMSO und TOWER für Erhöhung der ALT-Werte (basierend auf relativen Risiken)



4.3.1.3.11.7 Gastrointestinale Beschwerden

In der Fachinformation von Teriflunomid sind Gastrointestinale Beschwerden (Übelkeit und Durchfall) als ein sehr häufiges unerwünschtes Ereignis genannt [33].

In der TOWER- und in der TENERE-Studie sind sie als AESI prädefiniert.

Tabelle 4-110: Gastrointestinale Beschwerden – Ergebnisse für die gesamte Studienpopulation

Gastrointestinale Beschwerden									
Studie	Patientenzahl (Safety-Population) ^a	Primary SOC: Gastrointestinal Disorders Anzahl (prozentualer Anteil)	Relative Risk (RR) 95% KI [] p-Wert ^b Anzahl (prozentualer Anteil)	Odds Ratio (OR) 95% KI [] p-Wert ^b Anzahl (prozentualer Anteil)	HLGT: Gastrointestinal motility and defaecation conditions Anzahl (prozentualer Anteil)	HLT: Diarrhoea (excl. Infective) Anzahl (prozentualer Anteil)	PT: Diarrhoea Anzahl (prozentualer Anteil)	Relative Risk (RR) 95% KI [] p-Wert ^b	Odds Ratio (OR) 95% KI [] p-Wert ^b
TEM SO	T14 (n=358)	158 (44,1 %)	1,313 [1,090; 1,582] p = 0,0041	1,560 [1,153; 2,111] p = 0,0039	76 (21,2 %)	64 (17,9 %)	64 (17,9%)	2,011 [1,350; 2,995] p = 0,0006	2,231 [1,419; 3,509] p = 0,0005
	Pb (n=360)	121 (33,6%)			46 (12,8%)	32 (8,9 %)	32 (8,9%)		
TOWER	T14 (n=371)	121 (32,6 %)	1,038 [0,843; 1,277] p = 0,7268	1,056 [0,778; 1,434] p = 0,7268	52 (14,0 %)	41 (11,1%)	41 (11,1%)	1,520 [0,960; 2,404] p = 0,0739	1,584 [0,958; 2,620] p = 0,0732
	Pb (n=385)	121 (31,4 %)			45 (11,7%)	29 (7,5 %)	28 (7,3%)		

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

TENERE ^c	T14 (n=110)	41 (37,3 %)	1,394 [0,931; 2,088] p = 0,1066	1,629 [0,906; 2,927] p = 0,1029	24 (21,8 %)	23 (20,9 %)	23 (20,9 %)	2,640 [1,237; 5,632] p = 0,0121	3,073 [1,306; 7,233] p = 0,0101
	Rb (n=101)	27 (26,7 %)			13 (12,9 %)	8 (7,9 %)	8 (7,9 %)		
Studie	Patienten- zahl (Safety- Population)^a	Primary SOC: Gastro- intestinal Disorders Anzahl (prozentualer Anteil)		HLGT: Gastro- intestinal signs and symptoms Anzahl (prozentualer Anteil)	HLT:Nause a and Vomiting symtoms Anzahl (prozentual er Anteil)	PT: Nausea Anzahl (prozentual er Anteil)	Relative Risk (RR) 95% KI [] p-Wert^b	Odds Ratio (OR) 95% KI [] p-Wert^b	
TEM SO	T14 (n=358)	158 (44,1 %)		101 (28,2 %)	59 (16,5 %)	49 (13,7%)	1,895 [1,205; 2,980] p = 0,0056	2,037 [1,236; 3,359] p = 0,0053	
	Pb (n=360)	121 (33,6 %)		86 (23,9 %)	34 (9,4 %)	26 (7,2%)			
TOWER	T14 (n=371)	121 (32,6 %)		84 (22,6 %)	45 (12,1 %)	38 (10,2%)	1,160 [0,747; 1,801] p = 0,5091	1,178 [0,724; 1,916] p = 0,5090	
	Pb (n=385)	121 (31,4 %)		75 (19,5 %)	38 (9,9 %)	34 (8,8%)			
TENERE ^c	T14 (n=110)	41 (37,3 %)		25 (22,7 %)	16 (14,5 %)	10 (9,1 %)	2,295 [0,743; 7,089] p = 0,1487	2,425 [0,736; 7,993] p = 0,1455	
	Rb	27		14	5	4			

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

(n=101)	(26,7 %)	(13,9 %)	(5,0 %)	(4,0 %)
---------	----------	----------	---------	---------

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bH₀: RR = 1 bzw. H₀: OR = 1

^c99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

TEMZO

Für beide PTs (Übelkeit und Durchfall) ist das relative Risiko in der 14 mg Teriflunomid-Gruppe ungefähr doppelt so hoch wie in der Placebo-Gruppe (relatives Risiko bezogen auf den PT Diarrhoea: 2,011; 95 % KI: [1,350; 2,995] und RR bezogen auf den PT Nausea: 1,895; 95 % KI: [1,205; 2,980]). Gastrointestinale Beschwerden führten nur in seltenen Fällen zum Behandlungsabbruch (1 Patient brach die Therapie aufgrund von Durchfall ab, kein Patient beendete die Therapie wegen Übelkeit).

TOWER

Auch in der TOWER-Studie litten in der 14 mg Teriflunomid-Gruppe tendenziell mehr Patienten an Durchfall als in der Placebo-Gruppe, das relative Risiko war um das 1,5fache erhöht (relatives Risiko 1,520 95 % KI: [0,960; 2,404]), mit einem entsprechenden p-Wert von 0,0739 (nicht signifikant). Bezüglich des PT Übelkeit (Nausea) gab es keinen nachweisbaren Unterschied zwischen der 14 mg Teriflunomid- und der Placebo-Gruppe (RR bezogen auf den PT Nausea: 1,160; 95 % KI: [0,747; 1,801]). In der TOWER-Studie beendeten jeweils 3 Patienten der 14 mg Teriflunomid-Gruppe die Studie wegen Durchfall und Übelkeit.

TENERE

In der TENERE-Studie zeigte sich mit einem relativen Risiko von 2,640 (95 % KI: [1,237; 5,632]) im Vergleich von 14 mg Teriflunomid und Rebif[®] 44 µg ein signifikant selteneres Auftreten von Durchfall (PT Diarrhoea) in der Rebif[®] 44 µg-Gruppe. Übelkeit war unter 14 mg Teriflunomid ebenfalls häufiger als unter Rebif[®] 44 µg, der Unterschied war aber nicht signifikant (RR von 2,295 mit einem 95 % KI: [0,743; 7,089]). Die Konfidenzintervalle der relativen Risiken sind in den Auswertungen der TENERE-Studie wesentlich breiter als in der TEMZO- und der TOWER-Studie, was auf die geringere Patientenzahl zurückzuführen ist. Von den aufgetretenen Fällen von Übelkeit und Erbrechen waren keine schwerwiegend und die meisten Fälle an Durchfall waren moderater Natur [35].

Das Verzerrungspotential ist in Bezug auf Durchfall und Übelkeit (PTs Diarrhoea und Nausea) als hoch einzustufen, da es sich um Ereignisse handelt, die durch die subjektive Einschätzung der Patienten stark beeinflusst werden können. Dementsprechend ist das Ergebnis möglicherweise zu Ungunsten der oralen Therapieform (14 mg Teriflunomid) systematisch verzerrt. Von einem leicht erhöhten Risiko für Durchfall und Übelkeit unter der Therapie mit 14 mg Teriflunomid ist dennoch auszugehen.

Fazit

Gastrointestinale Beschwerden als unerwünschtes Ereignis gehörten zu den häufigsten unerwünschten Ereignissen der placebokontrollierten Studien TEMZO und TOWER. Durchfall und Übelkeit sind auch in der direkt vergleichenden Studie TENERE unter Teriflunomid häufiger aufgetreten als unter der zweckmäßigen Vergleichstherapie Rebif[®] 44

µg, ein erhöhtes Risiko für diese nicht schwerwiegende Nebenwirkung ist trotz des möglicherweise relevant verzerrten Ergebnisses anzunehmen.

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Die Studie TENERE lässt sich nicht mit den Studien TEMSO und TOWER zusammenfassen, während die letzteren beiden placebokontrollierte Studien sind, handelt es sich bei TENERE um eine aktivkontrollierte Studie, eine Metaanalyse ist nicht möglich.

Die Metaanalysen für Gastrointestinale Beschwerden der Studien TEMSO und TOWER sind im Folgenden dargestellt.

Metaanalyse für Gastrointestinale Beschwerden (SOC)

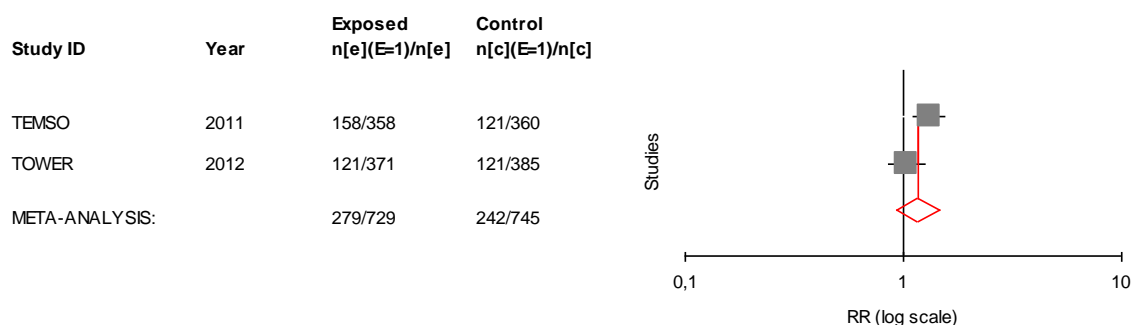
Tabelle 4-111: Metaanalyse-Ergebnisse für Gastrointestinale Beschwerden (SOC) aus TEMSO und TOWER – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse von TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	DL ^a :1,173 [0,931; 1,477] p = 0,1751 I ² =63,4%,p _Q =0,0982	DL ^a :1,285 [0,876; 1,883] p = 0,1996 I ² =68,5%,p _Q =0,0750

^aaufgrund der Heterogenität ist $I^2 > 0$, so dass die Metaanalyse mit Studie als zufälligem Effekt (nach DerSimonian und Laird (=DL)) andere Ergebnisse liefert als das Modell mit Studie als festem Effekt. Verwendet und interpretiert wurden in diesen Fällen ausschließlich die Ergebnisse nach DerSimonian und Laird

^bH₀: RR = 1 bzw. H₀: OR = 1

Abbildung 4-23: Forest-Plot der Metaanalyse von TEMSO und TOWER für Gastrointestinale Beschwerden (SOC) (basierend auf relativen Risiken)



Metaanalyse für Durchfall (PT Diarrhoea)

Tabelle 4-112: Metaanalyse-Ergebnisse für Durchfall (PT Diarrhoea) aus TEMSO und TOWER – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^a	Odds Ratio 95 % KI p-Wert ^a
Metaanalyse von TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	1,783 [1,320; 2,409] p = 0,0002 I ² =0%, p _Q =0,3659	1,914 [1,367; 2,680] p = 0,0002 I ² =0%, p _Q =0,3212

^aH₀: RR = 1 bzw. H₀: OR = 1

Abbildung 4-24: Forest-Plot der Metaanalyse von TEMSO und TOWER für Durchfall (PT Diarrhoea) (basierend auf relativen Risiken)

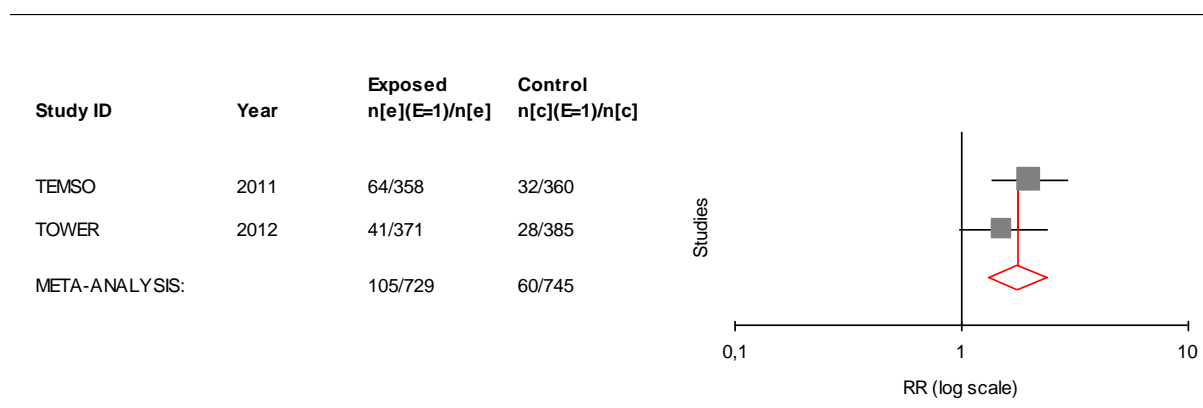
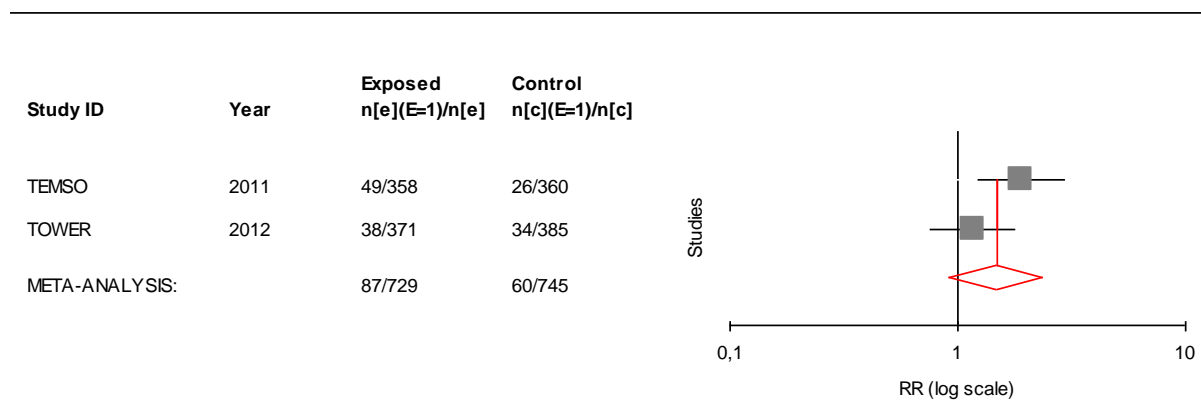
**Metaanalyse für Übelkeit (PT Nausea)**

Tabelle 4-113: Metaanalyse-Ergebnisse für Übelkeit (PT Nausea) aus TEMSO und TOWER –für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse von TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	DL ^a :1,478 [0,914; 2,392] p = 0,1114 I ² =57,0 %, p _Q = 0,1274	DL ^a : 1,544 [0,903; 2,641] p = 0,1126 I ² =57,8%, p _Q = 0,1239

^aaufgrund der Heterogenität ist I²>0, so dass die Metaanalyse mit Studie als zufälligem Effekt (nach DerSimonian und Laird (=DL)) andere Ergebnisse liefert als das Modell mit Studie als festem Effekt. Verwendet und interpretiert wurden in diesen Fällen ausschließlich die Ergebnisse nach DerSimonian und Laird^bH₀: RR = 1 bzw. H₀: OR = 1

Abbildung 4-25: Forest-Plot der Metaanalyse von TEMSO und TOWER für Übelkeit (PT Nausea) (basierend auf relativen Risiken)



4.3.1.3.11.8 Infektionen und Infestationen

Teriflunomid und Interferon beta greifen immunmodulatorisch in das Krankheitsgeschehen ein. Vor diesem Hintergrund sollten Infektionen und schweren Infektionen besondere Aufmerksamkeit gewidmet werden [2]. Infektionen werden als MedDRA[®]-Systemorganklasse (SOC) „Infektionen und Infestationen“ ausgewertet. Diese ist als ausreichend messsicher anzusehen, da es sich bei dieser SOC um eine sogenannte primary SOC handelt, d. h., dass bei der Klassifizierung von unerwünschten Ereignissen alle mit Infektionen zusammenhängenden Ereignisse dieser SOC zugeordnet werden.

Tabelle 4-114: Infektionen und Infestationen – Ergebnisse für die gesamte Studienpopulation

Studie	Patientenzahl (Safety- Population) ^a	Primary SOC: Infections and Infestations Anzahl (prozentualer Anteil)	Relative Risk (RR) 95% KI [] p-Wert ^b	Odds Ratio (OR) 95% KI [] p-Wert ^b
TEMSO	T14 (n=358)	222 (62,0%)	1,068 [0,948; 1,204] p = 0,2797	1,179 [0,875; 1,590] p = 0,2794
	Pb (n=360)	209 (58,1%)		
TOWER	T14 (n=371)	165 (44,5%)	0,869 [0,748; 1,010] p = 0,0666	0,764 [0,574; 1,018] p = 0,0657
	Pb (n=385)	197 (51,2%)		
TENERE ^c	T14 (n=110)	54 (49,1 %)	1,055 [0,795; 1,400] p = 0,7108	1,108 [0,645; 1,903] p = 0,7104
	Rb (n=101)	47 (46,5 %)		

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bH₀: RR = 1 bzw. H₀: OR = 1

^c99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

TEMZO & TOWER

In den beiden placebokontrollierten Studien TEMSO und TOWER konnte keine erhöhte Infektionsanfälligkeit durch die Therapie mit 14 mg Teriflunomid nachgewiesen werden. Dies zeigt sich durch ein jeweiliges relatives Risiko von 1,068 (95 % KI: [0,948; 1,204]) in der TEMSO- und 0,869 (95 % KI: [0,748; 1,010]) in der TOWER-Studie. Da innerhalb einer Studiendauer von 1 – 2 Jahren auch unabhängig von der Medikation im natürlichen Verlauf eine große Anzahl an Infektionen auftreten, sind die absoluten Häufigkeiten dieser Ereignisse mit 44 % bis 62 % sehr hoch. Dementsprechend kann die Aussage einer sicheren Anwendung nur vor dem Hintergrund getroffen werden, dass auch schwerwiegende unerwünschte Ereignisse, die der SOC „Infections and Infestations“ zugeordnet sind, in den beiden Behandlungsgruppen 14 mg Teriflunomid und Placebo mit ähnlicher Häufigkeit aufgetreten sind. In der TEMSO-Studie litten 2,5 % der Patienten in der 14 mg Teriflunomid-Gruppe und 2,2 % der Patienten in der Placebo- Gruppe an einer entsprechenden Infektion [32]. In der TOWER-Studie waren schwerwiegende Infektionen bei 3,0 % der Patienten in der 14 mg Teriflunomid-Gruppe und bei 2,9 % der Patienten in der Placebo-Gruppe aufgetreten, waren also auch über die Behandlungsgruppen gleichverteilt [31].

TENERE

In der TENERE-Studie sind Infektionen und Infestationen in den beiden Behandlungsgruppen 14 mg Teriflunomid und Rebif[®] 44 µg in ähnlicher Häufigkeit aufgetreten (das relative Risiko betrug 1,055 (95 % KI: [0,795; 1,400])).

Fazit

Es gibt keinen Hinweis darauf, dass das Infektionsrisiko unter Behandlung mit 14 mg Teriflunomid erhöht ist. Es konnte kein Unterschied zu der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg festgestellt werden.

In diesem Zusammenhang, der Aufrechterhaltung der Aufgaben des Immunsystems unter immunmodulatorischer Therapie, interessant ist das Ergebnis der TERIVA-Studie. Diese Studie untersuchte die Immunreaktion auf saisonale Grippeimpfungen von mit Teriflunomid-behandelten RMS-Patienten, die aus den Verlängerungsstudien der Phase II-Studie und der TEMSO-Studie (mediane Behandlungsdauer 5,7 Jahre) stammten. Die Teriflunomid-Patienten entwickelten, wie die Interferon-beta-Patienten einer Referenzpopulation, eine angemessene Immunreaktion auf die Grippeimpfung [33, 68].

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die

Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Die Studie TENERE lässt sich nicht mit den Studien TEMSO und TOWER zusammenfassen, während die letzteren beiden placebokontrollierte Studien sind, handelt es sich bei TENERE um eine aktivkontrollierte Studie, eine Metaanalyse ist hier nicht möglich.

Die Metaanalyse von TEMSO und TOWER bezüglich Infektionen und Infestationen zeigt mit Werten von $I^2=77,5\%$ und $p_Q=0,035$ ein heterogenes Bild. Das heterogene Bild resultiert aus der höheren Rate an Infektionen, die in der Placebo-Gruppe der TOWER-Studie aufgetreten ist, einen Vorteil für Teriflunomid kann man daraus jedoch nicht ableiten. Da die „Richtung“ des Ergebnisses somit kein Unterschied zwischen den Interventionen ist, kann man durchaus von gleichgerichteten Effekten sprechen. Die Heterogenität der Metaanalyse liefert kein Grund das Ergebnis der Einzelstudien in Frage zu stellen.

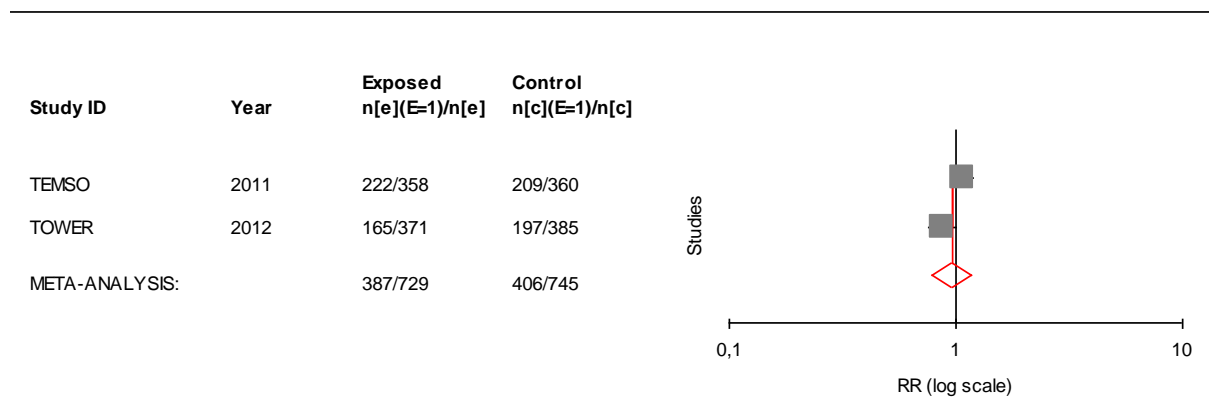
Tabelle 4-115: Metaanalyse-Ergebnisse für Infektionen und Infestationen (SOC) aus TEMSO und TOWER –für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse von TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	DL ^a : 0,969 [0,792; 1,185] p = 0,7559 $I^2=77,5\%$, $p_Q=0,0350$	DL ^a : 0,947 [0,619; 1,449] p = 0,8029 $I^2=76,3\%$, $p_Q=0,0400$

^aaufgrund der Heterogenität ist $I^2>0$, so dass die Metaanalyse mit Studie als zufälligem Effekt (nach DerSimonian und Laird (=DL)) andere Ergebnisse liefert als das Modell mit Studie als festem Effekt. Verwendet und interpretiert wurden in diesen Fällen ausschließlich die Ergebnisse nach DerSimonian und Laird

^b H_0 : RR = 1 bzw. H_0 : OR = 1

Abbildung 4-26: Forest-Plot der Metaanalyse von TEMSO und TOWER für Infektionen und Infestationen (SOC) (basierend auf relativen Risiken)



4.3.1.3.11.9 Bluthochdruck, Kopfschmerzen und periphere Neuropathie

Weitere unerwünschte Ereignisse von besonderem Interesse sind Bluthochdruck, Kopfschmerzen und periphere Neuropathie. Dabei waren Bluthochdruck und periphere Neuropathie in den Studienprotokollen prädefinierte Endpunkte. Im Rahmen der Termine zu „adverse event reportings“ wurden gezielt Fragen betreffend peripherer Neuropathie gestellt. Kopfschmerzen werden zusätzlich erhoben, weil diese in der Fachinformation der zweckmäßigen Vergleichstherapie Rebif® 44 µg als „sehr häufig“ aufgeführt sind [34].

Tabelle 4-116: Bluthochdruck – Ergebnisse für die gesamte Studienpopulation

Studie	Patienten- zahl (Safety- Popu- lation) ^a	Primary SOC: Vascular Disorders Anzahl (prozentu- aler Anteil)	HLGT: Vascular hyper- tensive Disorders Anzahl (prozentu- aler Anteil)	HLT: Vascluar hyper- tensive Disorders NEC ^b Anzahl (prozentu- aler Anteil)	PT: Hyper- tension Anzahl (prozentu- aler Anteil)	Relative Risk (RR) 95% KI [] p-Wert ^c	Odds Ratio (OR) 95% KI [] p-Wert ^c
TEM SO	T14 (n=358)	34 (9,5 %)	14 (3,9 %)	14 (3,9 %)	13 (3,6%)	2,179 [0,837; 5,669] p = 0,1104	2,223 [0,836; 5,916] p = 0,1096
	Pb (n=360)	20 (5,6 %)	6 (1,7 %)	6 (1,7 %)	6 (1,7%)		
TOWER	T14 (n=371)	26 (7,0 %)	16 (4,3 %)	15 (4,0%)	15 (4,0%)	1,946 [0,835; 4,535] p = 0,1231	1,986 [0,832; 4,740] p = 0,1224
	Pb (n=385)	19 (4,9 %)	9 (2,3 %)	8 (2,1%)	8 (2,1%)		
TENERE ^d	T14 (n=110)	8 (7,3 %)	6 (5,5 %)	5 (4,5 %)	5 (4,5 %)	1,148 [0,317; 4,156] p = 0,8338	1,155 [0,301; 4,425] p = 0,8337
	Rb (n=101)	7 (6,9 %)	4 (4,0 %)	4 (4,0 %)	4 (4,0 %)		

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bNEC = not elsewhere classified

^c(H₀: RR = 1) bzw. (H₀: OR = 1)

^d99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Tabelle 4-117: Kopfschmerzen und Periphere Neuropathie – Ergebnisse für die gesamte Studienpopulation

Studie	Patienten- zahl (Safety- Population) ^a	Primary SOC: Nervous system disorders Anzahl (prozentualer Anteil)	HLGT: Headaches Anzahl (prozentualer Anteil)	Relative Risk (RR) 95% KI [] p-Wert ^b	Odds Ratio (OR) 95% KI [] p-Wert ^b	HLGT: Peripheral neuropathies Anzahl (prozentualer Anteil)	Relative Risk (RR) 95% KI [] p-Wert ^b	Odds Ratio (OR) 95% KI [] p-Wert ^b
TEMSO	T14 (n=358)	151 (42,2 %)	74 (20,7%)	1,006 [0,755; 1,340] p = 0,9697	1,007 [0,701; 1,446] p = 0,9697	15 (4,2%)	5,028 [1,468; 17,22] p = 0,0101	5,204 [1,493; 18,14] p = 0,0096
	Pb (n=360)	145 (40,3 %)	74 (20,6%)			3 (0,8%)		
TOWER	T14 (n=371)	122 (32,9 %)	54 (14,6%)	1,057 [0,744; 1,502] p = 0,7557	1,067 [0,709; 1,606] p = 0,7557	13 (3,5%)	1,124 [0,520; 2,432] p = 0,7661	1,129 [0,508; 2,507] p = 0,7661
	Pb (n=385)	122 (31,7 %)	53 (13,8%)			12 (3,1%)		
TENERE ^c	T14 (n=110)	30 (27,3 %)	18 (16,4 %)	0,570 [0,338; 0,961] p = 0,0349	0,486 [0,250; 0,944] p = 0,0331	5 (4,5 %)	4,591 [0,546; 38,63] p = 0,1608	4,762 [0,547; 41,48] p = 0,1576
	Rb (n=101)	42 (41,6 %)	29 (28,7 %)			1 (1,0 %)		

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bH₀: RR = 1 bzw. H₀: OR = 1

^c99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

TEMSO, TOWER und TENERE

In keiner der drei Studien TEMSO, TOWER und TENERE hat sich bezüglich der unerwünschten Ereignisse Bluthochdruck und Kopfschmerzen ein signifikantes Ergebnis zu Gunsten oder Ungunsten einer der Interventionen ergeben. Periphere Neuropathien sind in der TEMSO-Studie in der 14 mg Teriflunomid-Gruppe häufiger aufgetreten als in der Placebo-Gruppe.

Bluthochdruck

Das Risiko für Bluthochdruck scheint mit relativen Risiken von 2,179 (95 % KI: [0,837; 5,669]) in der TEMSO-Studie und 1,946 (95 % KI: [0,835; 4,535]) in der TOWER-Studie unter Therapie mit 14 mg Teriflunomid höher zu sein als unter Placebo. Der Umfang der Studien war nicht groß genug, um eine entsprechende Aussage mit ausreichender statistischer Sicherheit zu machen. Im Sinne einer qualitätsgesicherten Anwendung ist ein entsprechendes Monitoring des Blutdrucks in der Fachinformation verankert [33]. Im Vergleich mit der zweckmäßigen Vergleichstherapie Rebif® 44 µg konnte kein Unterschied gezeigt werden.

Periphere Neuropathie

Periphere Neuropathie ist in der TEMSO-Studie in der 14 mg Dosierung signifikant häufiger aufgetreten als in der Placebo-Gruppe (relatives Risiko von 5,028; 95 % KI: [1,468; 17,22]). In der TOWER-Studie konnte für Teriflunomid kein erhöhtes Risiko bezüglich dieses unerwünschten Ereignisses festgestellt werden. In der TENERE-Studie sind in der 14 mg Teriflunomid-Gruppe mehr periphere Neuropathien aufgetreten als unter Rebif® 44 µg [33]. Der Unterschied zur zweckmäßigen Vergleichstherapie war jedoch nicht statistisch signifikant (relatives Risiko von 4,591; 95 % KI: [0,546; 38,63]).

Ein Symptom für periphere Neuropathie kann das Auftreten von Kribbeln oder anderen Sensibilitätsstörungen in den Extremitäten, sogenannte Parästhesien sein. Diese Parästhesien sind weniger klar gefasst als eine periphere Neuropathie und stellen selbst auch ein sehr häufiges Symptom der MS dar [69]. Parästhesien sind in der Fachinformation zu Teriflunomid (AUBAGIO®) als sehr häufiges unerwünschtes Ereignis aufgeführt, ein statistisch signifikanter Unterschied zwischen Teriflunomid und Placebo oder Teriflunomid und Rebif® 44 µg konnte bei den Patienten in der Studie jedoch nicht gezeigt werden (s. Tabelle 4-118), so dass sich die detaillierte Auswertung im Rahmen dieses Dokuments auf den PT „peripheral neuropathies“ beschränkt.

Tabelle 4-118: Parästhesie – Ergebnisse für die gesamte Studienpopulation

Studie	Patientenzahl (Safety- Population) ^a	PT Paraesthesia Anzahl (prozentualer Anteil)	Relative Risk (RR) 95% KI [] p-Wert ^b	Odds Ratio (OR) 95% KI [] p-Wert ^b
TEMSO	T14 (n=358)	35 (9,8 %)	1,173 [0,737; 1,868] p = 0,5010	1,192 [0,715; 1,987] p = 0,5008

	Pb (n=360)	30 (8,3 %)		
TOWER	T14 (n=371)	22 (5,9 %)	0,951 [0,543; 1,666] p = 0,8613	0,948 [0,522; 1,722] p = 0,8613
	Pb (n=385)	24 (6,2 %)		
TENERE ^c	T14 (n=110)	11 (10,0 %)	1,263 [0,529; 3,013] p = 0,5994	1,292 [0,498; 3,352] p = 0,5989
	Rb (n=101)	8 (7,9%)		

^as. Definition der Safety-Population unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)

^bH₀: RR = 1 bzw. H₀: OR = 1

^c99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Kopfschmerzen

In den placebokontrollierten Teriflunomid-Studien konnte für 14 mg Teriflunomid keine gegenüber Placebo erhöhte Rate an Kopfschmerzen beobachtet (HLGT Headaches) werden (s. Tabelle 4-117).

Kopfschmerzen sind in der TENERE-Studie in der 14 mg Teriflunomid-Gruppe mit einem relativen Risiko von 0,570 (95 % KI: [0,338; 0,961]) seltener aufgetreten als in der Rebif[®] 44 µg-Gruppe.

Kopfschmerzen sind kein typisches Symptom der MS (s. Modul 3), aber es gibt Hinweise, dass Kopfschmerzen sowohl zu Beginn als auch im Verlauf der Erkrankung häufiger auftreten als in der nicht betroffenen Bevölkerung. Dabei sind Spannungskopfschmerzen und Migräne ohne Aura die häufigsten Formen, wobei auch Neuralgien und Clusterkopfschmerzen beschrieben worden sind (in diesen Fällen könnte auch die Lokalisation der MS-Läsionen eine maßgebliche Rolle spielen) [70]. Auch krankheitsmodifizierende Wirkstoffe wie die Beta-Interferone, können Kopfschmerzen auslösen oder verstärken. Mit Hinblick auf das Ziel, den Patienten eine bessere Compliance zu ermöglichen, ist die Einschätzung des Kopfschmerz-auslösenden Potentials einer Dauertherapie ein wesentlicher Bestandteil der Bewertung. Der statistisch signifikant Vorteil in Bezug auf dieses unerwünschte Ereignis stellt dementsprechend einen für die Patienten im praktischen Alltag äußerst relevanten Vorteil für Teriflunomid dar.

Fazit:

In Bezug auf die unerwünschten Ereignisse Bluthochdruck und periphere Neuropathie konnte weder ein erhöhtes Risiko für die Therapie mit 14 mg Teriflunomid mit entsprechender

statistischer Sicherheit gezeigt werden noch ein Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie Rebif® 44 µg.

In Bezug auf Kopfschmerzen hat 14 mg Teriflunomid einen signifikanten Vorteil gegenüber Rebif® 44 µg gezeigt. Dies ist konsistent mit den Angaben in der Fachinformation von Rebif® 44 µg, in der Kopfschmerzen als sehr häufiges unerwünschtes Ereignis aufgeführt sind [34]. Für MS-Patienten, die durch ihre Erkrankung stark in ihrer Lebensqualität eingeschränkt sind und die unter Müdigkeit und Konzentrationsstörungen leiden, ist eine Vermeidung von Kopfschmerzen ein wesentlicher Vorteil einer Therapie.

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Einzelergebnisse mithilfe von Metaanalysen quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalysen (in der Regel als Forest-Plot) dar. Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Die Studie TENERE lässt sich nicht mit den Studien TEMSO und TOWER zusammenfassen, während die letzteren beiden placebokontrollierte Studien sind, handelt es sich bei TENERE um eine aktivkontrollierte Studie, eine Metaanalyse ist hier nicht möglich.

Bluthochdruck

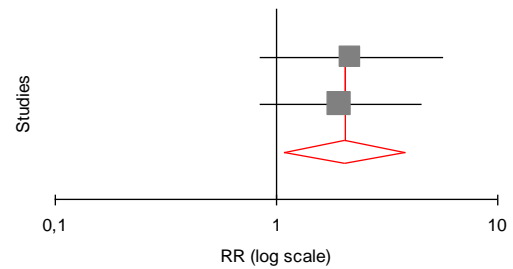
Tabelle 4-119: Metaanalyse-Ergebnisse für Bluthochdruck aus TEMSO und TOWER –für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^a	Odds Ratio 95 % KI p-Wert ^a
Metaanalyse von TEMSO & TOWER	Teriflunomid 14 mg versus Placebo	2,045 [1,085; 3,854] p = 0,0269 I ² =0%, p _Q =0,8622	2,087 [1,089; 3,999] p = 0,0266 I ² =0%, p _Q =0,8657

^aH₀: RR = 1 bzw. H₀: OR = 1

Abbildung 4-27: Forest-Plot der Metaanalyse von TEMSO und TOWER für Bluthochdruck (basierend auf relativen Risiken)

Study ID	Year	Exposed n[e](E=1)/n[e]	Control n[c](E=1)/n[c]
TEMZO	2011	13/358	6/360
TOWER	2012	15/371	8/385
META-ANALYSIS:		28/729	14/745

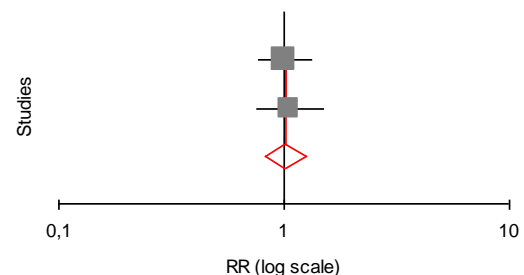


Kopfschmerzen

Die Metaanalyse der Studienergebnisse aus TEMZO und TOWER zu Kopfschmerzen sind Bestandteil des indirekten Vergleichs und unter 4.3.2.1.4.5 dargestellt.

Abbildung 4-28: Forest-Plot der Metaanalyse von TEMZO und TOWER für Kopfschmerzen (basierend auf relativen Risiken)

Study ID	Year	Exposed n[e](E=1)/n[e]	Control n[c](E=1)/n[c]
TEMZO	2011	74/358	74/360
TOWER	2012	54/371	53/385
META-ANALYSIS:		128/729	127/745



Periphere Neuropathie

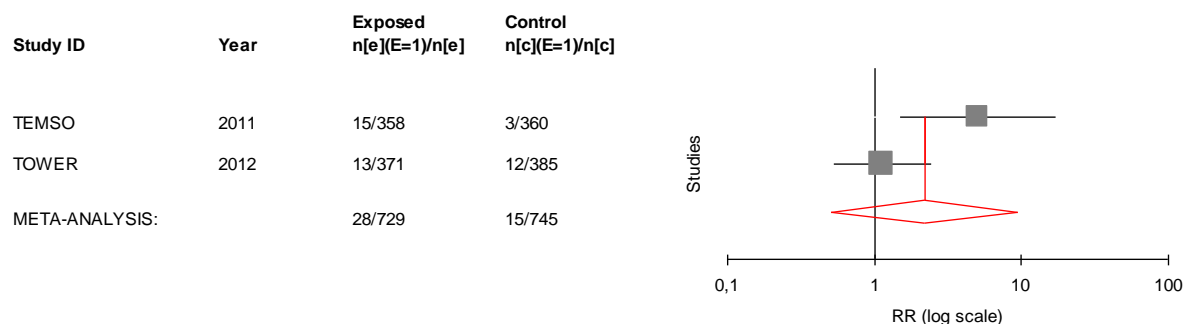
Tabelle 4-120: Metaanalyse-Ergebnisse für periphere Neuropathie (SOC) aus TEMZO und TOWER – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse von TEMZO & TOWER	Teriflunomid 14 mg versus Placebo	DL ^a : 2,195 [0,510; 9,447] p = 0,2911 I ² = 75,5 %, p _Q = 0,0433	DL ^a : 2,241 [0,505; 9,940] p = 0,2885 I ² = 75,5 %, p _Q = 0,0432

^aaufgrund der Heterogenität ist I²>0, so dass die Metaanalyse mit Studie als zufälligem Effekt (nach DerSimonian und Laird (=DL)) andere Ergebnisse liefert als das Modell mit Studie als festem Effekt. Verwendet und interpretiert wurden in diesen Fällen ausschließlich die Ergebnisse nach DerSimonian und Laird

^bH₀: RR = 1 bzw. H₀: OR = 1

Abbildung 4-29: Forest-Plot der Metaanalyse von TEMSO und TOWER für periphere Neuropathie (basierend auf relativen Risiken)



4.3.1.3.11.10 Zusammenfassung der Ergebnisse zu unerwünschten Ereignissen von besonderem Interesse (AESI)

Vergleich der 14 mg Teriflunomid-Dosierung mit Placebo

Die häufigsten unerwünschten Ereignisse, die mit der Teriflunomid-Behandlung in Verbindung gebracht werden konnten, waren Diarrhö, eine Erhöhung der ALT-Werte, Übelkeit und verminderte Haardichte (HLT Alopecias).

Tabelle 4-121: 14 mg Teriflunomid versus Placebo (Ergebnisse der TEMSO- und der TOWER-Studie)

AESI	Verzerrungs- potenzial	Vorteil Teriflunomid	Kein nachweisbarer Unterschied	Nachteil Teriflunomid
Reaktionen an der Injektionsstelle		e n t f ä l l t		
Grippeähnliche Symptome			x	
Verminderte Haardichte (HLT Alopecias)	niedrig			x
Leberfunktionsstörungen (inklusive Erhöhung der ALT-Werte)	niedrig			x
Gastrointestinale Beschwerden	niedrig			x
Neutropenie	niedrig			x
Infektionen und Infestationen	niedrig		x	
Überempfindlichkeit/Hautreaktionen	niedrig		x	
Bluthochdruck	niedrig		x	
Periphere Neuropathie	niedrig			x
Kopfschmerzen	niedrig		x	

Gegenüber Placebo ist Teriflunomid in Bezug auf alle „unerwünschten Ereignisse von besonderem Interesse“ entweder gleichwertig oder aber mit Nachteilen verbunden. In einer placebokontrollierten Studie entspricht dies dem erwarteten Ergebnis, ein von diesem Bild abweichendes Ergebnis wäre kritisch zu hinterfragen. Unerwünschte Ereignisse, die unter Therapie mit Teriflunomid signifikant häufiger auftreten als unter Placebo sind eine verminderte Haardichte (HLT Alopecias), Leberfunktionsstörungen, Gastrointestinale Beschwerden, Neutropenie und periphere Neuropathie. Jede dieser Nebenwirkungen war in ihrer Ausprägung relativ mild und keines dieser Ereignisse führt bei mehr als 3 % der Patienten zu einem Abbruch der Therapie.

Neutropenie und eine Erhöhung der ALT-Werte können ein erster Hinweis für eine erhöhte Infektanfälligkeit bzw. eine Leberschädigung sein. Im Rahmen einer qualitätsgesicherten Anwendung ist ein Monitoring der entsprechende Werte vorgesehen und auch in der Fachinformation verankert (ausführliche Darstellung in Modul 3.4 Anforderungen an eine qualitätsgesicherte Anwendung).

Insgesamt ist 14 mg Teriflunomid ein gut verträglicher Wirkstoff mit einem positiven Nutzen-Risiko-Verhältnis.

Vergleich der 14 mg Teriflunomid-Dosierung mit der zweckmäßigen Vergleichstherapie Rebif® 44 µg

Tabelle 4-122: 14 mg Teriflunomid versus Rebif® 44 µg (Ergebnisse der TENERE-Studie)

AESI	Verzerrungs- potential	Vorteil Teriflunomid	Kein nachweisbarer Unterschied	Nachteil Teriflunomid
Reaktionen an der Injektionsstelle	hoch ^a	x ^b		
Grippeähnliche Symptome	hoch ^a	x		
Verminderte Haardichte (HLT Alopecias)	niedrig			x
Leberfunktions- störungen (inklusive Erhöhung der ALT- Werte)	niedrig	x		
Gastrointestinale Beschwerden	hoch			x
Neutropenie	niedrig		x	
Infektionen und Infestationen	niedrig		x	
Bluthochdruck	niedrig		x	
Periphere	niedrig		x	

Neuropathie (inklusive Parästhesien)		
Kopfschmerzen	hoch	x

^aes wird nicht davon ausgegangen, dass die Ergebnisse für Reaktionen an der Injektionsstelle, grippeähnliche Symptome oder Kopfschmerzen relevant verzerrt sind. Das Verzerrungspotential wird dennoch als „hoch“ eingestuft, weil es sich nicht um vorab definierte AESI handelt.

^bDie Größe des Unterschieds lässt sich für Reaktionen an der Injektionsstelle nicht quantifizieren, da es sich aber um eine „sehr häufige“ Nebenwirkung von Rebif 44 µg handelt [34], die bei Therapie mit Teriflunomid nicht auftreten können, ist der Unterschied als bedeutsame Vermeidung eingestuft.

Alle unerwünschten Ereignisse, die bei der Therapie mit 14 mg Teriflunomid und Rebif[®] 44 µg als charakteristisch anzusehen sind, sind als nicht schwerwiegende Nebenwirkungen einzustufen. Von den unerwünschten Ereignissen, die im Vergleich zu Placebo gehäuft auftraten zeigen sich im Vergleich zu der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg für die unerwünschten Ereignisse verminderte Haardichte (HLT Alopecias) und Durchfall Nachteile für die 14 mg Teriflunomid-Therapie.

In der System Organ Class „Gastrointestinal Disorders“ insgesamt und bezogen auf Übelkeit gab es keinen signifikanten Nachteil der 14 mg Teriflunomid-Therapie, in der TOWER-Studie gab es in Bezug auf diese unerwünschten Ereignisse auch keinen Nachteil gegenüber Placebo. Periphere Neuropathien, Neutropenie und mit der Leberfunktion assoziierte unerwünschte Ereignisse wurden zwar als Sicherheitsaspekte identifiziert, die bei der Therapie mit Teriflunomid zu beachten sind, im Vergleich zu der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg ergab sich jedoch kein signifikanter Nachteil. In Bezug auf unerwünschte Ereignisse, die mit der Leberfunktion assoziiert sind (vor allem eine Erhöhung der ALT-Werte) zeigt 14 mg Teriflunomid sogar deutliche Vorteile gegenüber Rebif[®] 44 µg.

Auch unter Berücksichtigung der möglichen unerwünschten Ereignisse, die bei der Behandlung mit Teriflunomid auftreten können, hat die Therapie mit dem zu bewertenden Arzneimittel für die Dauertherapie einer chronischen Erkrankung erhebliche patientenrelevante Vorteile. Dies zeigt sich darin, dass die unter Rebif[®] 44 µg-Therapie sehr häufig auftretenden Reaktionen an der Injektionsstelle nicht beobachtbar sind und beträchtlich weniger grippeähnliche Symptome auftreten. Des Weiteren treten Kopfschmerzen in statistisch signifikantem Ausmaß weniger auf.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.12 Subgruppenanalysen – RCT

Für die tabellarische Darstellung der Ergebnisse aus Subgruppenanalysen gelten die gleichen Anforderungen wie für die tabellarische Darstellung von Ergebnissen aus Gesamtpopulationen in Abschnitt 4.3.1.3.

Beschreiben Sie die Ergebnisse von Subgruppenanalysen (einschließlich der Interaktionsterme). Stellen Sie dabei die Ergebnisse in den Subgruppen zunächst für jede einzelne Studie in tabellarischer Form dar. Diese Anforderung gilt sowohl für Subgruppenanalysen auf Basis individueller Patientendaten als auch für solche auf Basis aggregierter Daten.

Sofern die vorliegenden Studien bzw. Daten für eine Metaanalyse geeignet sind, fassen Sie die Ergebnisse mithilfe einer Metaanalyse quantitativ zusammen und stellen Sie die Ergebnisse der Metaanalyse (als Forest-Plot) dar.

Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Metaanalyse durchführen bzw. wenn Sie nicht alle Studien in die Metaanalyse einschließen.

Subgruppenanalysen haben den Sinn, die Ergebnisse einer Studie hinsichtlich potentieller Effektmodifikatoren zu untersuchen. Es besteht allgemeiner wissenschaftlicher Konsens darüber, dass die Interpretation von Subgruppenanalysen Einschränkungen hat und die Ergebnisse vor allem dann vorrangig deskriptiven Charakter haben, wenn keine Adjustierung für multiples Testen stattgefunden hat [17, 71, 72]. Für die Interpretierbarkeit von Subgruppenanalysen müssen demnach die folgenden Voraussetzungen erfüllt sein [72]:

- Die Subgruppenanalysen müssen Teil einer adäquat durchgeführten aussagekräftigen Studie sein.
- Die Subgruppenanalysen sollten a priori im Studienprotokoll geplant worden sein.
- Eine Korrektur für multiples Testen sollte durchgeführt und die Subgruppenanalysen nur auf wenige beschränkt sein.
- Eine alleinige Auswertung der separaten Subgruppen ist nicht geeignet, eine angemessene statistische Methodik, insbesondere Interaktionstests, ist gefordert.
- Die Subgruppenbildende Variable sollte eine Baseline-Variable darstellen.

In den vorliegenden Studien TEMSO, TOWER und TENERE sind nicht alle diese Kriterien erfüllt. Es handelt sich zwar überwiegend um vorab geplante Subgruppenanalysen, die Bestandteil einer adäquat durchgeführten Studie sind, aber es sind sehr viele Subgruppenanalysen durchgeführt worden, ohne dass eine Adjustierung für multiples Testen vorgenommen wurde. Darüber hinaus ist keine der Teriflunomid-Studien gepowert, um in Subgruppen signifikante Ergebnisse zu erzielen. Dementsprechend dienen alle angegebenen Subgruppenergebnisse lediglich der Identifizierung möglicher Effektmodifikatoren, ein Nachweis für Subgruppeneffekte kann auf Basis der vorliegenden Daten nicht erbracht werden.

Die in den Studienprotokollen a priori geplanten Subgruppenanalysen sind entsprechend den Anforderungen an das vorliegende Dossier in jedem Fall zu berichten. Für die Darstellung des

Nutzens und des Zusatznutzens von 14 mg Teriflunomid sind drei randomisierte kontrollierte Studien identifiziert worden, die die Einschlusskriterien erfüllen:

- TENERE
- TEMSO
- TOWER

Für diese drei Studien sind entsprechend der Nutzendossiervorlage die Ergebnisse der a priori geplanten Subgruppenanalysen zu berichten. Darüber hinaus sollen folgende Faktoren als potentielle Effektmodifikatoren berücksichtigt werden:

- Geschlecht
- Alter
- Krankheitsschwere bzw. – stadium

Diese Faktoren waren mit den a priori geplanten Subgruppenanalysen bereits abgedeckt und entsprechende Subgruppenanalysen mussten für keine der drei Studien posthoc durchgeführt werden.

4.3.1.3.12.1 Ergebnisse der Subgruppenanalysen der Studie TENERE

In der Studie TENERE waren für den primären Endpunkt „Time to Failure“ und den wesentlichen sekundären Endpunkt „jährliche Schubrate“ folgende prädefinierte Subgruppenanalysen vorgesehen:

- Demografische Faktoren (Alter, Geschlecht, Region)
- Klinische Krankheitsaktivität (Schubaktivität, Baseline EDSS-Wert)

Da die vorherige Einnahme krankheitsmodifizierender Wirkstoffe in der MS ein Effektmodifikator sein könnte, wurden die entsprechenden Subgruppenanalysen zusätzlich durchgeführt [73].

Für die Subgruppenanalysen des primären Endpunktes „Time to Failure“ wurde ein Cox Proportional-Hazards-Modell angewendet, für den sekundären Endpunkt „jährliche Schubrate“ ein Poisson-Modell, in das jeweils die Subgruppenzugehörigkeit und der Behandlung-Subgruppen-Interaktionsterm zusätzlich aufgenommen wurden. Jeder Subgruppenfaktor wurde separat analysiert. Das Modell beinhaltete somit die Faktoren Behandlung, Baseline EDSS-Wert ($\leq 3,5$ / $> 3,5$), Region, Subgruppe und Behandlung-Subgruppe-Interaktion und im Falle des Poisson-Modells zusätzlich die log-transformierte standardisierte Studiendauer als Offset-Variable.

Therapieversagen (Time to Failure)

Tabelle 4-123: Subgruppenanalysen für den primären Endpunkt Time to Failure in TENERE

Studie	Patientenzahl	Hazard Ratio [95 % KI] 14 mg Teriflunomid gegenüber Rebif®	Interaktionsterm
TENERE	Geschlecht		
	Männer	T14 (n = 33) Rb (n = 33)	0,669 [0,315; 1,425] p = 0,5314
	Frauen	T14 (n = 78) Rb (n = 71)	0,936 [0,559; 1,569]
	Altersgruppe		
	<38 Jahre	T14 (n = 62) Rb (n = 60)	0,902 [0,528; 1,538] p = 0,8121
	≥38 Jahre	T14 (n = 49) Rb (n = 44)	0,816 [0,406; 1,639]
	Region		
	Osteuropa	T14 (n = 41) Rb (n = 35)	1,190 [0,557; 2,542] p = 0,3957
	Westeuropa und Afrika	T14 (n = 64) Rb (n = 62)	0,822 [0,477; 1,416]
	Amerika	T14 (n = 6) Rb (n = 7)	0,228 [0,035; 1,463]
	Baseline EDSS-Wert		
	≤3,5	T14 (n = 95) Rb (n = 94)	0,815 [0,521; 1,276] p = 0,4759
	>3,5	T14 (n = 16) Rb (n = 10)	1,277 [0,305; 5,346]
	Anzahl Schübe in den vorangegangenen 2 Jahren		
	≤1	T14 (n = 48) Rb (n = 50)	0,890 [0,436; 1,817] p = 0,9534
	2	T14 (n = 41) Rb (n = 30)	0,782 [0,365; 1,673]
	3	T14 (n = 20) Rb (n = 18)	1,143 [0,491; 2,663]
	≥4	T14 (n = 2)	0,566 [0,060; 5,361]

	Rb (n = 6)		
	Vorbehandlung		
ja	T14 (n = 13)	0,488 [0,145; 1,639]	p = 0,4513
	Rb (n = 25)		
nein	T14 (n = 98)	0,920 [0,577; 1,468]	
	Rb (n = 79)		

Jährliche Schubrate

Tabelle 4-124: Subgruppenanalysen für den sekundären Endpunkt jährliche Schubrate in TENERE

Studie	Patientenzahl	Relative Risk [95 % KI] 14 mg Teriflunomid gegenüber Rebif®		Interaktionsterm
TENERE	Geschlecht			
	Männer	T14 (n = 33)	NA	p = 0,8985
		Rb (n = 33)		
	Frauen	T14 (n = 78)	1,159 [0,496; 2,705]	
		Rb (n = 71)		
	Altersgruppe			
	<38 Jahre	T14 (n = 62)	1,113 [0,496; 2,500]	p = 0,7345
		Rb (n = 60)		
	≥38 Jahre	T14 (n = 49)	1,421 [0,470; 4,299]	
		Rb (n = 44)		
	Region			
	Osteuropa	T14 (n = 41)	1,378 [0,424; 4,479]	NA
		Rb (n = 35)		
	Westeuropa und Afrika	T14 (n = 64)	1,233 [0,554; 2,740]	
	Rb (n = 62)			
Amerika	T14 (n = 6)	NA		
	Rb (n = 7)			
Baseline EDSS-Wert				
≤3,5	T14 (n = 95)	1,162 [0,585; 2,309]	p = 0,8194	
	Rb (n = 94)			
>3,5	T14 (n = 16)	1,565 [0,247; 9,929]		
	Rb (n = 10)			
Anzahl Schübe in den vorangegangenen 2 Jahren				

≤1	T14 (n = 48)	NA	NA
	Rb (n = 50)		
2	T14 (n = 41)	1,387 [0,343; 5,600]	
	Rb (n = 30)		
3	T14 (n = 20)	1,467 [0,574; 3,745]	
	Rb (n = 18)		
≥4	T14 (n = 2)	NA	
	Rb (n = 6)		
Vorbehandlung			
ja	T14 (n = 13)	1,068 [0,268; 4,252]	p = 0,8844
	Rb (n = 25)		
nein	T14 (n = 98)	1,248 [0,578; 2,695]	
	Rb (n = 79)		

Kommentar:

Weder für den primären Endpunkt „Time to Failure“ noch für den sekundären Endpunkt „jährliche Schubrate“ ergab sich in der Auswertung der gesamten Studienpopulation ein signifikanter Vorteil für 14 mg Teriflunomid, dieses Bild bestätigt sich auch in den Subgruppenanalysen; keiner der durchgeführten Interaktionstests liefert den Hinweis auf eine bedeutsame Interaktion. Ein auffällig abweichendes Ergebnis zeigt sich lediglich in der Subgruppe „Amerika“ bei der Untersuchung regional spezifischer Unterschiede, wo im Ergebnis zwar auch kein signifikanter Unterschied zu belegen war, jedoch die Hazard Ratio mit 0,228 deutlich niedriger lag als in den übrigen Subgruppen. Da es sich bei der betrachteten Subgruppe um eine sehr kleine Gruppe (n < 10 in jedem Arm) handelt, ist dieses Ergebnis kein Beleg für eine Interaktion.

4.3.1.3.12.2 Ergebnisse der Subgruppenanalysen der Studie TEMSO

In der TEMSO-Studie waren Subgruppenanalysen für die zwei wichtigsten klinischen Endpunkte, den primären Endpunkt Schübe (operationalisiert als jährliche Schubrate) und den wesentlichen sekundären Endpunkt Behinderungsprogression (operationalisiert als Zeit bis zur Behinderungsprogression) präspezifiziert. Diese vorab geplanten Subgruppenanalysen betrafen:

- Demografische Faktoren (Alter, Geschlecht, Region, Ethnische Zugehörigkeit)
- Klinische Krankheitsaktivität (Schubaktivität, MS-Form, Baseline EDSS-Wert)

- MRT-Parameter (Gadolinium-anreichernde Läsionen, Burden of Disease)

Für die Subgruppenanalysen des primären Endpunktes „jährliche Schubrate“ wurde ein Poisson-Modell angewendet, für den wesentlichen sekundären Endpunkt „Zeit bis zur Behinderungsprogression“ ein Cox Proportional-Hazards-Modell, in das jeweils die Subgruppenzugehörigkeit und der Behandlung-Subgruppen-Interaktionsterm zusätzlich aufgenommen wurden. Jeder Subgruppenfaktor wurde separat analysiert. Das Modell beinhaltet somit die Faktoren Behandlung, Baseline EDSS-Wert ($\leq 3,5$ / $> 3,5$), Region, Subgruppe und Behandlung-Subgruppe-Interaktion und im Falle des Poisson-Modells zusätzlich die log-transformierte standardisierte Studiendauer als Offset-Variable.

Jährliche Schubrate

Tabelle 4-125: Subgruppenanalysen für den Endpunkt jährliche Schubrate in TEMSO

Studie	Patientenzahl	Relative Risk ^a [95 % KI] 14 mg Teriflunomid gegenüber Placebo	Interaktionsterm ^b
TEMSO	Geschlecht		
	Männer	T14 (n=104) Pb (n=88)	0,786 [0,517; 1,193] p = 0,6173
	Frauen	T14 (n=254) Pb (n=275)	0,667 [0,521; 0,855]
	Alter		
	<38 Jahre	T14 (n=173) Pb (n=156)	0,638 [0,469; 0,868] p = 0,5499
	≥38 Jahre	T14 (n=185) Pb (n=207)	0,719 [0,542; 0,955]
	Ethnische Zugehörigkeit		
	White	T14 (n=346) Pb (n=356)	0,691 [0,558; 0,855] NA
	Black	T14 (n=1) Pb (n=3)	NA
	Asian/Oriental	T14 (n=8) Pb (n=1)	NA
	Multiracial	T14 (n=2) Pb (n=1)	NA
	Other	T14 (n=0) Pb (n=1)	NA
	Region		

Osteuropa	T14 (n=108) Pb (n=114)	0,800 [0,549; 1,165]	p = 0,1425
Westeuropa	T14 (n=166) Pb (n=170)	0,559 [0,417; 0,749]	
Amerika	T14 (n=83) Pb (n=82)	0,896 [0,535; 1,502]	
Baseline EDSS-Wert			
≤3,5	T14 (n=276) Pb (n=281)	0,606 [0,470; 0,780]	p = 0,0656
>3,5	T14 (n=82) Pb (n=82)	0,914 [0,621; 1,345]	
Anzahl Schübe in den vorangegangenen 2 Jahren			
≤1	T14 (n=71) Pb (n=71)	0,436 [0,230; 0,824]	p = 0,3913
2	T14 (n=191) Pb (n=186)	0,699 [0,523; 0,934]	
3	T14 (n=70) Pb (n=76)	0,610 [0,389; 0,959]	
≥4	T14 (n=26) Pb (n=30)	0,908 [0,529; 1,559]	
MS-Subtyp			
Secondary progressive and progressive relapsing MS	T14 (n = 26) Pb (n = 34)	0,985 [0,447; 2,172]	p = 0,3202
Relapsing remitting MS	T14 (n = 332) Pb (n = 329)	0,661 [0,531; 0,824]	
Vorbehandlung			
Ja	T14 (n = 102) Pb (n = 90)	0,602 [0,427; 0,849]	p = 0,5305
Nein	T14 (n = 256) Pb (n = 273)	0,702 [0,538; 0,916]	
Baseline-Wert: Anzahl Gadolinium-anreichernder Läsionen			
0	T14 (n = 230) Pb (n = 222)	0,732 [0,555; 0,966]	p = 0,7121
≥1	T14 (n = 125)	0,678 [0,497; 0,925]	

		Pb (n = 137)		
		Baseline-Wert: Burden of Disease		
<13 ml	T14 (n = 183)	0,721 [0,533; 0,976]	p = 0,5752	
	Pb (n = 183)			
≥13 ml	T14 (n = 172)	0,640 [0,475; 0,862]		
	Pb (n = 175)			

^aErhalten aus einem Poisson-Modell, das in der jeweiligen Subgruppe separat durchgeführt wurde, mit der Anzahl bestätigter Schübe, die zwischen Randomisierung und letzter Dosierung aufgetreten sind, als Responsevariable, Behandlung, EDSS-Baseline und Region als Kovariaten und log-transformierten standardisierten Studiendauern als Offset-Variable

^bErhalten aus einem Poisson-Modell mit der Anzahl bestätigter Schübe, die zwischen Randomisierung und letzter Dosierung aufgetreten sind, als Responsevariable, Behandlung, EDSS-Baseline, Region, Subgruppe und Behandlung-Subgruppe-Interaktion als Kovariaten und log-transformierten standardisierten Studiendauern als Offset-Variable.

Der vorteilhafte Effekt von 14 mg Teriflunomid auf die jährliche Schubrate war über die verschiedenen Patientensubgruppen konsistent. Mit Interaktionstermen von $p = 0,1425$ und $p = 0,0656$ können die Region und der Baseline-EDSS-Wert als mögliche Effektmodifikatoren identifiziert werden. Die Studie war nicht gepowert, um Subgruppen zu untersuchen, so dass die Daten nur auf einer deskriptiven Basis betrachtet werden können. Obwohl es das Vorliegen von Effektmodifikatoren nicht mit Sicherheit ausgeschlossen werden kann, zeigt 14 mg Teriflunomid in allen Behandlungsgruppen einen überlegenen Effekt.

Zeit bis zur Behinderungsprogression

Tabelle 4-126: Subgruppenanalysen für den Endpunkt Zeit bis zur Behinderungsprogression in TEMSO

Studie	Patientenzahl	Hazard Ratio ^a [95 % KI] 14 mg Teriflunomid gegenüber Placebo	Interaktionsterm ^b
TEMSO	Geschlecht		
	Männer	T14 (n=104) Pb (n=88)	0,585 [0,326; 1,051] p = 0,5067
	Frauen	T14 (n=254) Pb (n=275)	0,745 [0,502; 1,105]
	Alter		
	<38 Jahre	T14 (n=173) Pb (n=156)	0,579 [0,352; 0,953] p = 0,3292
	>38 Jahre	T14 (n=185) Pb (n=207)	0,811 [0,525; 1,252]
	Ethnische Zugehörigkeit		

White	T14 (n=346)	0,706 [0,507; 0,983]	NA
	Pb (n=356)		
Black	T14 (n=1)	NA	
	Pb (n=3)		
Asian/Oriental	T14 (n=8)	NA	
	Pb (n=1)		
Multiracial	T14 (n=2)	NA	
	Pb (n=1)		
Other	T14 (n=0)	NA	
	Pb (n=1)		
Region			
Osteuropa	T14 (n=108)	0,502 [0,258; 0,978]	p = 0,3138
	Pb (n=114)		
Westeuropa	T14 (n=170)	0,684 [0,442; 1,060]	
	Pb (n=167)		
Amerika	T14 (n=80)	1,114 [0,523; 2,371]	
	Pb (n=82)		
Baseline EDSS-Wert			
≤3,5	T14 (n=276)	0,819 [0,569; 1,178]	p = 0,0670
	Pb (n=281)		
>3,5	T14 (n=82)	0,351 [0,162; 0,765]	
	Pb (n=82)		
Anzahl Schübe in den vorangegangenen 2 Jahren			
≤1	T14 (n=71)	0,697 [0,311; 1,561]	p = 0,7955
	Pb (n=71)		
2	T14 (n=191)	0,783 [0,508; 1,208]	
	Pb (n=186)		
3	T14 (n=70)	0,492 [0,211; 1,147]	
	Pb (n=76)		
≥4	T14 (n=26)	0,628 [0,222; 1,776]	
	Pb (n=30)		
MS-Subtyp			
Secondary progressive and progressive relapsing MS	T14 (n = 26)	0,489 [0,128; 1,868]	p = 0,6002
	Pb (n = 34)		

Relapsing remitting MS	T14 (n = 332) Pb (n = 329)	0,721 [0,514; 1,011]	
Vorbehandlung			
Ja	T14 (n = 102) Pb (n = 90)	0,497 [0,270; 0,916]	p = 0,1490
Nein	T14 (n = 256) Pb (n = 273)	0,812 [0,551; 1,196]	
Baseline-Wert: Anzahl Gadolinium-anreichernder Läsionen			
0	T14 (n = 230) Pb (n = 222)	0,762 [0,504; 1,152]	p = 0,4987
≥1	T14 (n = 125) Pb (n = 137)	0,594 [0,343; 1,029]	
Baseline-Wert: Burden of Disease			
<13 ml	T14 (n = 183) Pb (n = 183)	0,676 [0,423; 1,082]	p = 0,9423
≥13 ml	T14 (n = 172) Pb (n = 175)	0,669 [0,420; 1,067]	

^aAbgeleitet aus einem Cox-proportional-Hazard-Modell, das in der jeweiligen Subgruppe separat durchgeführt wurde, mit Behandlung, EDSS-Baseline-Strata und Region als Kovariate.

^bAbgeleitet aus einem Cox-proportional-Hazard-Modell mit Behandlung, EDSS-Baseline-Strata, Region, Subgruppe und Behandlung-Subgruppe-Interaktion als Kovariate.

Auch bezüglich der Zeit bis zur Behinderungsprogression waren die Ergebnisse über die Subgruppen konsistent. Mit Interaktionstermen von $p = 0,0670$ und $p = 0,1490$ (zwischen 0,05 und 0,2) können der Baseline-EDSS-Wert und die Vorbehandlung als mögliche Effektmodifikatoren identifiziert werden. Die Studie war nicht gepowert, um Subgruppen separat zu untersuchen, so dass die Daten nur auf einer deskriptiven Basis betrachtet werden können. Obwohl es Hinweise auf das Vorliegen von Effektmodifikatoren gibt, zeigt 14 mg Teriflunomid in allen Behandlungsgruppen mit einer einzigen Ausnahme (Region Amerika) einen überlegenen Effekt. Eine plausible Erklärung für diese Ausnahme gibt es nicht, das Ergebnis dieser Subgruppe wurde von der FDA nicht als Grund für eine Einschränkung der Zulassung in den USA gewertet [74].

4.3.1.3.12.3 Ergebnisse der Subgruppenanalysen der Studie TOWER

Im Studienprotokoll der Studie TOWER waren für den primären Endpunkt, die jährliche Schubrate, und den wesentlichen sekundären Endpunkt, die Zeit bis zur Behinderungsprogression, folgende Subgruppenanalysen a priori festgelegt:

- Demografische Faktoren (Alter, Geschlecht, Region, Ethnische Zugehörigkeit)

- Klinische Krankheitsaktivität (Schubaktivität, MS-Form, Baseline EDSS-Wert, Vorbehandlung)

Für die Subgruppenanalysen des primären Endpunktes „jährliche Schubrate“ wurde ein Poisson-Modell angewendet, für den wesentlichen sekundären Endpunkt „Zeit bis zur Behinderungsprogression“ ein Cox Proportional-Hazards-Modell, in das jeweils die Subgruppenzugehörigkeit und der Behandlung-Subgruppen-Interaktionsterm zusätzlich aufgenommen wurden. Jeder Subgruppenfaktor wurde separat analysiert. Das Modell beinhaltet somit die Faktoren Behandlung, Baseline EDSS-Wert ($\leq 3,5$ / $> 3,5$), Region, Subgruppe und Behandlung-Subgruppe-Interaktion und im Falle des Poisson-Modells zusätzlich die log-transformierte standardisierte Studiendauer als Offset-Variable.

Jährliche Schubrate

Tabelle 4-127: Subgruppenanalysen für den Endpunkt jährliche Schubrate in TOWER

Studie	Patientenzahl	Relative Risk ^a [95 % KI] 14 mg Teriflunomid gegenüber Placebo	Interaktionsterm ^b
TOWER	Geschlecht		
	Männer	<u>T14 (n = 113)</u> Pb (n = 115)	0,706 [0,486; 1,027] p = 0,4257
	Frauen	<u>T14 (n = 257)</u> Pb (n = 273)	0,595 [0,455; 0,777]
	Alter		
	<38 Jahre	<u>T14 (n = 173)</u> Pb (n = 177)	0,600 [0,447; 0,805] p = 0,5916
	>38 Jahre	<u>T14 (n = 197)</u> Pb (n = 211)	0,679 [0,493; 0,936]
	Ethnische Zugehörigkeit		
	Kaukasier/Weiße	<u>T14 (n = 312)</u> Pb (n = 317)	0,677 [0,536; 0,855] NA
	Schwarze	<u>T14 (n = 7)</u> Pb (n = 7)	NA
	Asiaten/Orientalen	<u>T14 (n = 48)</u> Pb (n = 60)	0,332 [0,168; 0,657]
	Andere	<u>T14 (n = 3)</u> Pb (n = 4)	NA
	Region		
	Osteuropa	<u>T14 (n = 115)</u> Pb (n = 116)	0,725 [0,508; 1,033] p = 0,1703

Westeuropa und Afrika	T14 (n = 120) Pb (n = 121)	0,634 [0,423; 0,950]	
Amerika	T14 (n = 81) Pb (n = 84)	0,731 [0,448; 1,193]	
Asien und Australien	T14 (n = 54) Pb (n = 67)	0,336 [0,182; 0,623]	
Baseline EDSS-Wert			
≤3,5	T14 (n = 275) Pb (n = 293)	0,622 [0,475; 0,813]	p = 0,7401
>3,5	T14 (n = 95) Pb (n = 95)	0,657 [0,449; 0,960]	
Anzahl Schübe in dem vorangegangenen Jahr			
≤1	T14 (n = 244) Pb (n = 259)	0,669 [0,512; 0,876]	p = 0,6703
2	T14 (n = 99) Pb (n = 105)	0,587 [0,388; 0,888]	
≥3	T14 (n = 26) Pb (n = 23)	0,474 [0,239; 0,939]	
Anzahl Schübe in den vorangegangenen 2 Jahren			
≤1	T14 (n = 121) Pb (n = 128)	0,666 [0,439; 1,011]	p = 0,8895
2	T14 (n = 154) Pb (n = 162)	0,604 [0,429; 0,851]	
3	T14 (n = 50) Pb (n = 56)	0,570 [0,384; 0,845]	
≥4	T14 (n = 43) Pb (n = 42)	0,760 [0,400; 1,444]	
MS-Subtyp			
Secondary progressive and progressive relapsing MS	T14 (n = 4) Pb (n = 10)	NA	p = 0,3139
Relapsing remitting MS	T14 (n = 364) Pb (n = 378)	0,628 [0,506; 0,779]	
Vorbehandlung			
Ja	T14 (n = 126)	0,781 [0,558; 1,094]	p = 0,1328

		Pb (n = 135)		
	Nein	T14 (n = 244)	0,561 [0,424; 0,742]	
		Pb (n = 253)		
		Gebiet		
	USA	T14 (n = 68)	0,692 [0,401; 1,192]	p = 0,7078
		Pb (n = 72)		
	Nicht USA	T14 (n = 302)	0,625 [0,492; 0,794]	
		Pb (n = 316)		

^aErhalten aus einem Poisson-Modell, das in der jeweiligen Subgruppe separat durchgeführt wurde, mit der Anzahl bestätigter Schübe, die zwischen Randomisierung und letzter Dosierung aufgetreten sind, als Responsvariable, Behandlung, EDSS-Baseline und Region als Kovariaten und log-transformierten standardisierten Studiendauern als Offset-Variable.

^bErhalten aus einem Poisson-Modell mit der Anzahl bestätigter Schübe, die zwischen Randomisierung und letzter Dosierung aufgetreten sind, als Responsvariable, Behandlung, EDSS-Baseline, Region, Subgruppe und Behandlung-Subgruppe-Interaktion als Kovariaten und log-transformierten standardisierten Studiendauern als Offset-Variable.

Insgesamt ergeben sich aus den Subgruppenergebnissen des Endpunktes „jährliche Schubrate“ kein schlüssiger Hinweis für eine bedeutsame Interaktion. Mit Interaktionstermen von $p = 0,1703$ und $p = 0,1328$ könnten jedoch die Region und die Vorbehandlung möglicherweise Effektmodifikatoren sein.

Zeit bis zur Behinderungsprogression

Tabelle 4-128: Subgruppenanalysen für den Endpunkt Zeit bis zur Behinderungsprogression (12 week sustained disability progression) in TOWER

Studie	Patientenzahl	Hazard ratio [95 % KI] 14 mg Teriflunomid gegenüber Placebo		Interaktionsterm
TOWER	Geschlecht			
	Männer	T14 (n = 113)	0,337 [0,166; 0,685]	p = 0,0169
		Pb (n = 115)		
	Frauen	T14 (n = 257)	0,944 [0,593; 1,501]	
		Pb (n = 273)		
	Alter			
<38 Jahre	T14 (n = 173)	0,564 [0,325; 0,978]	p = 0,3859	
	Pb (n = 177)			
>38 Jahre	T14 (n = 197)	0,818 [0,478; 1,400]		
	Pb (n = 211)			
Ethnische Zugehörigkeit				

Kaukasier/Weiße	T14 (n = 312) Pb (n = 317)	0,692 [0,460; 1,040]	p = 0,9750
Schwarze	T14 (n = 7) Pb (n = 7)	NA	
Asiaten/Orientalen	T14 (n = 48) Pb (n = 60)	0,508 [0,135; 1,916]	
Andere	T14 (n = 3) Pb (n = 4)	NA	
Region			
Osteuropa	T14 (n = 115) Pb (n = 116)	1,054 [0,439; 2,534]	p = 0,7433
Westeuropa und Afrika	T14 (n = 120) Pb (n = 121)	0,656 [0,366; 1,175]	
Amerika	T14 (n = 81) Pb (n = 84)	0,638 [0,286; 1,422]	
Asien und Australien	T14 (n = 54) Pb (n = 67)	0,508 [0,179; 1,444]	
Baseline EDSS-Wert			
≤3,5	T14 (n = 275) Pb (n = 293)	0,704 [0,459; 1,081]	p = 0,7339
>3,5	T14 (n = 95) Pb (n = 95)	0,601 [0,256; 1,412]	
Anzahl Schübe in dem vorangegangenen Jahr			
≤1	T14 (n = 244) Pb (n = 259)	0,997 [0,614; 1,618]	p = 0,0179
2	T14 (n = 99) Pb (n = 105)	0,409 [0,186; 0,901]	
≥3	T14 (n = 26) Pb (n = 23)	0,127 [0,023; 0,697]	
Anzahl Schübe in den vorangegangenen 2 Jahren			
≤1	T14 (n = 121) Pb (n = 128)	1,587 [0,771; 3,266]	p = 0,0454
2	T14 (n = 154) Pb (n = 162)	0,520 [0,279; 0,969]	
3	T14 (n = 50) Pb (n = 56)	0,507 [0,189; 1,359]	

≥4	T14 (n = 43) Pb (n = 42)	0,211 [0,064; 0,698]	
MS-Subtyp			
Secondary progressive and progressive relapsing MS	T14 (n = 4) Pb (n = 10)	NA	p = 0,4888
Relapsing remitting MS	T14 (n = 364) Pb (n = 378)	0,663 [0,449; 0,979]	
Vorbehandlung			
Ja	T14 (n = 126) Pb (n = 135)	0,477 [0,254; 0,894]	p = 0,1703
Nein	T14 (n = 244) Pb (n = 253)	0,833 [0,511; 1,356]	
Gebiet			
USA	T14 (n = 68) Pb (n = 72)	0,686 [0,293; 1,607]	p = 0,9321
Nicht USA	T14 (n = 302) Pb (n = 316)	0,688 [0,448; 1,057]	

^aAbgeleitet aus einem Cox-proportional-Hazard-Modell, das in der jeweiligen Subgruppe separat durchgeführt wurde, mit Behandlung, EDSS-Baseline-Strata und Region als Kovariate.

^bAbgeleitet aus einem Cox-proportional-Hazard-Modell mit Behandlung, EDSS-Baseline-Strata, Region, Subgruppe und Behandlung-Subgruppe-Interaktion als Kovariate.

Entsprechend der Subgruppenanalysen könnten Geschlecht ($p = 0,0169$), Anzahl der Schübe im vergangenen Jahr ($p = 0,0179$), Anzahl der Schübe in den vergangenen zwei Jahren ($p = 0,0454$) möglicherweise Effektmodifikatoren sein. Besonders das Ergebnis bezüglich der Subgruppen, die die Schubaktivität betreffen, ist interessant, weil mit steigender Anzahl an Schüben zu Beginn der Studie ein monotoner Trend für einen stärkeren Effekt von 14 mg Teriflunomid zu beobachten ist. Ein entsprechender Effekt konnte in den Subgruppenanalysen des Endpunktes Behinderungsprogression in der TEMSO-Studie jedoch nicht beobachtet werden. Vor dem Hintergrund des allgemein deskriptiven Charakters der vorliegenden Subgruppenanalysen und der fehlenden Reproduzierbarkeit dieses Ergebnisses in einer zweiten Studie wird der Unterschied in Bezug auf die Schubaktivität nicht in die Nutzenbewertung von 14 mg Teriflunomid einbezogen.

4.3.1.3.12.4 Zusammenfassende Auswertung der Subgruppenanalysen

In den Studienprotokollen zu TENERE, TEMSO und TOWER ist das Prüfen auf eine qualitative Interaktion in solchen Fällen vorgesehen, in denen der Interaktionsterm einen Wert unter $p = 0,05$ angenommen hat.

In der Studie TENERE liegt keine Subgruppenanalyse vor, deren Interaktionsterm auf das Vorliegen relevanter Effektmodifikation schließen lässt. Allerdings sind in die TENERE-Studie nur eine relativ geringe Anzahl an Patienten eingeflossen, weswegen auch die Stichprobengrößen für die Subgruppenanalysen nicht die notwendige Power bereitstellen, um das Aufdecken moderater Unterschiede zu ermöglichen und es konnten darüber hinaus nicht alle a priori geplanten Subgruppenanalysen durchgeführt werden.

Subgruppenanalysen der größer angelegten Studien TEMSO und TOWER haben größere Chancen, das Vorliegen relevanter Effektmodifikatoren aufzudecken. Die Ergebnisse sind für die Bewertung des Zusatznutzens jedoch nur eingeschränkt verwertbar, denn:

- Ein Effektmodifikator im Vergleich Teriflunomid gegen Placebo muss nicht zwangsläufig einen Einfluss auf den Vergleich von Teriflunomid gegen die zweckmäßige Vergleichstherapie Rebif® 44 µg haben.
- Die große Anzahl an a priori geplanten Subgruppenanalysen (multiples Testen) erhöht die „Wahrscheinlichkeit, dass die Ergebnisse in irgendeiner Subgruppe statistische Signifikanz erreichen, obwohl es sich in Wahrheit um ein zufälliges Ergebnis handelt“ [17]
- Auch die Studien TEMSO und TOWER waren nicht gepowert, um signifikante Ergebnisse in Subgruppen zu erhalten.

Es können sich demnach falsch positive oder falsch negative Ergebnisse ergeben. Deswegen werden die Ergebnisse aus TEMSO und TOWER vergleichend nebeneinander gestellt.

Jährliche Schubrate

Tabelle 4-129: Vergleichende Gegenüberstellung der Subgruppenanalysen aus TEMSO und TOWER- Jährliche Schubrate

Jährliche Schubrate				
Subgruppe	TEMSO	TOWER	TEMSO	TOWER
	≥0,05	≥0,05	<0,05	<0,05
Geschlecht	x	x		
Alter	x	x		
Ethnie	NA	NA	NA	NA
Region	x	x		
Baseline EDSS-Wert	x	x		
Anzahl Schübe in den vorangegangenen 2 Jahren	x	x		
Anzahl Schübe im vorangegangenen Jahr		x		
MS-Subtyp	x	x		

Vorbehandlung	x	x		
Baseline-Wert: Anzahl Gd-anreichernder Läsionen	x			
Baseline-Wert: Burden of Disease	x			
Gebiet (USA / nicht USA)		x		

In TEMSO wurden die Region und der Baseline EDSS-Wert als mögliche Effektmodifikatoren auf die jährliche Schubrate identifiziert, in der TOWER-Studie waren es entsprechend die Region und die Vorbehandlung. In allen genannten Fällen handelte es sich um eine quantitative, nicht aber um eine qualitative Interaktion, d. h. der positive Effekt von 14 mg Teriflunomid auf die jährliche Schubrate war in allen Fällen gegeben. Lediglich für die Region gab es in beiden Studien den Hinweis auf eine Interaktion. Die entsprechenden Ergebnisse lassen sich jedoch nicht vergleichen, weil die betrachteten Regionen unterschiedlich sind.

Insgesamt geben die Daten keinen eindeutigen Hinweis auf eine Effektmodifikation bezüglich der jährlichen Schubrate.

Zeit bis zur Behinderungsprogression

Zeit bis zur Behinderungsprogression				
Subgruppe	TEMSO	TOWER	TEMSO	TOWER
	$\geq 0,05$	$\geq 0,05$	$< 0,05$	$< 0,05$
Geschlecht	x			x
Alter	x	x		
Ethnie	NA	x	NA	
Region	x	x		
Baseline EDSS-Wert	x	x		
Anzahl Schübe in den vorangegangenen 2 Jahren	x			x
Anzahl Schübe im vorangegangenen Jahr				x
MS-Subtyp	x	x		
Vorbehandlung	x	x		
Baseline-Wert: Anzahl Gd-anreichernder Läsionen	x			
Baseline-Wert: Burden of Disease	x			

Gebiet (USA / nicht USA)		x		
--------------------------	--	---	--	--

Für den Endpunkt „Zeit bis zur Behinderungsprogression“ gab es in der TOWER-Studie mit dem Geschlecht und der Anzahl an Schüben Subgruppen mit einem p-Wert des Interaktionstests unter 0,05. Für die entsprechenden Subgruppen gab es in der TEMSO-Studie jedoch keinen Hinweis auf das Vorliegen einer relevanten Effektmodifikation.

Mit einem p-Wert zwischen 0,05 und 0,2 wurde sowohl in TEMSO als auch in TOWER die Vorbehandlung als möglicher Effektmodifikator identifiziert. Auch hier handelt es sich um eine quantitative, nicht aber um eine qualitative Interaktion.

4.3.1.3.13 Zusammenfassung der Ergebnisse aus randomisierten kontrollierten Studien

Der vorliegende Abschnitt soll einen Überblick über die Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen aus randomisierten kontrollierten Studien geben. Die Zusammenfassung soll Aussagen zu allen in Abschnitt 4.3.1.3 präsentierten Endpunkten und Subgruppenanalysen enthalten. Dabei sollen, soweit verfügbar, numerische Ergebnisse aus Metaanalysen einschließlich Konfidenzintervallen dargestellt werden.

Fassen Sie die Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen aus randomisierten kontrollierten Studien zusammen.

4.3.1.3.13.1 Zusammenfassung der Ergebnisse zum medizinischen Nutzen aus randomisierten kontrollierten Studien

Zur Darstellung des Nutzens von 14 mg Teriflunomid konnten zwei placebokontrollierte und eine aktivkontrollierte randomisierte klinische Studien identifiziert werden:

- Placebokontrolliert:
 - TEMSO (**T**eriflunomide **M**ultiple **S**clerosis **O**ral)
 - TOWER (**T**eriflunomide **O**ral in people **W**ith relapsing multi**P**le **S**cle**R**osis)
- Aktivkontrolliert (Komparator Interferon beta-1a, Rebif® 44 µg):
 - TENERE

Die Demonstration der Wirksamkeit von Teriflunomid in der Dosierung von 14 mg/Tag in der Behandlung von schubförmig-remittierender Multipler Sklerose basiert wesentlich auf den robusten und konsistenten Ergebnissen von 2 pivotalen Phase 3-Studien (TEMSO/EFC6049 und TOWER/EFC10531). Mit den placebokontrollierten Studien TEMSO und TOWER lässt sich die gute Wirksamkeit von 14 mg Teriflunomid zeigen und die spezifischen unerwünschten Ereignisse verlässlich abschätzen, worauf eine fundierte Nutzen-Risiko-Abschätzung beruht. In der aktivkontrollierten Studie TENERER/EFC10891 zeigt sich hingegen der beträchtliche Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie. Der Nutzen von Interferon beta-1a, das durchgehend in nationalen [62] und internationalen

[75-77] Leitlinien empfohlen wird und einen sehr hohen Stellenwert in der täglichen Versorgung der MS-Patienten hat, ist unbestritten. Der direkte Vergleich mit diesem Therapiestandard, in dem 14 mg Teriflunomid im Gesamtbild überlegen ist, belegt auch dessen Nutzen für die MS-Patienten in der Zielpopulation.

Der Übersichtlichkeit halber werden im Folgenden die Ergebnisse zum medizinischen Nutzen lediglich für die placebokontrollierten Studien TEMSO und TOWER im Detail zusammengefasst, eine ausführlichere Zusammenfassung der Ergebnisse aus TENERE folgen im anschließenden Abschnitt 4.3.1.3.13.2, der sich mit dem Zusatznutzen befasst.

Tabelle 4-130: 14 mg Teriflunomid versus Placebo (Ergebnisse der TEMSO- und der TOWER-Studie)

Endpunkt	Studie	Effektschätzer KI [] p-Wert	Ergebnis
Mortalität			
Während der Studie verstorbene Patienten	Metaanalyse aus TEMSO und TOWER (gesamte Studienpopulation)	Risikodifferenz 0,0007 [-0,0039; 0,0054] p = 0,7539	keine erhöhte Mortalität
Morbidität			
Jährliche Schubrate	Metaanalyse aus TEMSO und TOWER (RRMS-Patienten)	Rate Ratio 0,644 [0,552; 0,751] p < 0,0001	Sehr gute Effektivität in den klinischen Endpunkten
Schubfreie Patienten	Metaanalyse aus TEMSO und TOWER (RRMS-Patienten)	Relatives Risiko 1,264 [1,154; 1,385] p < 0,0001	
Zeit bis zur Behinderungsprogression	Metaanalyse aus TEMSO und TOWER (RRMS-Patienten)	Hazard Ratio 0,696 [0,539; 0,898] p = 0,0053	
EDSS-Score	Metaanalyse aus TEMSO und TOWER (gesamte Studienpopulation)	Mittelwertsdifferenz 0,062 [-0,070; 0,193] p = 0,3582	
MSFC-Z-Score	TEMSO (gesamte Studienpopulation)	LS-Mean-Differenz 0,048 [-0,056; 0,152] p = 0,3646	
Fatigue (Gesamtwert)	TEMSO (gesamte Studienpopulation)	LS-Mean-Differenz -0,497 [-4,959; 3,965] p = 0,8271	
	TOWER (gesamte Studienpopulation)	LS-Mean-Differenz -2,754 [-7,047; 1,539]	

p = 0,2083			
Therapieversagen	entfällt, da in TEMSO und TOWER die entsprechenden Erhebungen fehlen		
Krankheitslast	TEMSO (gesamte Studienpopulation)	LS-Mean-Differenz -0,089 [-0,137; -0,041] p = 0,0003	
Aktive Läsionen („unique active lesions“)	TESMO (gesamte Studienpopulation)	Relatives Risiko 0,306 [0,228; 0,411] p < 0,0001	
Patienten-zufriedenheit			
TSQM	entfällt, da in TEMSO und TOWER die entsprechenden Erhebungen fehlen		
Lebensqualität			
SF-36 (physische Dimension)	TEMSO (gesamte Studienpopulation)	LS-Mean-Differenz 0,387 [-0,879; 1,653] p = 0,549	Kein nachgewiesener Vorteil in der Lebensqualität
	TOWER (gesamte Studienpopulation)	LS-Mean-Differenz 0,977 [-0,123; 2,077] p = 0,0817	
SF-36 (psychische Dimension)	TEMSO (gesamte Studienpopulation)	LS-Mean-Differenz 0,655 [-0,957; 2,266] p = 0,4255	
	TOWER (gesamte Studienpopulation)	LS-Mean-Differenz 1,479 [-0,122; 3,080] p = 0,0702	
EQ-5D (Lebensqualitäts-index)	TEMSO (gesamte Studienpopulation)	LS-Mean-Differenz 0,004 [-0,028; 0,036] p = 0,8138	
EQ-5D (Visuelle Analogskala)	TEMSO (gesamte Studienpopulation)	LS-Mean-Differenz -2,064 [-5,469; 1,340] p = 0,2343	
Unerwünschte Ereignisse			
Gesamtrate unerwünschter Ereignisse	Metaanalyse aus TEMSO und TOWER (gesamte Studienpopulation)	Relatives Risiko 1,038 [0,998; 1,079] p = 0,0642	Die Rate der Patienten, die die Therapie aufgrund von unerwünschten Ereignissen abgebrochen haben, ist gegenüber Placebo nur leicht erhöht.
Gesamtrate schwerwiegender unerwünschter Ereignisse	Metaanalyse aus TEMSO und TOWER (gesamte Studienpopulation)	Relatives Risiko 1,110 [0,853; 1,444] p = 0,4379	
Gesamtrate unerwünschter	Metaanalyse aus	Relatives Risiko 1,843	

Ereignisse, die zu einem Behandlungsabbruch führten	TEMZO und TOWER (gesamte Studienpopulation)	[1,006; 3,376] p = 0,0477	
Unerwünschte Ereignisse von besonderem Interesse			
Reaktionen an der Injektionsstelle	entfällt, da in TEMZO und TOWER Reaktionen an der Injektionsstelle nicht auftreten konnten (keine Injektion)		Teriflunomid 14 mg hat ein konsistentes und handhabbares Sicherheitsprofil.
Grippeähnliche Symptome	Metaanalyse aus TEMZO und TOWER (gesamte Studienpopulation)	Relatives Risiko 0,908 [0,467; 1,767] p = 0,7763	Eine Bestimmung des Blutbildes und ein Monitoring der Leberwerte sind Teil der qualitätsgesicherten Anwendung von Teriflunomid (s. Modul 3.4)
Verminderte Haardichte (HLT Alopecias)	Metaanalyse aus TEMZO und TOWER (gesamte Studienpopulation)	Relatives Risiko 3,355 [2,258; 4,985] p < 0,0001	
ALT-Wert-Erhöhung	Metaanalyse aus TEMZO und TOWER (gesamte Studienpopulation)	Relatives Risiko 1,875 [1,376; 2,555] p < 0,0001	
Überempfindlichkeiten/ Hautreaktionen	Metaanalyse aus TEMZO und TOWER (gesamte Studienpopulation)	Relatives Risiko 1,213 [0,964; 1,527] p = 0,0997	
Gastrointestinale Beschwerden	Metaanalyse aus TEMZO und TOWER (gesamte Studienpopulation)	Relatives Risiko 1,173 [0,931; 1,477] p = 0,1751	
Durchfall	Metaanalyse aus TEMZO und TOWER (gesamte Studienpopulation)	Relatives Risiko 1,783 [1,320; 2,409] p = 0,0002	
Übelkeit	Metaanalyse aus TEMZO und TOWER (gesamte Studienpopulation)	Relatives Risiko 1,478 [0,914; 2,392] p = 0,1114	
Neutropenie	Metaanalyse aus TEMZO und TOWER (gesamte Studienpopulation)	Relatives Risiko 4,020 [1,949; 8,291] p = 0,0002	
Lymphopenie	Metaanalyse aus TEMZO und TOWER (gesamte Studienpopulation)	Relatives Risiko 1,791 [0,377; 8,502] p = 0,4632	
Leukopenie	Metaanalyse aus TEMZO und TOWER (gesamte Studienpopulation)	Relatives Risiko 4,580 [0,992; 21,16] p = 0,0513	

Infektionen und Infestationen	Metaanalyse aus TEMSO und TOWER (gesamte Studienpopulation)	Relatives Risiko 0,969 [0,792; 1,185] p = 0,7559
Bluthochdruck	Metaanalyse aus TEMSO und TOWER (gesamte Studienpopulation)	Relatives Risiko 2,045 [1,085; 3,854] p = 0,0269
Periphere Neuropathie	Metaanalyse aus TEMSO und TOWER (gesamte Studienpopulation)	Relatives Risiko 2,195 [0,510; 9,447] p = 0,2911
Kopfschmerzen	Metaanalyse aus TEMSO und TOWER (gesamte Studienpopulation)	Relatives Risiko 1,026 [0,822; 1,281] p = 0,8209

Mortalität

In der TEMSO-Studie sind keine teilnehmenden Patienten verstorben. In der TOWER-Studie gab es über die Behandlungsgruppen verteilt 4 Todesfälle. Ein direkter Zusammenhang mit der Studienmedikation ist für keinen dieser Fälle anzunehmen.

Morbidität

Für die Beurteilung krankheitsmodifizierender Wirkstoffe in der MS-Therapie gibt es zwei wesentliche klinische Endpunkte, die Schübe (eine gängige Operationalisierung ist die jährliche Schubrate) und die Behinderungsprogression (üblicherweise operationalisiert als Zeit bis zur Behinderungsprogression). Für beide Endpunkte konnte mit 14 mg Teriflunomid eine therapeutisch relevante Verbesserung erreicht werden.

In den placebokontrollierten Studien TEMSO und TOWER konnte die jährliche Schubrate bei RRMS-Patienten im Schnitt um über 30 % signifikant gesenkt werden (33,9 %; 95 % KI: [17,6 %; 46,9 %] in der TEMSO-Studie und 37,2 %; 95 % KI: [22,1 %; 49,4 %] in der TOWER-Studie). Die Studien TEMSO und TOWER wurden bezüglich dieses Endpunkts in einer Metaanalyse zusammengefasst (s. Kapitel 4.3.2), mit dem Ergebnis einer Reduktion der jährlichen Schubrate um 35,6 % (95 % KI: [24,9 %; 44,8 %]).

In der TEMSO-Studie wurde die Behinderungsprogression bei RRMS-Patienten gegenüber Placebo deutlich verlangsamt (Hazard Ratio von 0,721; 95 % KI: [0,514; 1,011]). In der TOWER-Studie gab es in Bezug auf diesen Endpunkt einen signifikanten Vorteil für 14 mg Teriflunomid (Hazard Ratio von 0,663; 95 % KI: [0,449; 0,979]). Daraus ergibt sich in der metaanalytischen Zusammenfassung insgesamt eine signifikante Verlangsamung der Behinderungsprogression (Hazard Ratio 0,696; 95 % KI: [0,539; 0,898], p = 0,0053).

Damit ist für 14 mg Teriflunomid in Bezug auf die beiden relevanten klinischen Endpunkte in der Beurteilung eines krankheitsmodifizierenden MS-Wirkstoffs für die Zielpopulation (erwachsene Patienten mit RRMS) ein deutlicher klinischer Nutzen nachgewiesen worden.

In einem weiteren Effektivitätsparameter, der Bewertung der durch die Patienten empfundenen Fatigue konnte kein signifikanter Vorteil gezeigt werden.

14 mg Teriflunomid ist in der Lage, die Krankheitslast signifikant zu senken, das Ausmaß der entzündlichen Herde im zentralen Nervensystem geht deutlich zurück. Dies ist zwar kein unmittelbar physisch spürbarer Effekt, für die Patienten und ihr Leben mit der unheilbaren Krankheit aber von ausgesprochen großer Relevanz.

Lebensqualität

In Bezug auf die Lebensqualität konnte mit keinem der beiden etablierten Messinstrumente SF-36 oder EQ-5D eine signifikante Überlegenheit der Therapie gegenüber Placebo gezeigt werden. Eine mögliche Erklärung für dieses Ergebnis ist, dass sich die Lebensqualität von der MS-Populationen dieser beiden Studien nicht so schnell in einem messbaren Ausmaß veränderte, dass dies im zeitlichen Rahmen der klinischen Studien erfasst werden konnte. Die Auswertung der TOWER-Studie zeigte jedoch bereits eine Tendenz für eine verbesserte Lebensqualität durch die Therapie.

Patientenzufriedenheit

Patientenzufriedenheit ist weder in TEMSO noch in TOWER erhoben worden.

Unerwünschte Ereignisse

Teriflunomid zeigte in beiden Studien eine gute Verträglichkeit, weder die Gesamtrate an unerwünschten Nebenwirkungen (relatives Risiko von 1,038; 95 % KI: [0,998; 1,079] (Ergebnis der Metaanalyse von TEMSO und TOWER)), noch die Gesamtrate schwerwiegender unerwünschter Nebenwirkungen (relatives Risiko von 1,110; 95 % KI: [0,853; 1,444] (Ergebnis der Metaanalyse von TEMSO und TOWER)) unterschied sich signifikant von der Placebo-Gruppe. Lediglich die Anzahl an Studienabbrechern aufgrund von Nebenwirkungen war in der TOWER-Studie im Vergleich zu Placebo erhöht (relatives Risiko von 2,508; 95 % KI: [1,593; 3,948]). In der TEMSO-Studie konnte mit einem relativen Risiko von 1,352 und einem zugehörigen 95 %-Konfidenzintervall von [0,856; 2,137] auch in Bezug auf die unerwünschten Ereignisse, die zu einem Therapieabbruch führen, kein signifikanter Unterschied zwischen den Behandlungsarmen (14 mg Teriflunomid und Placebo) festgestellt werden.

Spezifische unerwünschte Ereignisse, die signifikant häufiger auftraten als unter Placebo, waren gastrointestinale Beschwerden, Neutropenie, periphere Neuropathie, eine Verminderte Haardichte (HLT „Alopecias“) und eine Erhöhung der Alanin-Amino-Transferase-Werte (ALT-Werte). Bei den aufgetretenen Ereignissen, die dem High Level Term (HLT) „Alopecias“ zugeordnet werden, handelt es sich nicht um einen kompletten Haarverlust,

sondern um eine reversibel verminderte Haardichte. Entsprechend sind gastrointestinale Beschwerden und diese Alopezien nur in Einzelfällen so unangenehm gewesen, dass sie zu einem Therapieabbruch führten. Neutropenien und Leberwerte sind bei einer qualitätsgesicherten Anwendung entsprechend der Fachinformation zu überwachen (s. Modul 3.4) und stellen somit keine wirklichen Sicherheitsbedenken für die Behandlung mit 14 mg Teriflunomid dar.

Medizinischer Nutzen - Fazit

14 mg Teriflunomid hat einen gut belegten Nutzen bei der Therapie erwachsener Patienten mit schubförmig-remittierender MS. Der Wirkstoff wird gut vertragen und hat ein positives Nutzen-Risiko-Verhältnis, das Grundlage für die arzneimittelrechtliche Zulassung von Teriflunomid im Anwendungsgebiet A ist.

Auch der Zusatznutzen (s. Abschnitt 4.3.1.3.13.2) gegenüber dem Goldstandard in der MS-Therapie, dessen Nutzen im Versorgungsalltag als belegt gilt, zeigt den medizinischen Nutzen von Teriflunomid 14 mg (AUBAGIO®).

4.3.1.3.13.2 Zusammenfassung der Ergebnisse zum medizinischen Zusatznutzen aus randomisierten kontrollierten Studien

Zur Darstellung des Zusatznutzens von 14 mg Teriflunomid konnte eine randomisierte klinische Studie identifiziert werden, die das zu bewertende Arzneimittel mit der zweckmäßigen Vergleichstherapie (Rebif® 44 µg) vergleicht:

- TENERE

Des Weiteren konnten zur Darstellung des Zusatznutzens von 14 mg Teriflunomid zwei placebokontrollierte randomisierte klinische Studien identifiziert werden, über die in einem indirekten Vergleich der Zusatznutzen gezeigt werden kann:

- TEMSO
- TOWER

Die Ergebnisse der Studien TEMSO und TOWER sind im Detail im Abschnitt 4.3.1.3.13.1 zusammengefasst, so dass im Folgenden lediglich die Ergebnisse der TENERE-Studie ergänzend dargestellt werden. Die Ergebnisse des indirekten Vergleichs folgen in Abschnitt 4.3.2.

Tabelle 4-131: 14 mg Teriflunomid versus Rebif® 44 µg (Ergebnisse der TENERE-Studie^{a)})

Endpunkt	Effektschätzer 95 % KI [] p-Wert	Ergebnis
Mortalität		
Gesamtmortalität	entfällt (keine Patienten mit dem Ereignis)	Für keine der beiden Interventionen besteht ein erhöhtes Mortalitätsrisiko
Morbidität		
Schübe (jährliche Schubrate)	Rate Ratio 1,197 [0,623; 2,299] p = 0,5896	Keine verbesserte Wirksamkeit.
Behinderungsprogression (Zeit bis zur Behinderungsprogression)	Hazard Ratio 0,982 [0,399; 2,420] p = 0,9690	
Therapieversagen (Time to Failure)	Hazard Ratio 0,861 [0,564; 1,314] p = 0,5953	
Fatigue	Mittelwertsdifferenz -5,003 [-12,31; 2,307] p = 0,1789	
Lebensqualität		
Lebensqualität (SF-36 oder EQ-5D)	entfällt, da in TENERE die entsprechende Erhebung fehlt	
Patientenzufriedenheit		
Allgemeine Zufriedenheit (TSQM)	Mittelwertsdifferenz 7,84 [1,458; 14,23] p = 0,0162	Signifikante Verbesserung der Patientenzufriedenheit, beträchtlicher Zusatznutzen.
Zufriedenheit in Bezug auf die Wirksamkeit (TSQM)	Mittelwertsdifferenz 3,835 [-3,082; 10,75] p = 0,2759	
Zufriedenheit in Bezug auf die Verträglichkeit (TSQM)	Mittelwertsdifferenz 21,77 [16,48; 27,07] p < 0,0001	
Zufriedenheit in Bezug auf die Anwenderfreundlichkeit (TSQM)	Mittelwertsdifferenz 27,96 [23,35; 32,56] p < 0,0001	
Unerwünschte Ereignisse		
Gesamtrate unerwünschter Ereignisse	RR 0,966 [0,904; 1,031] p = 0,2946	Signifikant weniger Patienten, die aufgrund von unerwünschten Ereignissen, die Studienmedikation abbrechen, geringer Zusatznutzen.
Gesamtrate schwerwiegender	RR 0,787	

Endpunkt	Effektschätzer 95 % KI [] p-Wert	Ergebnis
unerwünschter Ereignisse	[0,274; 2,264] p = 0,6568	
Gesamtrate unerwünschter Ereignisse, die zu einem Behandlungsabbruch führten	RR 0,501 [0,262; 0,959] p = 0,0369	
Unerwünschte Ereignisse von besonderem Interesse		
Reaktionen an der Injektionsstelle	RR nicht berechenbar, da das Ereignis unter 14 mg Teriflunomid nicht auftritt	Die beiden Wirkstoffe unterscheiden sich in dem Risiko für bestimmte unerwünschte Ereignisse.
Grippeähnliche Symptome	RR 0,051 [0,017; 0,158] p < 0,0001	Signifikant häufiger unter Teriflunomid-Behandlung:
Haarverdünnung (HLT Alopecias)	RR 20,2 [2,773; 147,1] p = 0,003	Verminderte Haardichte (HLT Alopecias) und Durchfall
Mit der Leberfunktion assoziierte unerwünschte Ereignisse	RR 0,25 [0,13; 0,45] p < 0,0001	Signifikant häufiger unter Interferon beta-1a-Behandlung:
Erhöhung der ALT-Werte	RR 0,326 [0,173; 0,613] p = 0,0005	Reaktionen an der Injektionsstelle, Grippeähnliche Symptome, mit der Leberfunktion assoziierte Ereignisse (inklusive ALT-Erhöhung) und Kopfschmerzen
Durchfall	RR 2,640 [1,237; 5,632] p = 0,0121	Die Sicherheitsprofile der beiden Wirkstoffe müssen gegeneinander abgewogen werden. Insgesamt zeigt sich eine bessere Verträglichkeit von Teriflunomid und damit ein beträchtlicher Zusatznutzen in diesem Bereich.
Übelkeit und Erbrechen	RR 2,295 [0,743; 7,089] p = 0,1487	
Neutropenie	RR 0,612 [0,178; 2,107] p = 0,4364	
Infektionen und Infestationen	RR 1,055 [0,795; 1,400] p = 0,7108	
Bluthochdruck	RR 1,148 [0,317; 4,156] p = 0,8338	
Periphere Neuropathie	RR 4,591 [0,546; 38,63] p = 0,1608	
Kopfschmerzen	RR 0,570	

Endpunkt	Effektschätzer 95 % KI [] p-Wert	Ergebnis
	[0,338; 0,961] p = 0,0349	

^a99,1 % der eingeschlossenen Patienten sind RRMS-Patienten

Mortalität

In der TENERE-Studie sind keine Todesfälle aufgetreten

Morbidität

Um den Vorteil einer Therapie im klinischen Alltag im Rahmen einer Studie möglichst valide abzubilden, wurde in der TENERE-Studie als primärer Endpunkt ein kombinierter Endpunkt, die Zeit bis zum Therapieversagen („Time to Failure“), gewählt. In diesem Endpunkt konnte der Vorteil von 14 mg Teriflunomid gegenüber Rebif[®] 44 µg nicht nachgewiesen werden (Hazard Ratio von 0,861; 95 % KI: [0,564; 1,314]). Die Berechnung der Fallzahlen basierte auf der Annahme, dass 14 mg Teriflunomid in beiden Teilendpunkten (Therapieversagen aufgrund eines Schubes und Therapieabbruch aus jedwedem anderen Grund) einen Vorteil gegenüber Rebif 44 µg hat. Die Ergebnisse sprechen jedoch für eine vergleichbare Effektivität und einen Vorteil lediglich in Bezug auf die Verträglichkeit und Anwenderfreundlichkeit, so dass der postulierte Unterschied zu groß und die resultierende Fallzahl für ein signifikantes Ergebnis zu gering war.

In der TENERE-Studie ist der wesentliche klinische Endpunkt „Jährliche Schubrate“ als sekundärer Endpunkt definiert. In Bezug auf diesen Endpunkt konnte keine Überlegenheit von 14 mg Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg gezeigt werden (relatives Risiko von 1,197; 95 % KI: [0,623; 2,299]).

Der zweite wesentliche klinische Endpunkt für krankheitsmodifizierende Wirkstoffe in der MS-Therapie ist die Zeit bis zur Behinderungsprogression. Dieser Endpunkt wurde in der TENERE-Studie posthoc ausgewertet. Die Ergebnisse sind kritisch zu hinterfragen, weil die Studiendauer für eine valide Erfassung dieses Endpunkts nicht angelegt war und das Verzerrungspotential für diese posthoc-Analyse dementsprechend als hoch einzustufen ist. Es konnte kein Unterschied zwischen den Interventionen gezeigt werden (Hazard Ratio von 0,982; 95 % KI: [0,399; 2,420]).

Lebensqualität

Die gesundheitsbezogene Lebensqualität ist in der TENERE-Studie nicht bewertet worden.

Patientenzufriedenheit

Mit dem für die Patientenzufriedenheit validierten Messinstrument TSQM (treatment satisfaction questionnaire for medication) konnte eine signifikant größere Zufriedenheit der

Patienten mit 14 mg Teriflunomid als mit Rebif[®] 44 µg gezeigt werden (mittlere Differenz des Scores der allgemeine Zufriedenheit von 7,843; 95 % KI: [1,458; 14,23]).

In den weiteren Bereichen ergibt sich das folgende Bild:

- Kein signifikant besseres Ergebnis im Bereich Effektivität (3,835; 95 % KI: [-3,082; 10,75])
- Signifikant besseres Abschneiden von 14 mg Teriflunomid im Bereich Verträglichkeit (21,77; 95 % KI: [16,48; 27,07])
- Signifikant besseres Abschneiden von 14 mg Teriflunomid im Bereich Anwenderfreundlichkeit (27,96; 95 % KI: [23,35; 32,56])

Das Ergebnis in den einzelnen Bereichen des TSQM ist konsistent mit den Studienergebnissen, die sich für Teriflunomid bei der Auswertung anderer Endpunkte ergeben haben:

- Die Patienten nehmen keine signifikante Verbesserung im Bereich der Wirksamkeit wahr, was mit einer vergleichbaren Effektivität von 14 mg Teriflunomid und Rebif[®] 44 µg bezogen auf die Morbiditätsendpunkte Schübe und Behinderungsprogression korreliert.
- Die Patienten empfinden 14 mg Teriflunomid als besser verträglich als die zweckmäßige Vergleichstherapie. Dies ist in guter Übereinstimmung mit einer geringeren Rate an unerwünschten Ereignissen, die zu einem Therapieabbruch führten, und der bedeutsamen Vermeidung unerwünschter Ereignisse von besonderem Interesse (s. 4.1.8.1.5).
- Die Patienten bewerten 14 mg Teriflunomid als deutlich besser in der Anwenderfreundlichkeit, wodurch der Vorteil der oralen Applikationsform für die Patienten belegt wird.

Dies zeigt, dass die Patienten mit dem zu bewertenden Arzneimittel 14 mg Teriflunomid insgesamt zufriedener waren als mit der zweckmäßigen Vergleichstherapie und Rebif[®] 44 µg

Unerwünschte Ereignisse

Sowohl Teriflunomid als auch Interferon beta-1a sind gut verträgliche Wirkstoffe. Die Gesamtrate unerwünschter Ereignisse (relatives Risiko von 0,966; 95 % KI: [0,904; 1,031]) und die Gesamtrate schwerwiegender unerwünschter Ereignisse (relatives Risiko von 0,787; 95 % KI: [0,274; 2,264]) waren zwischen den Behandlungsgruppen nicht signifikant unterschiedlich.

Neutropenie und eine Erhöhung der ALT-Werte sind für 14 mg Teriflunomid, jedoch auch für Interferon beta-1a, 44 µg, s. c., (Rebif[®] 44 µg) häufige unerwünschte Ereignisse, die kritisch

zu beobachten sind. Im direkten Vergleich (TENERE) zeigte sich bezüglich der Erhöhung der ALT-Werte ein deutlicher, statistisch signifikanter Vorteil für Teriflunomid (relatives Risiko von 0,326 (95 % KI: [0,173; 0,613])); das Ergebnis bezogen auf mit der Leberfunktion assoziierte unerwünschte Ereignisse ist ähnlich (relatives Risiko von 0,25 (95 % KI: [0,13; 0,45])).

Ereignisse, die dem High Level Term (HLT) „Alopecias“ (Verminderte Haardichte) zuzuordnen sind, traten unter 14 mg Teriflunomid häufiger auf (relatives Risiko von 20,2; 95 % KI: [2,773; 147,1]), dagegen leiden die Rebif[®] 44 µg - Patienten an Kopfschmerzen (relatives Risiko von 0,570; 95 % KI: [0,338; 0,961]), grippeähnlichen Symptomen (relatives Risiko 0,051; 95 % KI: [0,017; 0,158]) und Reaktionen an der Injektionsstelle (21,8 % der Patienten, 95% KI: [14,2 %; 31,1%]). Reaktionen an der Injektionsstelle, die unter Rebif[®] 44 µg sehr häufig sind, treten bei der Therapie mit 14 mg Teriflunomid naturgemäß nicht auf, da es sich um eine Filmtablette handelt.

Insgesamt ergibt sich eine bedeutsame Vermeidung unerwünschter Ereignisse durch die Therapie mit 14 mg Teriflunomid. Dies entspricht nach §5 Absatz 7 Satz 2 der AM-NutzenV [9] einem beträchtlicher Zusatznutzen von Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg.

Fazit:

Das Ergebnis der TENERE-Studie ist ein beträchtlicher Zusatznutzen von 14 mg Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg aufgrund von vergleichbarer Effektivität und von bedeutsamen Vorteilen im Nebenwirkungsprofil und der Patientenzufriedenheit.

Aus den Studiendaten kann kein Beleg einer äquivalenten Effektivität gezogen werden, weil die Studie zum Nachweis der Überlegenheit geplant und ein hierarchisch vorgeschalteter Nichtunterlegenheitsnachweis nicht vorgesehen war. Die Aussage der vergleichbaren Effektivität soll daher durch einen indirekten Vergleich gestützt werden, in den die größer angelegten placebokontrollierten Studien TEMSO und TOWER einfließen können (s. Abschnitt 4.3.2).

4.3.2 Weitere Unterlagen

4.3.2.1 Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn indirekte Vergleiche als Nachweis für einen Zusatznutzen herangezogen werden sollen. Das ist dann möglich, wenn keine direkten Vergleichsstudien für das zu bewertende Arzneimittel gegenüber der zweckmäßigen Vergleichstherapie vorliegen oder diese keine ausreichenden Aussagen über den Zusatznutzen zulassen.

Mit der Studie TENERE ist ein direkter Vergleich des zu bewertenden Wirkstoffes Teriflunomid mit der zweckmäßigen Vergleichstherapie Interferon beta-1a, s. c. (Rebif®), vorhanden. Dieser direkte Vergleich hat das folgende Ergebnis: Für Teriflunomid konnte in Bezug auf den primären Endpunkt der Studie („Zeit bis zum Therapieversagen“, s. Abschnitt 4.3.1.3.3) keine Überlegenheit im Vergleich zu Interferon beta-1a, s. c. (Rebif®) gezeigt werden. Teriflunomid 14 mg und Interferon beta-1a, 44 µg, s. c. (Rebif® 44 µg) scheinen im Hinblick auf ihre klinische Wirksamkeit (Effekte auf Schubrate und Behinderungsprogression) vergleichbar zu sein, es konnte aber nicht mit statistisch ausreichender Präzision demonstriert werden. In Bezug auf relevante Nebenwirkungen (grippeähnliche Symptome, Injektionen an der Reaktionsstelle und mit der Leberfunktion assoziierte Ereignisse) und die Abbruchrate aufgrund von Nebenwirkungen zeigt Teriflunomid signifikante Vorteile. Die Ergebnisse der TENERE-Studie und der daraus ableitbare beträchtliche Zusatznutzen für Teriflunomid 14 mg soll durch einen indirekten Vergleich weiter gestützt werden.

4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche

Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung zu Studien für indirekte Vergleiche. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen analog Abschnitt 4.3.1.1 zur Verfügung. Benennen Sie

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Literaturrecherche*
- *Studien aus der Suche in Studienregistern*
- *Resultierender Studienpool aus den einzelnen Suchschritten*

4.3.2.1.1.1 Studien des pharmazeutischen Unternehmers

Benennen Sie in der nachfolgenden Tabelle alle Studien, die an die Zulassungsbehörde übermittelt wurden (Zulassungsstudien), sowie alle Studien, für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war.

Von den randomisierten kontrollierten Studien, die der Zulassungsbehörde im Zulassungsdossier übermittelt wurden, sollen die Studien, deren Studienberichte im Abschnitt 5.3.5 des Zulassungsdossiers enthalten sind, aufgeführt werden. Darüber hinaus sollen alle

randomisierten kontrollierten Studien, für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, aufgeführt werden.

Benennen Sie nur randomisierte, kontrollierte Studien, die ganz oder teilweise innerhalb des in diesem Dokument beschriebenen Anwendungsgebiets durchgeführt wurden. Fügen Sie dabei für jede Studie eine neue Zeile ein.

Folgende Informationen sind in der Tabelle darzulegen: Studienbezeichnung, Angabe „Zulassungsstudie ja/nein“, Studienstatus (abgeschlossen, abgebrochen, laufend), Studiendauer und Therapiearme. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile.

Tabelle 4-132: Liste der Studien des pharmazeutischen Unternehmers – RCT^a mit dem zu bewertenden Arzneimittel

Studienbezeichnung	Zulassungsstudie (ja/nein)	Status (abgeschlossen / abgebrochen / laufend)	Studiendauer ^b	Therapiearme
HMR1726D/2001 (Phase II Studie als Monotherapie)	ja	abgeschlossen	36 Wochen	Teriflunomid 7 mg Teriflunomid 14 mg Placebo
HMR1726D/2002 ^e (LTS6048, Verlängerung der Phase II Monotherapie-Studie)	ja	laufend	528 Wochen (oder bis Teriflunomid in dem Land, in dem der Patient lebt, kommerziell erhältlich ist)	Teriflunomid 7 mg Teriflunomid 14 mg
HMR1726D-2003 (PDY6045, Phase II Studie als Zusatztherapie)	ja	abgeschlossen	24 Wochen	Teriflunomid 7 mg/ Interferon beta Teriflunomid 14 mg/ Interferon beta Placebo/ Interferon beta
HMR1726D/2004 (PDY6046)	ja	abgeschlossen	24 Wochen	Teriflunomid 7 mg/ Glatirameracetat Teriflunomid 14 mg/ Glatirameracetat Placebo/ Glatirameracetat
HMR1726D/2005 (LTS6047, Verlängerung von PDY6045 und PDY6046)	ja	abgeschlossen	24 Wochen	Teriflunomid 7 mg/ Beta-Interferon Teriflunomid 14 mg/ Interferon beta Placebo/ Interferon beta Teriflunomid 7 mg/ Glatirameracetat Teriflunomid 14 mg/ Glatirameracetat Placebo/ Glatirameracetat
TEMPO (HMR1726D/3001, EFC6049)	ja	abgeschlossen	108 Wochen	Teriflunomid 7 mg, Teriflunomid 14 mg, Placebo
TOWER (EFC 10531)	ja	abgeschlossen	48 bis 154 Wochen	Teriflunomid 7 mg, Teriflunomid 14 mg, Placebo
TENERE (EFC10891)	ja	abgeschlossen	48 bis 118 Wochen	Teriflunomid 7 mg, Teriflunomid 14 mg, Interferon beta-1a, s.c. bis zu 44 µg

Studienbezeichnung	Zulassungsstudie (ja/nein)	Status (abgeschlossen / abgebrochen / laufend)	Studiendauer ^b	Therapiearme
Verlängerung der TEMSO-Studie ^e (HMR1726D/3004, LTS6050)	ja	laufend	288 Wochen (oder bis Teriflunomid in dem Land, in dem er Patient lebt, kommerziell erhältlich ist)	Teriflunomid 7 mg Teriflunomid 14 mg
TOPIC (HMR1726D-3005, EFC6260)	nein	laufend ^c	24 bis 108 Wochen	Teriflunomid 7 mg Teriflunomid 14 mg Placebo
TERACLES (EFC6058)	nein	abgeschlossen ^d	48 – 152 Wochen	Teriflunomid 7 mg Teriflunomid 14 mg Placebo
Verlängerung der TOPIC-Studie	nein	laufend	bis Teriflunomid in dem Land, in dem der Patient lebt, kommerziell erhältlich ist	Teriflunomid 7 mg Teriflunomid 14 mg
Verlängerung der TOWER-Studie ^e	nein	laufend	bis Teriflunomid in dem Land, in dem der Patient lebt, kommerziell erhältlich ist	Teriflunomid 14 mg
Verlängerung der TENERE-Studie ^e	nein	laufend	bis Teriflunomid in dem Land, in dem der Patient lebt, kommerziell erhältlich ist	Teriflunomid 14 mg

^aStudien, die keine RCT sind und Phase I-Studien sind nicht mit aufgeführt, dies umfasst u. a. Dosisfindungsstudien an gesunden Probanden oder auch z. B. TERIVA, eine Studie, in der die Immunantwort auf eine Grippeimpfung bei Patienten unter Teriflunomid-Behandlung untersucht wird, bei der es sich aber nicht um eine randomisierte, kontrollierte Studie handelt.

^bbezieht sich in allen Fällen auf die reine Behandlungsdauer, d. h. auf die Studiendauer ohne die Randomisierungs- und Eliminationsphasen

^cDie aktuellen Revisionen der MS-Diagnosekriterien erlauben eine frühere Diagnose der RMS, was das medizinische Umfeld, in dem klinische Studien zu CIS und MS generell durchgeführt werden, verändert hat. Vor diesem Hintergrund hat das Steering Committee die Entscheidung getroffen, die Rekrutierung zu stoppen und so die Anzahl der Patienten, die einer Behandlung mit Placebo ausgesetzt sind, möglichst gering zu halten. [39]

^dFragestellung in der TERACLES-Studie war die Effektivität von Teriflunomid im Vergleich zu Placebo jeweils in Kombination mit Interferon beta bei Patienten mit RMS, die unter laufender Interferon-beta-Therapie weiterhin klinische Aktivität zeigten. Mit Blick auf die aktuellen Entwicklungen im Bereich der MS-Therapie wurde die Kombination einer oralen und einer parenteralen Darreichungsform als ein im klinischen Alltag wenig präferiertes Therapieschema eingestuft. Dies spiegelte sich auch in einer sehr schleppenden Patientenrekrutierung für die Studie wider. Aus diesen Gründen wurde die Studie vorzeitig beendet. [40]

^edie Verlängerungsstudien zu HMR1726/2001, TEMSO, TOWER und TENERE sind Verlängerungen, die unter Wegfall des vorherigen Komparators Placebo oder Interferon beta-1a weiterlaufen. Sie sind in der Liste der Studien des pharmazeutischen Unternehmers aufgeführt, da es sich um Verlängerungen der für die Zulassung und die Nutzenbewertung maßgeblich relevanten Studien handelt..

Geben Sie an, welchen Stand die Information in Tabelle 4-11 hat, d. h. zu welchem Datum der Studienstatus abgebildet wird. Das Datum des Studienstatus sollte nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Die Informationen in Tabelle 4-132 bilden den Studienstatus zum 29.07.2013 ab.

Geben Sie in der nachfolgenden Tabelle an, welche der in Tabelle 4-11 genannten Studien nicht für die Nutzenbewertung herangezogen wurden. Begründen Sie dabei jeweils die Nichtberücksichtigung. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-133: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	Begründung für die Nichtberücksichtigung der Studie
HMR1726D/2001 (Phase II Studie als Monotherapie)	Studiendauer zu kurz (36 Wochen)
HMR1726D/2002 (LTS6048, Verlängerung der Phase II Monotherapie-Studie)	kein passender Komparator (in der Verlängerung Wegfall des Placebo-Arms)
HMR1726D-2003 (PDY6045, Phase II Studie als Zusatztherapie)	Studie zur Kombinationstherapie (andere Indikation als in der Zulassung)
HMR1726D/2004 (PDY6046)	Studie zur Kombinationstherapie (andere Indikation als in der Zulassung)
HMR1726D/2005 (LTS6047, Verlängerung von PDY6045 und PDY6046)	Studie zur Kombinationstherapie (andere Indikation als in der Zulassung)
Verlängerung der TEMSO-Studie (HMR1726D/3004, LTS6050)	kein passender Komparator (in der Verlängerung Wegfall des Placebo-Arms)
TOPIC (HMR1726D-3005, EFC6260)	Studie bei CIS-Patienten (andere Indikation als in der Zulassung),
TERACLES (EFC6058)	Studie zur Kombinationstherapie (andere Indikation als in der Zulassung)
Verlängerung der TOPIC-Studie	Studie bei CIS-Patienten (andere Indikation als in der Zulassung), nicht kontrolliert
Verlängerung der TOWER-Studie	kein passender Komparator (in der Verlängerung Wegfall des Placebo-Arms)
Verlängerung der TENERE-Studie	kein passender Komparator (in der Verlängerung Wegfall des Interferon-Arms)
TENERE ^a	Keine placebokontrollierte Studie

^aTENERE ist für die indirekten Vergleich keine direkt relevante Studie, wird aber herangezogen, um das Ergebnis des indirekten Vergleichs auf Konsistenz zu prüfen.

Stellen Sie die abgebrochenen und laufenden Studien, die nicht in die Bewertung eingehen, in Anhang 4-D und Anhang 4-E dar.

4.3.2.1.1.2 Studien aus der bibliografischen Literaturrecherche

Sofern eine bibliografische Literaturrecherche durchgeführt wurde, beschreiben Sie nachfolgend das Ergebnis dieser Recherche. Illustrieren Sie den Selektionsprozess und das Ergebnis der Selektion mit einem Flussdiagramm. Geben Sie dabei an, wie viele Treffer sich insgesamt (d. h. über alle durchsuchten Datenbanken) aus der bibliografischen Literaturrecherche ergeben haben, wie viele Treffer sich nach Entfernung von Dubletten ergeben haben, wie viele Treffer nach Sichtung von Titel und, sofern vorhanden, Abstract als nicht relevant angesehen wurden, wie viele Treffer im Volltext gesichtet wurden, wie viele der im Volltext gesichteten Treffer nicht relevant waren (mit Angabe der Ausschlussgründe) und wie viele relevante Treffer verblieben. Geben Sie zu den relevanten Treffern an, wie vielen Einzelstudien diese zuzuordnen sind. Listen Sie die im Volltext gesichteten und ausgeschlossenen Studien unter Nennung des Ausschlussgrunds in Anhang 4-C.

Falls die Recherche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel gemeinsam mit der Recherche zu einer anderen Fragestellung (z. B. zu indirekten Vergleichen) durchgeführt wurde, unterteilen Sie die Angaben zu relevanten Treffern und Studien entsprechend diesen Fragestellungen.

[Anmerkung: „Relevanz“ bezieht sich in diesem Zusammenhang auf die im Abschnitt 4.2.2 genannten Kriterien für den Einschluss von Studien in die Nutzenbewertung.]

Orientieren Sie sich bei der Erstellung des Flussdiagramms an dem nachfolgenden Beispiel.

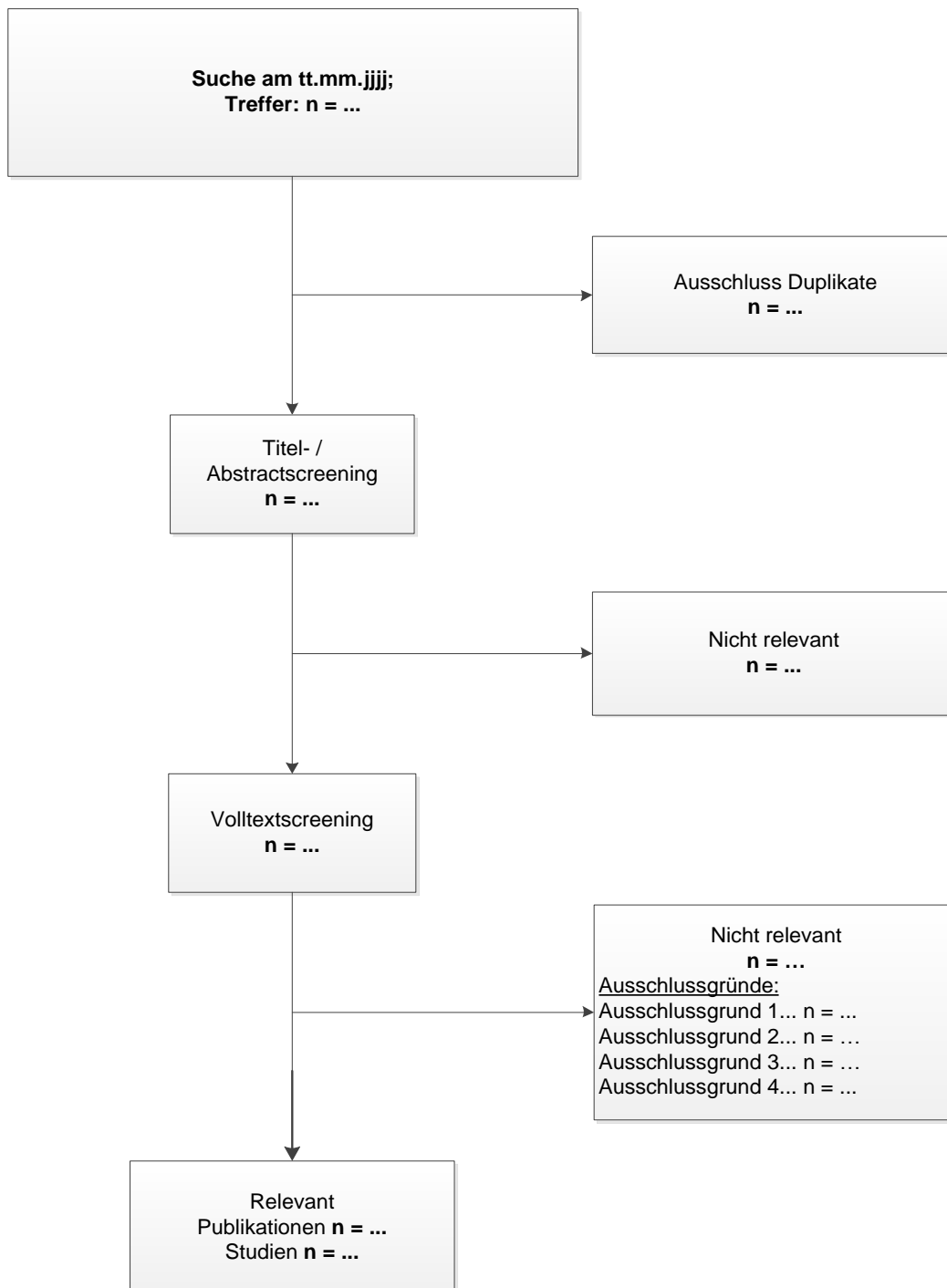


Abbildung 30: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten, kontrollierten Studien mit dem zu bewertenden Arzneimittel

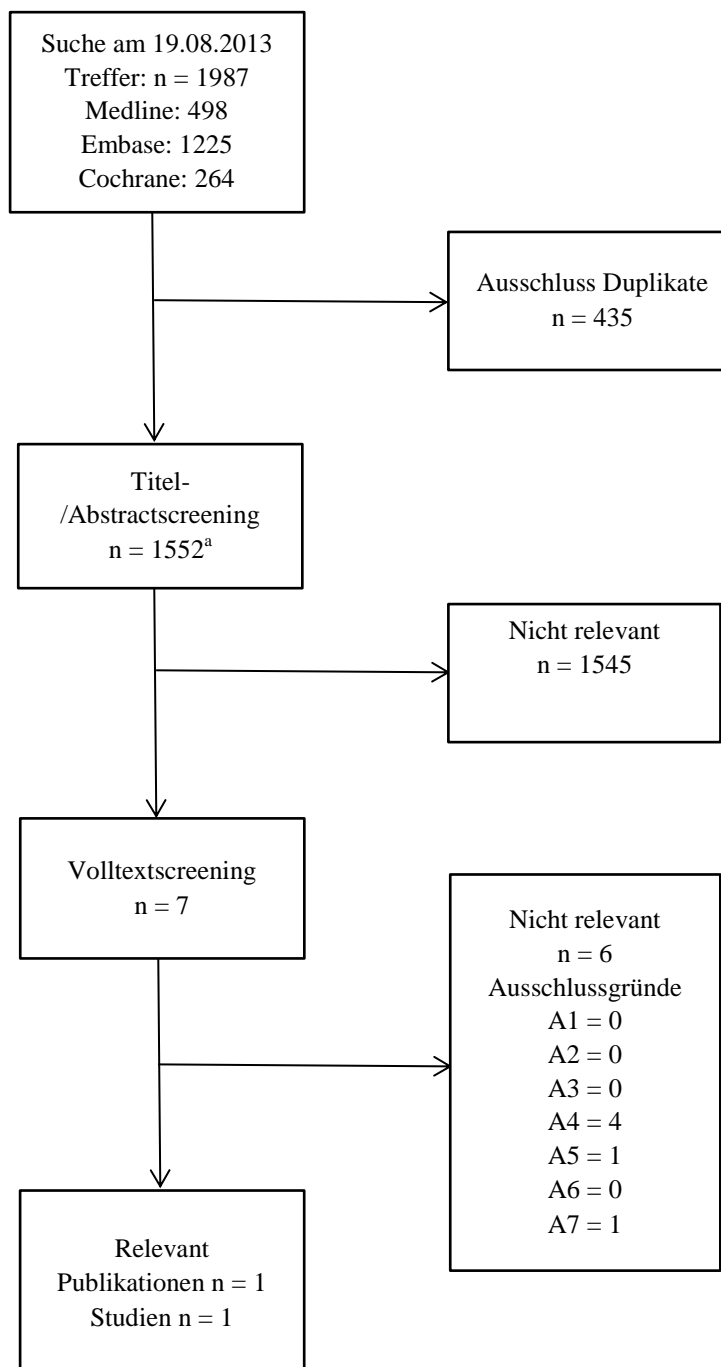
Die systematische Literaturrecherche für das zu bewertende Arzneimittel Teriflunomid ist in Kapitel 4.3.1 dargestellt. Es wurden keine den Studienpool des pharmazeutischen Unternehmers ergänzende Studien gefunden. Aus der systematischen Literaturrecherche nach randomisierten kontrollierten Studien, die Teriflunomid mit Placebo vergleichen, resultierten die in der folgenden Tabelle gelisteten relevanten Treffer.

Tabelle 4-134: Relevante Studien aus der bibliografischen Literaturrecherche – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	Publikation(en)
TEMSO	O'Connor et al. 2011 [35], Miller et al. 2012 [43], O'Connor et al. 2013 [42], Wolinsky et al. 2013 [42]

Das in Abbildung 4-31 dargestellte Flussdiagramm der systematischen Literaturrecherche bezieht sich auf die Suche nach randomisierten kontrollierten Studien, die die zweckmäßige Vergleichstherapie Interferon beta-1a, s.c., dreimal pro Woche, 44 µg (Rebif®) mit Placebo vergleichen. Diese Suche ist **nicht** Bestandteil von Modul 4.3.1 und für den indirekten Vergleich zusätzlich notwendig.

Abbildung 4-31: Flussdiagramm der bibliografischen Literaturrecherche in MEDLINE, EMBASE und den Cochrane-Datenbanken



^aEine geringe Anzahl an Duplikaten wurde erst im Rahmen des Titel-/Abstractscreenings als solche identifiziert, sind in den 1552 Treffern noch enthalten und wurden im Anschluss als „nicht relevant“ aussortiert.

Aus der systematischen Literaturrecherche nach randomisierten kontrollierten Studien, die die zweckmäßige Vergleichstherapie Interferon beta-1a, s. c. (Rebif[®]), mit Placebo vergleichen, resultierten die in der folgenden Tabelle gelisteten relevanten Treffer. Die Ausschlussgründe

A1 bis A7 nach denen die Studien ausgeschlossen wurden sind in Kapitel 4.2.2.2 in Tabelle 4-9 dargestellt.

Tabelle 4-135: Relevante Studien aus der bibliografischen Literaturrecherche – RCT mit der zweckmäßigen Vergleichstherapie

Studienbezeichnung	Publikation(en)
PRISMS	Ebers et al. 1998 [78]

Publikationen der Verlängerungen der PRISMS-Studie wurden im Titel-/Abstractscreening ausgeschlossen. Nach der initialen Studiendauer von 2 Jahren wurden die Patienten der Placebo-Gruppe in die 22µg- und 44µg Interferon-Behandlungsarme randomisiert, so dass die Verlängerung ohne Placebo-Kontrolle die Einschlusskriterien für relevante Studien nicht mehr erfüllt.

Weitere Publikationen (exclusive der in Tabelle 4-135 genannten) der Studie PRISMS, die lediglich einzelne Endpunkte berichten, die nicht als patientenrelevant eingestuft worden sind, oder zusätzliche Auswertungen zu Endpunkten darstellen, in denen die Art der Auswertung der Endpunkte mit den anderen Studien nicht vergleichbar ist, wurden in der Volltextsichtung ausgeschlossen [79-82].

Eine Publikation über hämatologisch Effekte von Rebif[®] [83] enthält Daten zur Häufigkeit von unerwünschten Ereignissen, die mit dem blutbildenden System zusammenhängen (z. B. Leukopenie, Lymphopenie und Neutropenie), aber lediglich als zusammenfassende Auswertung aus mehreren Studien, die bis auf die PRISMS-Studie nicht die Einschlusskriterien erfüllen. Diese Publikation wurde im Volltextscreening ausgeschlossen. In besagter Publikation wird für Rebif[®] ein gegenüber Placebo erhöhtes Risiko für unerwünschte Ereignisse, die mit dem blutbildenden System zusammenhängen, diskutiert. Auch mit Blick auf diese Publikation ist die im Dossier gewählte, von den vorliegenden Daten geleitete Auswertung (Berücksichtigung der Neutropenie im direkten Vergleich, sowie Analysen von Lymphopenie und Leukopenie in indirekten Vergleichen) angemessen.

4.3.2.1.1.3 Studien aus der Suche in Studienregistern

Beschreiben Sie in der nachfolgenden Tabelle alle relevanten Studien, die durch die Suche in Studienregistern identifiziert wurden. Geben Sie dabei an, in welchem Studienregister die Studie identifiziert wurde und welche Dokumente dort zur Studie jeweils hinterlegt sind (z. B. Studienregistereintrag, Bericht über Studienergebnisse etc.). Geben Sie auch an, ob die Studie in der Liste der Studien des pharmazeutischen Unternehmers enthalten ist (siehe Tabelle 4-11) und ob die Studie auch durch die bibliografische Literaturrecherche identifiziert wurde (sofern eine solche durchgeführt wurde). Fügen Sie für jede Studie eine neue Zeile ein.

[Anmerkung: „Relevanz“ bezieht sich in diesem Zusammenhang auf die im Abschnitt 4.2.2 genannten Kriterien für den Einschluss von Studien in die Nutzenbewertung.]

Orientieren Sie sich bei Ihren Angaben an der beispielhaften ersten Tabellenzeile.

Tabelle 4-136: Relevante Studien aus der Suche in Studienregistern – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	Identifikationsorte (Name der Studienregister und Angabe der Zitate ^a)	Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja / nein)	Studie durch bibliografische Literaturrecherche identifiziert (ja / nein / n. d. ^b)
TEM SO	clinicaltrials.gov [NCT00134563], Registereintrag und Ergebnisse [44, 45] ICTRP Search Portal [EUCTR2004-000555-42-NO], Registereintrag [46]; EU ClinicalTrialsRegister, Registereintrag [47]	ja	ja
TOWER	clinicaltrials.gov [NCT00751881], Registereintrag und Ergebnisse [48, 49]; ICTRP Search Portal [EUCTR2007-004452-36-GB], Registereintrag [50]; EU ClinicalTrialsRegister [51], Registereintrag; PharmNet.Bund, Registereintrag [52]	ja	nein
a: Zitat des Studienregistereintrags sowie, falls vorhanden, der im Studienregister aufgelisteten Berichte über Studiendesign und/oder -ergebnisse. b: n. d. = bibliografische Literaturrecherche nicht durchgeführt.			

Tabelle 4-137: Relevante Studien aus der Suche in Studienregistern – RCT mit der zweckmäßigen Vergleichstherapie (Rebif[®] (Interferon beta-1a, s. c.))

Studienbezeichnung	Identifikationsorte (Name der Studienregister und Angabe der Zitate ^a)	Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja / nein)	Studie durch bibliografische Literaturrecherche identifiziert (ja / nein / n. d. ^b)
Keine relevanten Treffer			
a: Zitat des Studienregistereintrags sowie, falls vorhanden, der im Studienregister aufgelisteten Berichte über Studiendesign und/oder -ergebnisse. b: n. d. = bibliografische Literaturrecherche nicht durchgeführt.			

4.3.2.1.1.4 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel und der zweckmäßigen Vergleichstherapie

Benennen Sie in der nachfolgenden Tabelle den aus den verschiedenen Suchschritten (Abschnitte 4.3.1.1.1, 4.3.1.1.2 und 4.3.1.1.3) resultierenden Pool relevanter Studien für das zu bewertende Arzneimittel, auch im direkten Vergleich zur zweckmäßigen Vergleichstherapie. Führen Sie außerdem alle für die Nutzenbewertung berücksichtigten Studien einschließlich der verfügbaren Quellen in Abschnitt 4.6 auf.

Folgende Informationen sind in der Tabelle darzulegen: Studienbezeichnung, Studienkategorie und verfügbare Datenquellen. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Hierbei sollen die Studien durch Zwischenzeilenüberschriften ggf. sinnvoll angeordnet werden, beispielsweise nach Therapieschema (Akut-/Langzeitstudien) und jeweils separat nach Art der Kontrolle (Placebo, zweckmäßige Vergleichstherapie, beides). Sollten Sie eine Strukturierung des Studienpools vornehmen, berücksichtigen Sie diese auch in den weiteren Tabellen in Modul 4.

Tabelle 4-138: Studienpool – RCT mit dem zu bewertenden Arzneimittel und der zweckmäßigen Vergleichstherapie (placebokontrollierte Studien)

Studie	Studienkategorie			verfügbare Datenquellen ^a		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	gesponserte Studie ^b (ja/nein)	Studie Dritter (ja/nein)	Studienbericht (ja/nein [Zitat])	Registereintrag ^c (ja/nein [Zitat])	Publikation (ja/nein [Zitat])
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo						
TEM SO	ja	ja	nein	ja [32]	ja[45, 47]	ja [35]
TOWER	ja	ja	nein	ja [31]	ja[49]	nein
Studien: Interferon beta-1a 44 µg, s.c., dreimal pro Woche, versus Placebo						
PRISMS	nein	nein	ja	nein	nein	Ja [78]
a: Bei Angabe „ja“ sind jeweils die Zitate der Datenquelle(n) mit anzugeben.						
b: Studie, für die der Unternehmer Sponsor war oder an der der Unternehmer anderweitig finanziell beteiligt war.						
c: Zitat der Studienregistereinträge sowie, falls vorhanden, der in den Studienregistern aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.						

4.3.2.1.2 Charakteristika der Studien für indirekte Vergleiche

Charakterisieren Sie nachfolgend die Studien, die für indirekte Vergleiche herangezogen wurden, und bewerten Sie deren Verzerrungspotenzial. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2 und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.

Tabelle 4-139: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel

Studie	Studiendesign <RCT, doppelblind/einfach verblindet/offen, parallel/cross-over etc.>	Population <relevante Charakteristika, z. B. Schweregrad>	Interventionen (Zahl der randomisierten Patienten)	Studiendauer <ggf. Run-in, Behandlung, Nachbeobachtung>	Ort und Zeitraum der Durchführung	Primärer Endpunkt; patientenrelevante sekundäre Endpunkte
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo						
TEMSO	RCT, doppelblind, parallel	Multiple Sklerose Patienten (McDonald's Kriterien) Erwachsene (18-55 Jahre) EDSS \leq 5,5 Mindestens 1 Schub im vorangegangenen Jahr oder 2 Schübe in den 2 vorangegangenen Jahren	Teriflunomid 7 mg (n = 365) Teriflunomid 14 mg (n = 358) Placebo (n = 363)	Behandlungsdauer: 108 Wochen Beobachtungsdauer: 128 Wochen (beinhaltet eine 4-Wochen- Screening-Periode und eine bis zu 16 Wochen dauernde Eliminationsphase (Wash-out) für Patienten, die nicht an der Verlängerung teilnehmen)	126 aktive Zentren in 21 Ländern: Österreich, Kanada, Chile, Tschechien, Dänemark, Estland, Finnland, Frankreich, Deutschland, Italien, Niederlande, Norwegen, Polen, Portugal, Russland, Schweden, Schweiz, Türkei, Ukraine, Großbritannien, USA 09/2004 – 9/2011	Jährliche Schubrate; Zeit bis zur Behinderungsprogressi on (EDSS), MRT- bezogene Endpunkte, Fatigue (FIS), unerwünschte Ereignisse
TOWER	RCT, doppelblind, parallel	Multiple Sklerose Patienten Erwachsene (18-55 Jahre) EDSS \leq 5,5 Schubförmiger Verlauf mit oder ohne Progression (RRMS, SPMS, PRMS) Mindestens 1 Schub im	Teriflunomid 7 mg (n = 408) Teriflunomid 14 mg (n = 372) Placebo (n = 389)	Behandlungsdauer : 48 bis 152 Wochen (je nach Zeitpunkt der Aufnahme in die Studie) Beobachtungsdauer: 56 bis 160 Wochen (beinhaltet eine 4- Wochen-Screening- Periode und eine 4 Wochen dauernde Eliminationsphase für Patienten die nicht ein	189 aktive Zentren in 26 Ländern: Australien, Österreich, Weißrussland, Belgien, Kanada, Chile, China, Tschechien, Estland, Frankreich, Deutschland, Griechenland, Mexiko, Niederlande, Philippinen, Polen, Rumänien, Slowakei, Spanien, Schweden,	Jährliche Schubrate; Zeit bis zur Behinderungsprogressi on (EDSS), Fatigue (FIS), Zeit bis zum ersten Schub, Anteil schubfreier Patienten, Anteil Patienten ohne Behinderungsprogressi on (nach 6 Monaten, 1 Jahr, 2 Jahren), Mittlere Änderung des EDSS, Lebensqualität

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

		vorangegangenen Jahr oder 2 Schübe in den 2 vorangegangenen Jahren		der Verlängerung teilnehmen)	Thailand, Tunesien, Türkei, Ukraine, Großbritannien, USA 08/2008 – 04/2012	(SF-36)
Studien: Interferon beta-1a 44 µg, s.c., dreimal pro Woche, versus Placebo						
PRISMS	RCT, doppelblind, parallel	MS-Patienten Erwachsene ^a EDSS 0-5,0 Verlaufsform RRMS	Interferon beta-1a, s. c., 22 µg (n = 189) Interferon beta-1a, s. c., 44 µg (n = 184) Placebo (n = 187)	2 Jahre	22 Studienzentren in neun Ländern: Australien, Belgien, Kanada, ,Finnland, Deutschland, Niederlande, Schweden, Schweiz, Großbritannien 05/1994 Start der Rekrutierung	Anzahl der Schübe im Verlauf der Studie; Zeit bis zum ersten und zum zweiten Schub, Anteil schubfreier Patienten, Behinderungsprogressi on (Zunahme des EDSS um mindestens 1 Punkt, anhaltend über 3 Monate), benötigte Steroidtherapie und Hospitalisierung, Krankheitsaktivität und Krankheitslast, Depressionen

^agenauere Angaben zum Alter der Patienten finden sich in der Publikation nicht

Tabelle 4-140: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel

Studie	<Gruppe 1>	<Gruppe 2>	<Gruppe 3>	<i>ggf. weitere Spalten mit Behandlungscharakteristika z. B. Vorbehandlung, Behandlung in der Run-in- Phase etc.</i>
<Studie 1>	xxx 250 µg, 1 Inhalation bid + Placebo 2 Inhalationen bid	yyy 200 µg, 2 Inhalationen bid + Placebo 1 Inhalation bid		Vorbehandlung: zzz 1000 µg pro Tag, 4 Wochen vor Studienbeginn Bedarfsmedikation: aaa
Studien: Teriflunomi 14 mg, einmal pro Tag, versus Placebo				
TEMSO	Teriflunomid 7 mg, oral qd	Teriflunomid 14 mg, oral qd	Placebo, oral qd	
TOWER	Teriflunomid 7 mg, oral qd	Teriflunomid 14 mg, oral qd	Placebo, oral qd	
Studien: Interferon beta-1a 44 µg, s.c., dreimal pro Woche, versus Placebo				
PRISMS	Interferon beta-1a, s. c., 22 µg tiw	Interferon beta-1a, s. c., 44 µg tiw	Placebo, s.c, tiw	

qd = einmal täglich, tiw = dreimal wöchentlich

Tabelle 4-141: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel

Studie Gruppe	N (randomisierte Patienten)	Alter (Jahre) Mittelwert \pm Standard-abweichung oder Median ^a	Geschlecht w /m (%)	Vorbehandlung mit MS-Medikation in den vorangegangenen 2 Jahren (ja/nein)	Anzahl Schübe in den vorangegangenen 2 Jahren Mittelwert \pm Standard-abweichung	Zeit seit der ersten MS-Diagnose/ Dauer der Erkrankung (in Jahren) ^b Mittelwert \pm Standard-abweichung	Zeit seit der ersten MS-Diagnose/ Dauer der Erkrankung (in Jahren) ^b Median	Baseline EDSS Mittelwert \pm Standard-abweichung	Verlaufsform RRMS/SP MS/PRMS (%)
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo									
TEMSO									
T7	366	37,4 \pm 9,0	69,7/30,3	102 (27,9 %)/264 (72,1 %)	2,3 \pm 1,2	5,28 \pm 5,36	3,75 (0,1 – 27,6)	2,68 \pm 1,34	91,0/4,6/4,4
T14	359	37,8 \pm 8,2	71,0/29,0	102 (28,4%)/257 (71,6 %)	2,2 \pm 1,0	5,58 \pm 5,48	3,67 (0,1 – 30,1)	2,67 \pm 1,24	92,8/3,3/3,9
Pb	363	38,4 \pm 9,0	75,8/24,2	90 (24,8 %)/273 (75,2 %)	2,2 \pm 1,0	5,13 \pm 5,59	3,50 (0,1 – 31,6)	2,68 \pm 1,34	90,6/6,1/3,3
TOWER									
T7	408	37,4 \pm 9,4	73,5/26,5	123 (30,1%)/285 (69,9%)	2,1 \pm 1,1	5,30 \pm 5,45	3,50 (0,0 – 29,1)	2,71 \pm 1,39	96,3/0,7/2,9
T14	372	38,2 \pm 9,4	69,4/30,6	126 (33,9%)/246 (66,1%)	2,1 \pm 1,2	5,27 \pm 5,90	3,25 (0,0 – 32,7)	2,71 \pm 1,35	98,9/0,5/0,5
Pb	389	37,9 \pm 9,3	70,2/29,8	135 (34,7%)/254 (65,3%)	2,1 \pm 1,1	4,92 \pm 5,66	3,17 (0,0 – 33,8)	2,69 \pm 1,36	97,4/1,0/1,5
Studien: Interferon beta-1a 44 μg, s.c., dreimal pro Woche, versus Placebo									
PRISMS									
Rb 22 μ g	189	34,8	67/33	Vorbehandlung mit immun-modulatorischen Wirkstoffen in den letzten 12 Monaten vor Studienbeginn war Ausschlusskriterium der Studie ^c	3,0 \pm 1,1		5,4 (3,0 – 11,2)	2,5 \pm 1,2	100/0/0
Rb 44 μ g	184	35,6	66/34		3,0 \pm 1,1		6,4 (2,9 – 10,3)	2,5 \pm 1,3	100/0/0
Pb	187	34,6	75/25		3,0 \pm 1,3		4,3 (2,4 – 8,4)	2,4 \pm 1,2	100/0/0

T7 = Teriflunomid 7 mg, T14 = Teriflunomid 14 mg, Pb = Placebo, Rb = Rebif[®] (Beta-Interferon-1a s. c. 22 μ g oder 44 μ g)

^aDie Mediane des Alters der Patienten in der TEMSO- und TOWER-Studie unterscheiden sich nicht maßgeblich von dem Mittelwert der Patienten (für alle Patienten 37,9 Jahre in TEMSO und 38,0 Jahre in TOWER) und wird daher hier nicht gesondert aufgeführt

Medizinischer Nutzen, medizinischer Zusatznutzen, Patientengruppen mit therap. bedeutsamem Zusatznutzen

^bIn den Studien TEMSO und TOWER wird zwischen der Zeit seit Diagnosestellung und Zeit seit den ersten MS-Symptomen differenziert, angegeben ist die Zeit seit der ersten Diagnose, der Median liegt für die Zeit seit den ersten MS-Symptomen für die Patienten der TEMSO-Studie bei 6,83 Jahren und für die TOWER-Studie bei 8,00 Jahren- welche Angabe mit den Angaben der Krankheitsdauer in PRISMS besser übereinstimmt, lässt sich aus den Angaben der Publikation nicht eindeutig ableiten.

^cNur 3 % der Patienten hatten vor diesem Cut-off-Zeitpunkt eine immunsuppressive Therapie erhalten.

Beschreiben Sie die Studien zusammenfassend. Sollte es Unterschiede zwischen den Studien geben, weisen Sie in einem erläuternden Text darauf hin.

Es können drei Studien identifiziert werden, die entweder die Effektivität und Verträglichkeit des zu bewertenden Arzneimittels (Teriflunomid 14 mg, oral, 1 x täglich) oder der zweckmäßigen Vergleichstherapie (Interferon beta-1a, 44 µg, 3 x wöchentlich) bei der Therapie von Patienten mit RRMS untersuchen.

Bei allen drei Studien TEMSO [32, 35, 43], TOWER und PRISMS [78] handelt es sich um doppelblinde, placebokontrollierte Studien.

Studiendauer

Die beiden Teriflunomid-Studien TEMSO und TOWER sind von ähnlicher Größe (ca. 1000 eingeschlossene Patienten), die Rekrutierung der Patienten startete 2004 bzw. 2008. Die beiden Studien sind ähnlich konzipiert, der wesentliche Unterschied der beiden Studien besteht in der Studiendauer. Die TEMSO-Studie lief über 2 Jahre, wogegen die Studiendauer in TOWER pro Patient unterschiedlich war und zwischen 48 und 152 Wochen betrug (Studiendauer = Rekrutierungszeitraum + 48 Wochen Behandlungszeitraum des letzten in die Studie eingeschlossenen Patienten).

Die Interferon-beta-1a-Studie PRISMS [78] untersucht etwas weniger Patienten (560), die Rekrutierung der Patienten erfolgte in den 1990er Jahren. Die Studiendauer in der PRISMS-Studie betrug 2 Jahre.

Alle drei Studien entsprechen von der Studiendauer den Anforderungen, die an Studien zu krankheitsmodifizierenden Wirkstoffen zur Therapie der MS gestellt werden [1, 2]. Die Unterschiede in den Studiendauern begründen keine Annahme einer Heterogenität zwischen den Studien bezüglich relativer Abstandsmaße wie z. B. Rate Ratios, Hazard Ratios, relativen Risiken oder Odds Ratios.

Endpunkte

Der primäre Endpunkt der Teriflunomid-Studien TEMSO und TOWER ist identisch (jährliche Schubrate), in beiden Studien ist die Zeit bis zur Behinderungsprogression der wesentliche sekundäre Endpunkt. In der PRISMS-Studie werden ebenfalls die Schübe als primärer Endpunkt erfasst, die Zeit bis zum ersten Schub und die Behinderungsprogression als sekundäre Endpunkte.

Damit entsprechen alle drei Studien den Empfehlungen der EMA [1, 2], wonach die Behinderungsprogression oder die Schubrate der primäre Endpunkt und der jeweils andere als wesentlicher sekundärer Endpunkt festgelegt werden sollte. Die beiden wesentlichen Effektivitätseindpunkte sind somit in allen eingeschlossenen Studien abgebildet und eine Vergleichbarkeit der Ergebnisse ist prinzipiell gegeben.

Patientenpopulation

- **Alter und Geschlecht**

Die Geschlechterverteilung ist in allen drei Studien ähnlich (zwischen 60 und 75 % der Patienten waren Frauen) und entspricht der Geschlechterverteilung in der Zielpopulation, ca. 70 % der MS-Patienten sind weiblich (s. Modul 3.2.1.3). Auch das mittlere Alter der Patienten zwischen 30 und 45 Jahren ist für Studien im Anwendungsgebiet A als repräsentativ anzusehen (s. Modul 3.2.1.3)..

- **Krankheitsschwere und -verlauf**

Die Patientenpopulationen wiesen in den Studien TEMSO, TOWER und PRISMS bei Studienbeginn eine ähnliche Schwere der Erkrankung auf. Die Patienten hatten bei Studienbeginn im Schnitt etwas mehr als 2 Schübe in den vorangegangenen 2 Jahren und einen mittleren EDSS von ungefähr 2,7, in PRISMS liegt dieser Wert mit 3,0 ein wenig höher. In allen drei Studien leiden 90 – 100 % der eingeschlossenen Patienten an schubförmig-remittierender MS. In den Teriflunomid-Studien TEMSO und TOWER litten jeweils über 90 % der eingeschlossenen Patienten an der Verlaufsform RRMS, in der PRISMS-Studie waren es 100 %. Dies ist in guter Übereinstimmung mit der Zulassung von Teriflunomid, die sich nur auf RRMS-Patienten bezieht.

Ein möglicher Grund für Heterogenität zwischen den Studien ist eine mögliche unterschiedliche Krankheitsschwere, die sich nicht an den oben aufgeführten Baseline-Charakteristika (Schubrate, EDSS-Wert,...) festmachen lässt. Unter dem Überbegriff „Studiendauer“ wurde dargestellt, dass die Patienten für die Interferon-beta-Studien mehr als 10 Jahre früher rekrutiert wurden als die Patienten der Teriflunomid-Studien. Die Diagnose-Kriterien, nach denen ein Patient als MS-Patient eingestuft wird, haben sich in der Zwischenzeit geändert und sind sensitiver geworden [14-16]. Darüber hinaus waren zum Zeitpunkt der Durchführung der Interferon-beta-Studien auf Grund fehlender Therapiemöglichkeiten die Patienten zum Großteil therapienaiv. Mit der breiten Einführung der Interferon beta-Präparate und Glatirameractat hat sich die Situation geändert, so dass ein ein etwas größerer Teil der Patienten der beiden Teriflunomid-Studien TEMSO und TOWER bereits vorbehandelt war (s. Tabelle 4-141) .

Die genannten Faktoren müssen nicht zwangsläufig zu wesentlichen Unterschieden in den Patientenpopulationen führen, doch wird eine gewisse Heterogenität zwischen „älteren“ und „neueren“ MS-Studien diskutiert [84, 85]. Diese Heterogenität wird maßgeblich an einem unterschiedlichen Verhalten der Placebo-Gruppen in den verschiedenen Studien festgemacht, was auch beim Vergleich der Teriflunomid-Studien mit PRISMS zu beobachten ist. So konnten in den jeweiligen Placebo-Gruppen von TEMSO und TOWER jährliche Schubraten von 0,54 und 0,50 beobachtet werden, ein sehr viel geringerer Wert als die in PRISMS entsprechend beobachtete jährliche Schubrate von 1,28. Je nach Art der statistischen Auswertung kann dieser Unterschied das Ergebnis in verschiedenem Ausmaß verzerren und ist in der Analyse kritisch zu bewerten. Relative Abstandsmaße wie Rate Ratios sind für

Verzerrungen jedoch weniger anfällig als absolute Abstandsmaße, weswegen auf die Präsentation der Ergebnisse zu Differenzen von Raten verzichtet wird.

Formulierung der Rebif®-Präparate

Im Laufe der Jahre ist die Formulierung der Rebif®-Präparate verändert worden, die sogenannte „Rebif® New Formulation“ kommt ohne humanes Serumalbumin oder fötales Kälberserum aus, wodurch die Verträglichkeit der Injektion verbessert und das immunogene Potential verringert wird [86]. Dementsprechend ist auch dies ein Punkt, in dem sich speziell Studien mit Rebif® als Intervention abhängig vom Zeitpunkt der Studiendurchführung unterscheiden, so ist die TENERE Studie mit der neuen Formulierung durchgeführt worden, PRISMS jedoch nicht. Für wirksamkeitsbezogene Endpunkte wird der Einfluss der Formulierung als gering eingestuft, das davon ausgehende Verzerrungspotential als niedrig. Für einzelne Endpunkte, die sich auf die Verträglichkeit beziehen (Reaktionen an der Injektionsstelle, grippeähnliche Symptome) ist der Einfluss jedoch nicht zu vernachlässigen und wird an entsprechender Stelle diskutiert.

4.3.2.1.3 Verzerrungspotenzial auf Studienebene

Bewerten Sie das Verzerrungspotenzial der RCT auf Studienebene mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-142: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung		Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial auf Studienebene
			Patient	Behandler			
<Studie 1>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo							
TEMPO	ja	ja	ja	ja	nein	nein	niedrig
TOWER	ja	ja	ja	ja	nein	nein	niedrig
Studien: Interferon beta-1a 44 µg, s.c., dreimal pro Woche, versus Placebo							
PRISMS	ja	ja	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Bei TEMSO, TOWER und PRISMS handelt es sich um randomisierte, doppelblinde, placebo-kontrollierte Studien. Es erfolgte eine adäquate Erzeugung der Randomisierungssequenz und Verdeckung der Gruppenzuteilung. Da die Studien doppelblind durchgeführt worden sind, ist diesbezüglich keine Verzerrung des Studienergebnisses zu erwarten. Es liegt kein Hinweis auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Faktoren vor. Damit ist das Verzerrungspotential auf Studienebene für die Studien TEMSO, TOWER und PRISMS als niedrig einzustufen.

4.3.2.1.4 Ergebnisse aus indirekten Vergleichen

Geben Sie in der folgenden Tabelle einen Überblick über die patientenrelevanten Endpunkte, auf denen Ihre Bewertung des medizinischen Nutzens und Zusatznutzens aus indirekten Vergleichen beruht. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Zeile. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-143: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche

	Mortalität	Morbidität				Adverse events	
		Schübe	Behinderungsprogression				
Studie	Während der Studie verstorbene Patienten	Jährliche Schubrate	Schubfreie Patienten/Patienten mit ≥ 1 Schub	Zeit bis zur Behinderungsprogression	Anteil Patienten mit Behinderungsprogression zu einem bestimmten Zeitpunkt	Mittlere Änderung EDSS (2 Jahre)	AEs, SAEs, Behandlungsabbrüche, relevante spezifische Aes ^a
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo							
TEMSO	Ja	Ja	Ja	Ja	Ja	Ja	Ja
TOWER	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Studien: Interferon beta-1a 44 µg, s.c., dreimal pro Woche, versus Placebo							
PRISMS	Ja	Ja	Ja	Ja	Nein	Ja	Ja
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Interferon beta-1a 44 µg, s.c., dreimal pro Woche							
TENERE	Ja	Ja	Ja	Ja	Ja	Nein	Ja

^aDer Bereich der unerwünschten Ereignisse ist zusammengefasst angegeben, die Einstufung einer Studie mit einem „ja“ für die mögliche Berücksichtigung des Endpunktes in dem indirekten Vergleich bedeutet nicht, dass dies für alle Kategorien (Gesamtrate der unerwünschten Ereignisse, Gesamtrate der schwerwiegenden unerwünschten Ereignisse, Gesamtrate der unerwünschten Ereignisse, die zum Behandlungsabbruch führten, definierter unerwünschte Ereignisse von besonderem Interesse) zutreffen muss, sondern, dass mindestens ein

Unterpunkt im indirekten Vergleich ausgewertet werden kann (s. Auswertung des Endpunktes „unerwünschte Ereignisse“)

Die Tabelle gibt eine Übersicht über die Endpunkte, die im indirekten Vergleich berücksichtigt werden können. Zusätzlich zu den Studien, die in den indirekten Vergleich einfließen, ist die direkt vergleichende Studie TENERE mit aufgenommen, weil die Ergebnisse dieser Studie herangezogen werden, um die Ergebnisse des indirekten Vergleichs auf Konsistenz mit dem direkten Vergleich zu prüfen.

Die Ergebnistabellen zu den einzelnen Endpunkten geben jeweils die Resultate sowohl auf Basis der relativen Risiken (RR) als auch auf Basis der Odds Ratios (OR) an. Die Diskussion in den erläuternden Texten bezieht sich der Einfachheit halber lediglich auf die relativen Risiken, würden die beiden Betrachtungsweisen zu abweichenden Ergebnissen führen, würde dies an der entsprechenden Stelle diskutiert.

In Modul 4.3.1 (Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel) sind mehr Endpunkte ausgewertet als in Tabelle 4-143 angegeben, weil nicht alle Endpunkte im indirekten Vergleich berücksichtigt werden können. Im Einzelnen handelt es sich hierbei um die folgenden Endpunkte:

- Time to Failure
- Fatigue (FIS)
- Lebensqualität (SF-36)
- Lebensqualität (EQ-5D)
- Patientenzufriedenheit (TSQM)

In den für den indirekten Vergleich eingeschlossenen Studien sind die Endpunkte „Time to Failure“ und Patientenzufriedenheit (TSQM) nicht mit aufgenommen, da diese Endpunkte lediglich in der TENERE-Studie berichtet sind.

Ähnliches gilt für die Lebensqualität, die in PRISMS nicht mit SF-36 oder EQ-5D erfasst worden ist.

Die Fatigue ist in der PRISMS-Studie zwar als Nebenwirkung aufgeführt, aber nicht mit dem validierten Messinstrument FIS (Fatigue Impact Scale) erfasst, so dass sich die Ergebnisse der Teriflunomid- und Beta-Interferon 1a- Studien hinsichtlich der Operationalisierung dieses Endpunktes unterscheiden und sich auf dieser Basis nicht vergleichen lassen.

Unerwünschte Ereignisse von besonderem Interesse (AESI = adverse events of special interest) werden später in einer gesonderten Endpunkt-Matrix aufgeschlüsselt (s. Modul 4.3.2.1.4.5). Unerwünschte Ereignisse von besonderem Interesse sind wichtig, um das Schadenspotential einer oder mehreren Interventionen abzuschätzen, so dass, wenn möglich,

auf diese Ereignisse auch auswertend eingegangen werden sollte. Eine solche Analyse ist jedoch mit dem Problem behaftet, dass für Teriflunomid charakteristische Effekte in den Beta-Interferon-Studien z. T. nicht berichtet sind, was zu einer relevanten Verzerrung der Ergebnisse führen kann. Dies ist kritisch zu betrachten und Teil der Diskussion des Endpunktes „unerwünschte Ereignisse“.

Es gibt keine zusätzlich zu berichtenden patientenrelevanten Endpunkte, die in der Beta-Interferon-Studie PRISMS ausgewertet ist, in den Teriflunomid-Studien TEMSO und TOWER jedoch fehlen.

Berücksichtigung der MS-Verlaufsform RRMS

Die Zulassung von Teriflunomid beschränkt sich auf Patienten mit RRMS. In den Studien TEMSO, TOWER und TENERE waren jedoch auch für Patienten mit anderen schubförmigen Formen der MS die Einschlusskriterien erfüllt. Patienten mit schubförmig verlaufender SPMS oder PRMS, d. h. Patienten, mit progressiv schubförmiger MS, konnten in die Studien eingeschlossen werden. De facto ist der Anteil der letztgenannten Patientengruppen in den Studien gering (0 % - 9,4 %), so dass der Einfluss auf das Studienergebnis zwar wahrscheinlich nicht groß, aber auch nicht gänzlich auszuschließen, ist. Der Einfluss der Verlaufsform auf unerwünschte Ereignisse – sofern es sich bei diesen nicht um typische Symptome der MS handelt – wird als niedrig eingestuft. Die Auswertung der gesamten Studienpopulation für Verträglichkeitsendpunkte ist somit zulässig.

Deswegen werden vorliegende Auswertungen der Wirksamkeitsendpunkte, die in den Studienberichten von TEMSO und TOWER auch für die Subgruppe der RRMS-Patienten gemacht werden, für diese Patientengruppe analysiert und diskutiert. In der TENERE-Studie waren lediglich drei Patienten eingeschlossen, die an SPMS oder PRMS litten (alle im 14 mg –Teriflunomid-Behandlungsarm) eine gesonderte Darstellung ohne diese Patienten liegt nicht vor, wird aufgrund des geringen Anteils nicht als ein das Ergebnis verzerrender Faktor gewertet.

Daraus resultiert, dass im indirekten Vergleich je nach Endpunkt unterschiedliche Populationen für die Analyse herangezogen werden (s. Tabelle 4-144); aus der PRISMS-Studie wird immer die gesamte Studienpopulation (100 % RRMS-Patienten) berücksichtigt.

Tabelle 4-144: Ausgewertete Studienpopulationen der Teriflunomid-Studien in den indirekten Vergleichen nach Endpunkt

Endpunkt	Ausgewertete Population
Jährliche Schubrate	TEMSO und TOWER: RRMS-Subpopulation TENERE: gesamte Studienpopulation
Zeit bis zur Behinderungsprogression	TEMSO und TOWER: RRMS-Subpopulation TENERE: gesamte Studienpopulation

Mortalität	TEMZO, TOWER und TENERE: gesamte Studienpopulation
Abbrüche aufgrund von unerwünschten Ereignissen	TEMZO, TOWER und TENERE: gesamte Studienpopulation
Grippeähnliche Symptome	TEMZO, TOWER und TENERE: gesamte Studienpopulation
Reaktionen an der Injektionsstelle	TEMZO, TOWER und TENERE: gesamte Studienpopulation
Kopfschmerzen	TEMZO, TOWER und TENERE: gesamte Studienpopulation
Erhöhung der ALT-Werte	TEMZO, TOWER und TENERE: gesamte Studienpopulation
Lymphopenie	TEMZO, TOWER und TENERE: gesamte Studienpopulation
Leukopenie	TEMZO, TOWER und TENERE: gesamte Studienpopulation

4.3.2.1.4.1 Schübe – indirekte Vergleiche aus RCT

Für die indirekten Vergleiche soll zunächst für jeden Endpunkt eine Übersicht über die verfügbaren Vergleiche gegeben werden. Anschließend soll die Darstellung der Ergebnisse in 3 Schritten erfolgen: 1) Bewertung des Verzerrungspotenzials auf Endpunktebene pro Studie, 2) tabellarische Darstellung der Ergebnisse der einzelnen Studien, 3) Darstellung des indirekten Vergleichs. Für die Punkte 1 und 2 gelten die gleichen Anforderungen wie für die Darstellung der Ergebnisse der direkten Vergleiche in Abschnitt 4.3.1.3.

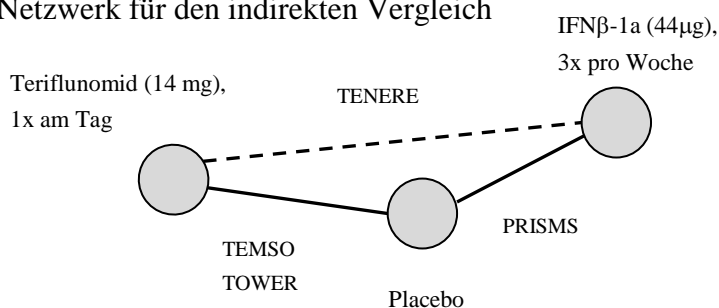
Geben Sie für den im vorliegenden Abschnitt präsentierten Endpunkt einen Überblick über die in den Studien verfügbaren Vergleiche.

Tabelle 4-145: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden

Anzahl Studien	Referenzen Studien	Intervention (14 mg Teriflunomid)	Placebo	Rebif® 44 µg (3 x wöchentlich)
2	TEMZO TOWER	• •	• •	
1	PRISMS		•	•

Stellen Sie zusätzlich die Netzwerkstruktur des indirekten Vergleichs grafisch dar.

Abbildung 4-32: Netzwerk für den indirekten Vergleich



Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-146: Operationalisierung von Schüben

Studie	Operationalisierung
	Schübe
TEMSO	Schübe werden operationalisiert als <ul style="list-style-type: none"> Anzahl bestätigter Schübe pro Patientenjahr („jährliche Schubrate“; ARR = annualized relapse rate) Patienten, die im Studienverlauf schubfrei waren oder Patienten mit ≥ 1 Schub Die Definition eines Schubes ist unten angegeben.
TOWER	Schübe werden operationalisiert als <ul style="list-style-type: none"> Anzahl bestätigter Schübe pro Patientenjahr („jährliche Schubrate“; ARR = annualized relapse rate) Patienten, die im Studienverlauf schubfrei waren oder Patienten mit ≥ 1 Schub Die Definition eines Schubes ist unten angegeben.
	Ein Schub war in TEMSO und TOWER wie folgt definiert: <ul style="list-style-type: none"> Auftreten eines neuen klinischen Zeichens/Symptoms oder die klinische Verschlechterung eines vorhandenen Symptoms (eines, das mindestens 30 Tage lang stabil war), das mindestens 24 h anhielt und nicht von Fieber begleitet war. Die Schübe wurden durch den behandelnden Neurologen, basierend auf einer objektiven Bewertung (EDSS/FS scores) eines unabhängigen Bewerter (der untersuchende Neurologe), bestätigt, wenn eines der folgenden Kriterien erfüllt war: Erhöhung um 1 Punkt in mindestens 2 FS-Funktionen oder eine Erhöhung um 2 Punkte in mindestens einer FS-Funktion (ausgenommen Darm/Blase und Gehirn) Eine Erhöhung um mindestens 0,5 Punkte im EDSS score (es sei denn, der EDSS war vorher 0, dann war eine Erhöhung um 1,0 Punkte gefordert) im Vergleich zu der vorherigen stabilen Einstufung.
PRISMS	Schübe werden operationalisiert als <ul style="list-style-type: none"> Anzahl bestätigter Schübe im Studienverlauf (die „jährliche Schubrate“, auch ARR = annualized relapse rate, wurde durch Division mit der Studiendauer von 2 Jahren abgeleitet) Patienten, die im Studienverlauf schubfrei waren oder Patienten mit ≥ 1 Schub Die Definition eines Schubes ist wie folgt: Schübe waren definiert als das Auftreten eines neuen Symptoms oder die Verschlechterung eines alten Symptoms über die letzten 24 h, das eine MS-Aktivität zugeordnet werden konnte und von einer stabilen Phase oder einer Verbesserung gefolgt wurde (in den folgenden 30 Tagen). Die Schwere eines Schubes wurde mittels der „Scribbs neurological rating scale“ oder der activities of daily living scale“ bewertet.

Die Operationalisierung der Schübe ist in allen drei Studien als ausreichend vergleichbar einzustufen, eine Vergleichbarkeit auf Basis der jährlichen Schubrate ist gegeben.

Im indirekten Vergleich eingeschlossene Operationalisierung des Endpunkts Schübe

Die „jährliche Schubrate“ ist der primäre Endpunkt der placebokontrollierten Teriflunomid-Studien und die gängigste Operationalisierung des Endpunktes Schübe. Nur die Auswertung der jährlichen Schubrate wird im vorliegenden Modul 4.3.2.1.3.1. dargestellt. Weitere

mögliche Darstellungen sind die Angabe einer Differenz von jährlichen Raten, die Auswertung der Zeit bis zum Schub, der Anteil Patienten, die einen Schub erlitten haben oder der Anteil an schubfreien Patienten. Es handelt sich bei diesen Darstellungen in jedem Fall um sekundäre Endpunkte, die darüber hinaus aufgrund von Unvollständigkeit der Daten oder größerer Heterogenität kritischer zu betrachten sind. Der Vollständigkeit halber sind alle durchgeführten Analysen im Anhang (Modul 5) des vorliegenden Dossiers zu finden [87].

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-147: Bewertung des Verzerrungspotenzials für Schübe in RCT für indirekte Vergleiche

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo						
TEMPO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
Studien: Interferon beta-1a 44 µg, s.c., dreimal pro Woche, versus Placebo						
PRISMS	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Für die Bewertung des Verzerrungspotentials hinsichtlich des Endpunktes Schübe in den Studien TEMPO und TOWER sei an dieser Stelle auf Modul 4.3.1 verwiesen.

Für die Studie PRISMS wird das Verzerrungspotential auf Studienebene ebenfalls als niedrig eingestuft. Die Endpunkterheber sind verblindet und das ITT-Prinzip adäquat umgesetzt. Es gibt keinen Hinweis auf ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential beeinflussende Faktoren. Das Verzerrungspotential wird für den Endpunkt „Schübe“ insgesamt als niedrig eingestuft.

Stellen Sie die Ergebnisse für den Endpunkt Schübe für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Jährliche Schubrate-Übersicht

Tabelle 4-148: Ergebnisse für jährliche Schubrate aus RCT des indirekten und direkten Vergleichs – für die RRMS-Subpopulation

Studie	Patientenzahl (ITT)	Adjustierte jährliche Schubrate ^a 95 % KI []	Rate Ratio 95 % KI [] P (H ₀ : RR = 1)	Relative Risikoreduktion vs. Placebo bzw. Rebif® 44 µg 95 % KI []
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo				
TEMPO	T14 (n=332) RRMS-Patienten	0,355 [0,294; 0,429]	0,661 [0,531; 0,824] p = 0,0002	33,9 %
	Pb (n=329) RRMS-Patienten	0,537 [0,462; 0,623]		
TOWER	T14 (n=364) RRMS-Patienten	0,310 [0,261; 0,369]	0,628 [0,506; 0,779] p < 0,0001	37,2 %
	Pb (n=378) RRMS-Patienten	0,495 [0,425; 0,576]		
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Interferon beta-1a 44 µg, s.c., dreimal pro Woche				
TENERE	T14 (n=111)	0,259 [0,153; 0,438]	1,197 [0,623; 2,299] p = 0,5896 ^d	-19,7 %
	Rb (n=104)	0,216 [0,113; 0,415]		
Studien: Interferon beta-1a 44 µg, s.c., dreimal pro Woche, versus Placebo				
PRISMS	Rb (n = 184)	0,865	0,67 ^b [0,56; 0,79] ^b p < 0,0001 ^c	33 % [21%; 44%]
	Pb (n = 187)	1,280		

T14 = Teriflunomid 14 mg; Pb = Placebo; Rb = Rebif® 44 µg

^aIn den Teriflunomid-Studien abgeleitet aus einem Poisson-Modell mit der Anzahl bestätigter Schübe, deren Beginn zwischen Randomisierung und letzter Dosis der Studienmedikation lag, als abhängiger Variable,

Behandlung, EDSS zur Baseline ($\leq 3,5$ / $> 3,5$) und Region als unabhängigen Variablen und der log-transformierten Behandlungsdauer als Offset-Variable

In der PRISMS-Studie abgeleitet aus einem verallgemeinerten linearen Modell (GLM) mit dem Logarithmus als Link-Funktion, wobei angenommen wurde, dass sich die Varianzen proportional zu den Mittelwerten verhalten. Die jährlichen Schubraten wurden durch Division mit der Studiendauer von 2 Jahren abgeleitet

^babgeleitet aus der in der Originalpublikation genannten relativen Risiko-Reduktion und dem zugehörigen 95% KI (s. Zelle unmittelbar rechts daneben)

^cberechnet aus $\ln(\text{RR})$ und den logarithmierten Grenzen des 95 % KI ($\text{SE} = 0,0878 \Rightarrow z = -4,5622 \Rightarrow p(\text{H}_0: \text{Rate Ratio}=1) < 0,0001$)

^dberechnet aus $\ln(\text{RR})$ und den logarithmierten Grenzen des 95 % KI ($\text{SE} = 0,3331 \Rightarrow z = 0,5399 \Rightarrow p(\text{H}_0: \text{Rate Ratio} = 1) = 0,5893$)

Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an die übliche Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.*
- *Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.*
- *Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere Widersprüche zwischen direkter und indirekter Evidenz.*

Für die Durchführung des indirekten Vergleiches gelten die folgenden Festlegungen:

Tabelle 4-149: Definitionen für den indirekten Vergleich

Analyse	Definition
Metaanalyse 1	Metaanalyse von TEMSO und TOWER
Indirekter Vergleich	Kombination von Metaanalyse 1 mit PRISMS (nach Bucher)
Metaanalyse 2	Metaanalyse von TENERE und dem indirekten Vergleich

Jährliche Schubrate – indirekter Vergleich

Tabelle 4-150: Ergebnisse für die jährliche Schubrate aus RCT mit dem zu bewertenden Arzneimittel- 1. indirekter Vergleich – für die RRMS-Subpopulation

Analyse	Komparatoren	Rate Ratio 95 % KI [] p-Wert ^a	Relative Risiko- Reduktion
Metaanalyse 1 (TEMSO & TOWER)	Teriflunomid 14 mg versus Placebo	0,644 [0,552; 0,751] p < 0,0001 I ² =0%, p _Q =0,7444	35,6 %
Indirekter Vergleich (Metaanalyse 1.1 + PRISMS)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,961 [0,763; 1,211] p = 0,7369	
Metaanalyse 2 (Indirekter Vergleich 1 & TENERE)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,985 [0,792, 1,224] p = 0,8911 I ² =0%, p _Q =0,5349	

^aH₀: Rate Ratio = 1

Metaanalyse 1

Die Ergebnisse von TEMSO und TOWER zeigen eine sehr gute Homogenität (I² = 0%, p_Q=0,7444) und lassen sich statistisch gut zusammenführen. Die beiden Studien sind analog zueinander konzipiert worden, die Patientenkollektive, die Operationalisierung und die Erfassung der jährlichen Schubrate entsprachen sich. Deswegen lassen sich die in den zwei Studien ermittelten jährlichen Schubraten auch aus medizinischer Sicht gut zusammenfassen.

Die Ergebnisse der Metaanalyse 1 (0,644; 95 % KI: [0,552; 0,751]; p < 0,0001) sind dem Ergebnis der PRISMS-Studie sehr ähnlich (Rate Ratio von 0,67; 95 % KI: [0,56; 0,79]; p(H₀: Rate Ratio = 1) < 0,0001).

Indirekter Vergleich

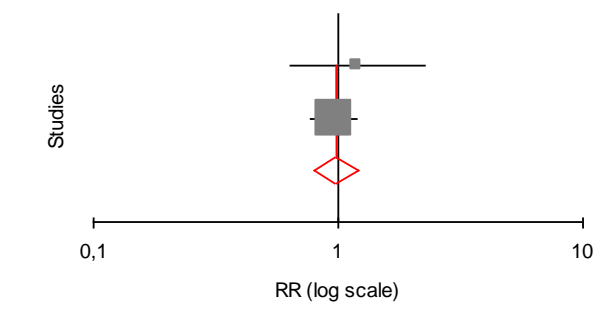
Durch den indirekten Vergleich wird die Annahme einer ähnlichen Reduktion der Schubrate durch 14 mg Teriflunomid und Rebif[®] 44 µg bestätigt. Hier liegt der Wert der Rate Ratio der jährlichen Schubrate für 14 mg Teriflunomid versus Rebif[®] 44 µg mit 0,961 (95 % KI: [0,763; 1,211]); p = 0,7369) nahe bei 1. Das Ergebnis des indirekten Vergleichs zeigt, dass die beiden Interventionen 14 mg Teriflunomid und Rebif[®] 44 µg die jährliche Schubrate in vergleichbarem Ausmaß senken.

Metaanalyse 2

Das Ergebnis dieses indirekten Vergleiches wurde über eine weitere Metaanalyse mit den Daten aus der direkt vergleichenden Studie TENERE auf Konsistenz geprüft (Metaanalyse 2). Die Ergebnisse zeigen eine hohe Homogenität ($I^2=0\%$, $p_Q=0,5349$) und auch hier einen Wert der Rate Ratio nahe bei 1 (Rate Ratio = 0,985; 95 % KI: [0,792; 1,224]; $p(H_0: \text{Rate Ratio} = 1) = 0,8911$).

Die Daten sind konsistent, das Ergebnis des indirekten Vergleichs, d. h. die vergleichbare Senkung der jährlichen Schubrate durch 14 mg Teriflunomid und Rebif[®] 44 µg steht nicht im Widerspruch zu dem Ergebnis der TENERE-Studie. Im Vergleich zur TENERE-Studie zeigt der indirekte Vergleich ein deutlich kürzeres Konfidenzintervall, d. h. eine deutlich verbesserte Präzision der Schätzung der Rate Ratio.

Abbildung 4-33: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Rate Ratio der jährlichen Schubrate



Fazit - Jährliche Schubrate

Die Ergebnisse des indirekten Vergleiches unterstützen die Aussage, dass die jährliche Schubrate durch 14 mg Teriflunomid und Rebif[®] 44 µg in vergleichbarem Ausmaß gesenkt wird. Die Präzision dieser Aussage wird dabei durch die indirekten Vergleiche deutlich gesteigert.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt, für den ein indirekter Vergleich vorgenommen wird, fortlaufend in einem eigenen Abschnitt dar.

4.3.2.1.4.2 Behinderungsprogression – indirekte Vergleiche aus RCT

Für die indirekten Vergleiche soll zunächst für jeden Endpunkt eine Übersicht über die verfügbaren Vergleiche gegeben werden. Anschließend soll die Darstellung der Ergebnisse in 3 Schritten erfolgen: 1) Bewertung des Verzerrungspotenzials auf Endpunktebene pro Studie, 2) tabellarische Darstellung der Ergebnisse der einzelnen Studien, 3) Darstellung des

indirekten Vergleichs. Für die Punkte 1 und 2 gelten die gleichen Anforderungen wie für die Darstellung der Ergebnisse der direkten Vergleiche in Abschnitt 4.3.1.3.

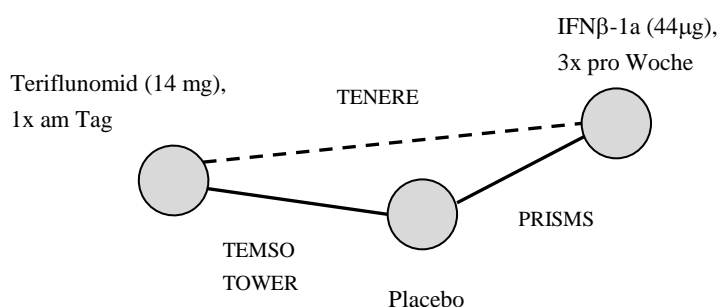
Geben Sie für den im vorliegenden Abschnitt präsentierten Endpunkt einen Überblick über die in den Studien verfügbaren Vergleiche.

Tabelle 4-151: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden

Anzahl Studien	Referenzen Studien	Intervention (14 mg Teriflunomid)	Placebo	Rebif® 44 µg (3 x wöchentlich)
2	TEM SO TOWER	• •	• •	
1	PRISMS		•	•

Stellen Sie zusätzlich die Netzwerkstruktur des indirekten Vergleichs grafisch dar.

Abbildung 4-34: Netzwerk für den indirekten Vergleich



Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-152: Operationalisierung von Behinderungsprogression

Studie	Operationalisierung
TEM SO	<p>Zeit bis zur 12 Wochen anhaltenden Behinderungsprogression (12week SDP = sustained disability progression):</p> <p>Definiert als die Zeit bis es zu einer 1-Punkt Erhöhung des EDSS-Scores im Vergleich zum Baseline-Wert gekommen ist (in den Fällen, in denen der Baseline-Score $\leq 5,5$ betrug) oder zu einer Erhöhung um mindestens 0,5 Punkte (in den Fällen mit einem Baseline-Score $> 5,5$). Diese Erhöhung des EDSS-Scores musste für mindestens 12 Wochen bestehen bleiben (12 Wochen anhaltende Behinderungsprogression).</p> <p>Die Zeit-bis-zum-Ereignis-Variable ist definiert als die Zeit in Tagen, die von dem Tag der Randomisierung bis zum Zeitpunkt der ersten Behinderungsprogression vergeht.</p>
TOWER	<p>Zeit bis zur 12 Wochen anhaltenden Behinderungsprogression (12week SDP = sustained disability progression):</p> <p>Definiert als die Zeit bis es zu einer 1-Punkt Erhöhung des EDSS-Scores im Vergleich zum Baseline-Wert gekommen ist (in den Fällen, in denen der Baseline-Score $\leq 5,5$ betrug) oder zu einer Erhöhung um mindestens 0,5 Punkte (in den Fällen mit einem Baseline-Score $> 5,5$). Diese Erhöhung des EDSS-Scores musste für mindestens 12 Wochen bestehen bleiben (12</p>

Wochen anhaltende Behinderungsprogression).

Die Zeit-bis-zum-Ereignis-Variable ist definiert als die Zeit in Tagen, die von dem Tag der Randomisierung bis zum Zeitpunkt der ersten Behinderungsprogression vergeht.

In jedem Fall, in dem eine Progression der Erkrankung festgestellt werden konnte, wurde der Patient dazu aufgefordert, erneut der Studienteilnahme zuzustimmen.

Bei jedem Arztbesuch wurde untersucht, ob der Patient eine anhaltende Krankheitsprogression erlitten hat.

- Eine erste anhaltende Krankheitsprogression (oder Behinderungsprogression) mit oder ohne Schub wurde definiert als eine persistierende Erhöhung von mindestens 1,0 Punkten auf der EDSS Skala bezogen auf den Baseline-Wert (bzw. 0,5 für Patienten, deren Baseline EDSS Score über 5,5 lag), die für mindestens 12 Wochen stabil war.
- Eine folgende anhaltende Behinderungsprogression war definiert als eine persistierende Erhöhung von mindestens 1,0 Punkten auf der EDSS Skala bezogen auf den Wert des letzten Arztbesuches (bzw. 0,5 für Patienten, deren letzter EDSS Score über 5,5 lag), die für mindestens 12 Wochen stabil war

Die Kombination von EDSS und FS (FS = Funktionssystem) Scores sind einer der ältesten und wahrscheinlich gebräuchlichsten Werkzeuge zur Erfassung des Krankheitsstatus in der MS. Basierend auf einer neurologischen Untersuchung werden die 7 Funktionssysteme (plus „andere“) eingestuft. Diese Einstufungen werden in Verbindung mit Beobachtungen und Informationen über den Gang und den Gebrauch von Gehilfen verwendet, um den EDSS festzulegen. Jeder der einzelnen FS Scores ist eine klinische Ordinalskala im Wertebereich von 0 bis 5 oder 6. Der EDSS ist eine klinische Ordinalskala im Wertebereich von 0 (normale neurologische Funktion) bis 10 (Tod durch MS) in Stufen von 0,5.

Der FS Score und der EDSS wurden durch den untersuchten Neurologen persönlich durchgeführt. Die Durchführungsdauer variierte abhängig vom Zustand des Patienten und der Fähigkeit des Untersuchers. Der gleiche Untersucher musste die Bewertung so gut wie möglich zu jeglicher Arztbesuch durchführen.

Bestätigung der Behinderung:

In den Fällen, wo ein Patient eine erste Behinderungsprogression während der Behandlungszeit erfahren hat und zum Zeitpunkt des Arztbesuchs, bei dem das Auswaschverfahren gestartet wurde, noch keine 12-Wochen-Bestätigung vorlag, wurde der Patient für eine EDSS-Bewertung in die Klinik bestellt. Wenn dann eine Behinderungsprogression bestätigt werden konnte, wurde der Patient für eine letzte Bewertung des EDSS-Grades zum Zeitpunkt 24 Wochen einbestellt.

Im Fall einer vorzeitigen Beendigung der Studie war die Effektivitätsbewertung entsprechend am Ende des Behandlungszeitraumes durchzuführen.

PRISMS	<p>Zeit bis zur Behinderungsprogression</p> <p>Eine Behinderungsprogression war definiert als eine Zunahme des EDSS um mindestens 1 Punkt, die über mindestens 3 Monate anhielt.</p>
--------	---

Operationalisierungen der Behinderungsprogression, die in den indirekten Vergleich eingeschlossen sind

Für den indirekten Vergleich kommt lediglich die „Zeit bis zur Behinderungsprogression (bestätigt nach 12 Wochen)“ als Operationalisierung in Frage. Andere Operationalisierungen des Endpunktes Behinderungsprogression sind z. B. eine mittlere Änderung des EDSS oder des MSFC-Scores und der Anteil an Patienten mit oder ohne Behinderungsprogression. Diese

Operationalisierungen sind in den vorliegenden Studien nicht gleichermaßen erfasst, so dass ein indirekter Vergleich auf Basis dieser Daten nicht möglich ist.

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-153: Bewertung des Verzerrungspotenzials für Zeit bis zur Behinderungsprogression in RCT für indirekte Vergleiche

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
TEMPO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
PRISMS	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Für die Bewertung des Verzerrungspotentials hinsichtlich des Endpunktes Behinderungsprogression in den Studien TEMPO und TOWER sei an dieser Stelle auf Modul 4.3.1.3.2 verwiesen.

Bei PRISMS handelt es sich um eine doppelblinde, randomisierte Studie. Das ITT-Prinzip ist adäquat umgesetzt, es gibt keine Hinweise auf eine ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Faktoren. Das Verzerrungspotential für den Endpunkt Behinderungsprogression wird für die Studie als niedrig eingestuft.

Stellen Sie die Ergebnisse für den Endpunkt Schübe für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Zeit bis zur Behinderungsprogression-Übersicht

Tabelle 4-154: Ergebnisse für Zeit bis zur Behinderungsprogression aus RCT des indirekten und direkten Vergleichs – für die RRMS-Subpopulation

Studie	Patientenzahl (ITT)	Anzahl Patienten mit Behinderungsprogression (Anteil)	Anzahl Patienten, die zensiert wurden (Anteil)	Hazard Ratio 95 % KI [] p-Wert ^a	Relative Risiko-Reduktion vs. Placebo bzw. Rebif® 44 µg
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo					
TEM SO	T14 (n= 332) RRMS Patienten	59 (17,8 %)	273 (82,2 %)	0,721 [0,514; 1,011]	27,9 %
	Pb (n = 329) RRMS Patienten	78 (23,7 %)	251 (76,3%)		
TOWER	T14 (n=364) RRMS- Patienten	42 (11,5%)	322 (88,5%)	0,663 [0,449; 0,979]	33,7 %
	Pb (n=378) RRMS- Patienten	64 (16,9%)	314 (83,1%)		
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Interferon beta-1a 44 µg, s. c., dreimal pro Woche					
TENERE	T14 (n=111)	10 (9,0 %)	101 (91,0 %)	0,982 [0,399; 2,420] p = 0,9690	1,8 %
	Rb (n=104)	9 (8,7 %)	95 (91,3 %)		
Studien: Interferon beta-1a 44 µg, s. c., dreimal pro Woche, versus Placebo					
PRISMS	Rb (n = 184)			0,62 [0,43; 0,91] p < 0,05	38 %

Pb (n = 187)

^aLog-Rank Test

Der indirekte Vergleich ist nur auf Grundlage von Hazard Ratios möglich, da Häufigkeitsangaben zur Behinderungsprogression in PRISMS [78] fehlen.

Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an die übliche Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.*
- *Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.*
- *Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere Widersprüche zwischen direkter und indirekter Evidenz.*

Für die Durchführung des indirekten Vergleiches gelten die folgenden Festlegungen:

Tabelle 4-155: Definitionen für den indirekten Vergleich

Analyse	Definition	
Indirekter Vergleich	Metaanalyse 1	Metaanalyse von TEMSO und TOWER
	Indirekter Vergleich	Kombination von Metaanalyse 1.1 mit PRISMS (nach Bucher)
	Metaanalyse 2	Metaanalyse von TENERE und dem indirekten Vergleich 1

Zeit bis zur Behinderungsprogression - Indirekter Vergleich

Tabelle 4-156: Ergebnisse für die Zeit bis zur Behinderungsprogression aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die RRMS-Subpopulation

Analyse	Komparatoren	Hazard Ratio 95 % KI [] p-Wert ^a
Metaanalyse 1 (TEMSO & TOWER)	Teriflunomid 14 mg versus Placebo	0,696 [0,539; 0,898] p = 0,0053 I ² =0%, p _Q =0,7501
Indirekter Vergleich (Metaanalyse 1 + PRISMS)	Teriflunomid 14 mg versus Rebif 44 µg	1,122 [0,713; 1,766] p = 0,6195

Metaanalyse 2 (Indirekter Vergleich & TENERE)	Teriflunomid 14 mg versus Rebif 44 µg	1,092 [0,728; 1,638] p = 0,6701 I ² =0%, p _Q =0,7965
---	--	---

^aH₀: HR = 1

Metaanalyse 1

Die Ergebnisse von TEMSO und TOWER zeigen eine sehr gute Homogenität (I²=0%, p_Q=0,7501) und lassen sich statistisch gut zusammenführen. Die beiden Studien sind analog zueinander konzipiert worden, die Patientenkollektive und die Operationalisierung der Behinderungsprogression sind aus medizinischer Sicht hinreichend ähnlich, um die Ergebnisse zusammenführen zu können.

Das Ergebnis der Metaanalyse 1 (Hazard Ratio von 0,696; 95 % KI: [0,539; 0,898]) entspricht einer relativen Risikoreduktion von 30,4 % und liegt etwas niedriger als der entsprechende Wert in der PRISMS-Studie (38 %).

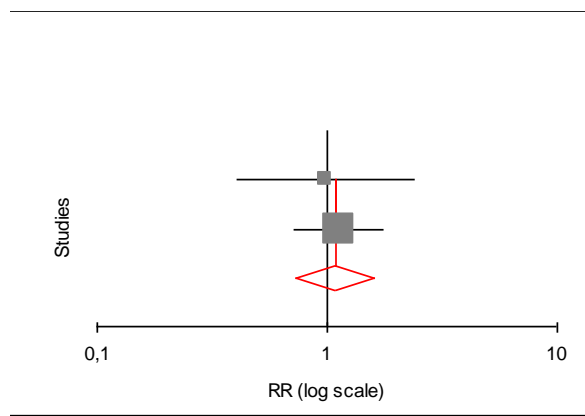
Indirekter Vergleich 1

Das Ergebnis des indirekten Vergleichs zeigt mit einer Hazard Ratio von 1,122 und einem 95 % KI von [0,713; 1,766], dass die beiden Interventionen die Behinderungsprogression in einem ähnlichen Ausmaß verlangsamen.

Metaanalyse 2

Das Ergebnis des indirekten Vergleichs bezogen auf die Zeit bis zur Behinderungsprogression wurde über eine weitere Metaanalyse mit den Daten aus der TENERE-Studie auf Konsistenz geprüft (Metaanalyse 2). Die Ergebnisse zeigen eine gute Homogenität (I²=0%, p_Q=0,7965). Mit einer Hazard Ratio von 1,092 und einem zugehörigen 95 % Konfidenzintervall von [0,728; 1,638] steht das Ergebnis des indirekten Vergleichs nicht im Widerspruch zu den Ergebnissen der direkt vergleichenden Studie TENERE und die Verlangsamung der Behinderungsprogression durch die beiden Interventionen kann als vergleichbar angesehen werden.

Abbildung 4-35: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Hazard Ratio der Zeit bis zur Behinderungsprogression



Zeit bis zur Behinderungsprogression - Fazit

Die Zeit bis zur Behinderungsprogression war der wesentliche sekundäre Endpunkt in den beiden placebokontrollierten Teriflunomid-Studien TEMSO und TOWER. Auch in der PRISMS-Studie war es ein wesentlicher sekundärer Endpunkt. Auch in diesem wesentlichen Zielkriterium konnte sowohl für 14 mg Teriflunomid als auch für die zweckmäßige Vergleichstherapie Rebif® 44 µg ein positiver Effekt in einem vergleichbaren Ausmaß gezeigt werden.

Im Vergleich zu TENERE zeigt der indirekte Vergleich ein deutlich kürzeres Konfidenzintervall, d. h. eine deutlich verbesserte Präzision der Schätzung der Hazard Ratio.

4.3.2.1.4.3 Mortalität – indirekte Vergleiche aus RCT

Für die indirekten Vergleiche soll zunächst für jeden Endpunkt eine Übersicht über die verfügbaren Vergleiche gegeben werden. Anschließend soll die Darstellung der Ergebnisse in 3 Schritten erfolgen: 1) Bewertung des Verzerrungspotenzials auf Endpunktebene pro Studie, 2) tabellarische Darstellung der Ergebnisse der einzelnen Studien, 3) Darstellung des indirekten Vergleichs. Für die Punkte 1 und 2 gelten die gleichen Anforderungen wie für die Darstellung der Ergebnisse der direkten Vergleiche in Abschnitt 4.3.1.3.

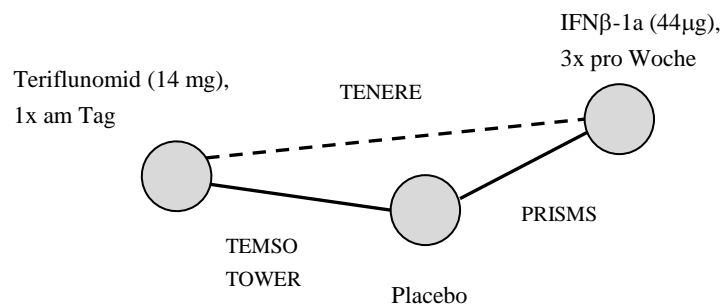
Geben Sie für den im vorliegenden Abschnitt präsentierten Endpunkt einen Überblick über die in den Studien verfügbaren Vergleiche.

Tabelle 4-157: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden

Anzahl Studien	Referenzen Studien	Intervention (14 mg Teriflunomid)	Placebo	Rebif® 44 µg (3 x wöchentlich)
2	TEMPO TOWER	• •	• •	
1	PRISMS		•	•

Stellen Sie zusätzlich die Netzwerkstruktur des indirekten Vergleichs grafisch dar.

Abbildung 4-36: Netzwerk für den indirekten Vergleich



Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-158: Operationalisierung von Mortalität

Studie	Operationalisierung
TEMPO	Mortalität wird bestimmt als die Anzahl an Todesfällen, die während der Behandlung aufgetreten sind.
TOWER	Mortalität wird bestimmt als die Anzahl an Todesfällen, die während der Behandlung aufgetreten sind.
PRISMS	Mortalität wird bestimmt als die Anzahl an Todesfällen, die während der Behandlung aufgetreten sind.

Die Operationalisierung der Mortalität ist in allen drei Studien identisch.

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-159: Bewertung des Verzerrungspotenzials für Mortalität in RCT für indirekte Vergleiche

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo						
TEMPO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
Studien: Interferon beta-1a 44 µg, s.c., dreimal pro Woche, versus Placebo						
PRISMS	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Für die Bewertung des Verzerrungspotentials hinsichtlich des Endpunktes Mortalität in den Studien TEMPO und TOWER sei an dieser Stelle auf Modul 4.3.1.3.8 verwiesen.

Für die Studie PRISMS wird das Verzerrungspotential auf Studienebene ebenfalls als niedrig eingestuft. Die Endpunkterheber sind verblindet und das ITT-Prinzip adäquat umgesetzt. Es

gibt keinen Hinweis auf ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential beeinflussende Faktoren. Das Verzerrungspotential wird für den Endpunkt Mortalität insgesamt als niedrig eingestuft.

Stellen Sie die Ergebnisse für den Endpunkt Mortalität für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Mortalität-Übersicht

Tabelle 4-160: Ergebnisse für Mortalität aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation

Studie	Patientenzahl (Safety-Population ^a)	Anzahl verstorbener Patienten (prozentualer Anteil)	Risiko-Differenz (RD) ^b 95 % KI [] p(H ₀ : R=1)
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo			
TEMZO	T14 (n=358)	0 (0%)	0,0000 [-0,0054; 0,0054] p = 1,0000
	Pb (n=360)	0 (0%)	
TOWER	T14 (n=371)	2 (0,5%)	0,0028 [-0,0062; 0,0118] p = 0,5439
	Pb (n=385)	1 (0,3%)	
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Interferon beta-1a 44 µg, s. c., dreimal pro Woche			
TENERE	T14 (n=110)	0 (0%)	0,0000 [-0,0184; 0,0184] p = 1,0000
	Rb (n=101)	0 (0%)	
Studien: Interferon beta-1a 44 µg, s. c., dreimal pro Woche, versus Placebo			
PRISMS	Rb (n = 184)	0 (0%)	-0,0053 [-0,0201; 0,0094] p = 0,4775
	Pb (n = 187)	1 (0,5 %)	

T14 = Teriflunomid 14 mg; Pb = Placebo; Rb = Rebif 44[®] µg

^as. Definition der Safety-Population für die Teriflunomid-Studien unter 4.3.1.3.9 (Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten)), für PRISMS ist keine Abweichung der

Safety-Population von der ITT-Population angegeben, so dass von der Annahme ausgegangen wird, dass zwischen den entsprechenden Populationen kein Unterschied besteht.

^bÜbergang zur Analyse von Risiko-Differenzen statt Risiko-Quotienten, da sowohl TEMSO als auch TENERE „Doppel-Null-Studien“ sind, d. h. keine Todesfälle in den zu vergleichenden Behandlungsgruppen auftraten.

Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an die übliche Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.*
- *Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.*
- *Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere Widersprüche zwischen direkter und indirekter Evidenz.*

Für die Durchführung des indirekten Vergleichs gelten die folgenden Festlegungen:

Tabelle 4-161: Definitionen für den indirekten Vergleich

Analyse	Definition
Metaanalyse 1	Metaanalyse von TEMSO und TOWER
Indirekter Vergleich	Kombination von Metaanalyse 1.1 mit PRISMS (nach Bucher)
Metaanalyse 2	Metaanalyse von TENERE und dem indirekten Vergleich

Mortalität - Indirekter Vergleich

Tabelle 4-162: Ergebnisse für Mortalität aus RCT mit dem zu bewertenden Arzneimittel-indirekter Vergleich 1 – für die gesamte Studienpopulation

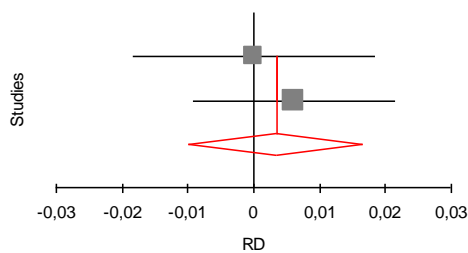
Analyse	Komparatoren	Risiko-Differenz (RD) 95 % KI p(H ₀ : RD=0)
Metaanalyse 1 (TEMSO & TOWER)	Teriflunomid 14 mg versus Placebo	0,0007 [-0,0039; 0,0054] p = 0,7539 I ² =0%, p _Q =0,6032
Indirekter Vergleich (Metaanalyse 1 + PRISMS)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,0061 [-0,0094; 0,0216] p = 0,4403
Metaanalyse 2 (Indirekter Vergleich & TENERE)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,0035 [-0,0098; 0,0168] p = 0,6094

$$I^2=0\%, p_Q=0,6195$$

In den vier Studien TEMSO, TOWER, TENERE und PRISMS sind erwartungsgemäß nur vereinzelt Todesfälle aufgetreten. Mit TEMSO und TENERE sind zwei Studien dabei, in denen kein Todesfall aufgetreten ist („Doppel-Null-Studien“). Deswegen ist ein Vergleich der Studien basierend auf relativen Risiken nicht möglich. Die Metaanalysen und der indirekte Vergleich basieren in diesem Fall auf den Risikodifferenzen.

Beide Metaanalysen (Metaanalyse 1 und Metaanalyse 2) zeigten eine sehr gute Homogenität der Ergebnisse. In allen Studien sind die aufgetretenen Todesfälle zufällig verteilt, ein Unterschied zwischen den Behandlungsgruppen ist nicht erkennbar.

Abbildung 4-37: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risikodifferenz der Mortalität



Mortalität - Fazit

Weder durch das zu bewertende Arzneimittel Teriflunomid 14 mg noch durch die zweckmäßige Vergleichstherapie Rebif® 44 µg wird in den vorliegenden Studien eine Erhöhung des Mortalitätsrisikos gezeigt, es konnte kein Unterschied zwischen den Interventionen festgestellt werden.

4.3.2.1.4.4 Unerwünschte Ereignisse (die zum Behandlungsabbruch führten) – indirekte Vergleiche aus RCT

Für die indirekten Vergleiche soll zunächst für jeden Endpunkt eine Übersicht über die verfügbaren Vergleiche gegeben werden. Anschließend soll die Darstellung der Ergebnisse in 3 Schritten erfolgen: 1) Bewertung des Verzerrungspotenzials auf Endpunktebene pro Studie, 2) tabellarische Darstellung der Ergebnisse der einzelnen Studien, 3) Darstellung des indirekten Vergleichs. Für die Punkte 1 und 2 gelten die gleichen Anforderungen wie für die Darstellung der Ergebnisse der direkten Vergleiche in Abschnitt 4.3.1.3.

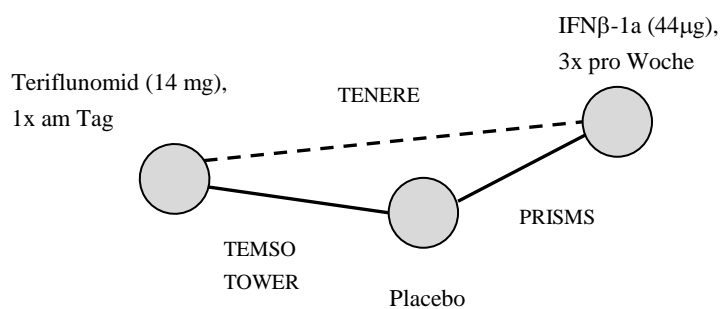
Geben Sie für den im vorliegenden Abschnitt präsentierten Endpunkt einen Überblick über die in den Studien verfügbaren Vergleiche.

Tabelle 4-163: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden

Anzahl Studien	Referenzen Studien	Intervention (14 mg Teriflunomid)	Placebo	Rebif® 44 µg (3 x wöchentlich)
2	TEMSO TOWER	• •	• •	
1	PRISMS		•	•

Stellen Sie zusätzlich die Netzwerkstruktur des indirekten Vergleichs grafisch dar.

Abbildung 4-38: Netzwerk für den indirekten Vergleich



Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-164: Operationalisierung von „Unerwünschte Ereignisse, die zum Abbruch der Behandlung führten“

Studie	Operationalisierung
TEMSO	Anzahl an Patienten, die die Behandlung aufgrund von unerwünschten Ereignissen abgebrochen haben
TOWER	Anzahl an Patienten, die die Behandlung aufgrund von unerwünschten Ereignissen abgebrochen haben
PRISMS	Anzahl an Patienten, die die Behandlung aufgrund von unerwünschten Ereignissen abgebrochen haben

Die Operationalisierung der unerwünschten Ereignisse, die zu einem Abbruch der Behandlung geführt haben, ist in allen vier Studien identisch.

Operationalisierungen der unerwünschten Ereignisse, die in den indirekten Vergleich eingeschlossen werden

In den indirekten Vergleichen sind lediglich die unerwünschten Ereignisse, die zu einem Behandlungsabbruch führten und ausgewählte unerwünschte Ereignisse von besonderem Interesse (s. u.) berücksichtigt.

Übliche weitere Operationalisierung der unerwünschten Ereignisse sind die Gesamtrate der unerwünschten Ereignisse und die Gesamtrate schwerwiegender unerwünschter Ereignisse. Diese Operationalisierungen können im indirekten Vergleich nicht berücksichtigt werden, weil die entsprechenden Angaben in den Publikationen der PRISMS- Studie fehlen. Für Teriflunomid finden sich die entsprechenden Auswertungen in Modul 4.3.1.3.9.

Unerwünschte Ereignisse von besonderem Interesse sind unter 4.3.2.1.2.5 gesondert aufgeführt.

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-165: Bewertung des Verzerrungspotenzials für „Unerwünschte Ereignisse, die zum Abbruch der Behandlung führten“ in RCT für indirekte Vergleiche

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo						
TEMPO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
Studien: Interferon beta-1a 44 µg, s.c., dreimal pro Woche, versus Placebo						
PRISMS	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Für die Bewertung des Verzerrungspotentials hinsichtlich des Endpunktes „Unerwünschte Ereignisse, die zu einem Behandlungsabbruch führten“ in den Studien TEMPO und TOWER sei an dieser Stelle auf Modul 4.3.1.3.9 verwiesen.

Für die Studie PRISMS wird das Verzerrungspotential auf Studienebene ebenfalls als gering eingestuft. Die Endpunkterheber sind verblindet und das ITT-Prinzip ist adäquat umgesetzt. Es gibt keinen Hinweis auf ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Faktoren. Das Verzerrungspotential wird für den Endpunkt „Unerwünschte Ereignisse, die zu einem Abbruch der Behandlung führten“ insgesamt als niedrig eingestuft.

Stellen Sie die Ergebnisse für den Endpunkt unerwünschte Ereignisse, die zu einem Behandlungsabbruch führten für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Unerwünschte Ereignisse, die zu einem Behandlungsabbruch führten-Übersicht

Tabelle 4-166: Ergebnisse für „Unerwünschte Ereignisse, die zu einem Behandlungsabbruch führten“ aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation

Studie	Patientenzahl (Safety- Population ^a)	Anzahl Patienten, die aufgrund von unerwünschten Ereignissen die Studie abgebrochen haben (prozentualer Anteil)	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo				
TEMSO	T14 (n=358)	39 10,9 %	1,352 [0,856; 2,137] p = 0,1963	1,395 [0,842; 2,311] p = 0,1956
	Pb (n=360)	29 8,1 %		
TOWER	T14 (n=371)	58 15,6 %	2,508 [1,593; 3,948] p < 0,0001	2,787 [1,692; 4,592] p < 0,0001
	Pb (n=385)	24 6,2 %		
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Interferon beta-1a 44 µg, s. c., dreimal pro Woche				
TENERE	T14 (n=110)	12 10,9 %	0,501 [0,262; 0,959] p = 0,0369	0,440 [0,205; 0,943] p = 0,0349
	Rb (n=101)	22 21,8 %		
Studien: Interferon beta-1a 44 µg, s. c., dreimal pro Woche, versus Placebo				
PRISMS	Rb (n = 184)	9 (4,9 %)	4,573 [1,002; 20,88] p = 0,0498	4,757 [1,014; 22,32] p = 0,0480
	Pb (n = 187)	2 (1,1 %)		

^as. Definition der Safety-Population für die Teriflunomid-Studien unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten), für PRISMS ist keine Abweichung der Safety-Population von der ITT-Population angegeben, so dass von der Annahme ausgegangen wird, dass zwischen den entsprechenden Populationen kein Unterschied besteht.

^bH₀: RR = 1 bzw. H₀: OR = 1

Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an die übliche Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.*
- *Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.*
- *Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere Widersprüche zwischen direkter und indirekter Evidenz.*

Für die Durchführung des indirekten Vergleiches gelten die folgenden Festlegungen:

Tabelle 4-167: Definitionen für den indirekten Vergleich

Analyse	Definition
Metaanalyse 1	Metaanalyse von TEMSO und TOWER
Indirekter Vergleich	Kombination von Metaanalyse 1 mit PRISMS (nach Bucher)
Metaanalyse 2	Metaanalyse von TENERE und dem indirekten Vergleich

Unerwünschte Ereignisse, die zu einem Behandlungsabbruch führten-indirekter Vergleich

Tabelle 4-168: Ergebnisse für „Unerwünschte Ereignisse, die zu einem Behandlungsabbruch führten“ aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse 1 (TEMSO & TOWER)	Teriflunomid 14 mg versus Placebo	DL ^a : 1,843 [1,006; 3,376] p = 0,0477 I ² =72%, p _Q =0,0604	DL ^a : 1,974 [1,002; 3,889] p = 0,0493 I ² =73%, p _Q =0,0561
Indirekter Vergleich (Metaanalyse 1 + PRISMS)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,403 [0,079; 2,067] p = 0,2759	0,415 [0,077; 2,245] p = 0,3072
Metaanalyse 2 (Indirekter Vergleich & TENERE)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,486 [0,266; 0,889] p = 0,0192 I ² =0%, p _Q =0,8086	0,435 [0,217; 0,873] p = 0,0191 I ² =0%, p _Q =0,9512

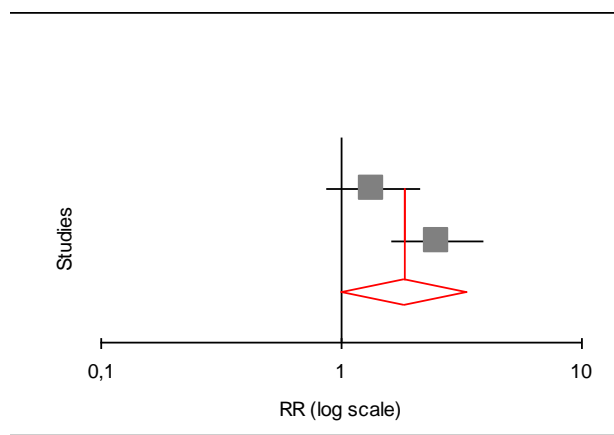
^aaufgrund der Heterogenität ist I²>0, so dass die Metaanalyse mit Studie als zufälligem Effekt (nach DerSimonian und Laird (=DL)) andere Ergebnisse liefert als das Modell mit Studie als festem Effekt. Verwendet und interpretiert wurden in diesen Fällen ausschließlich die Ergebnisse nach DerSimonian und Laird

^bH₀: RR = 1 bzw. H₀: OR = 1

Metaanalyse 1

Die Ergebnisse von TEMSO und TOWER zeigen, dass sich die beiden placebokontrollierten Studien in Bezug auf diesen Endpunkt unterscheiden. Die metaanalytische Zusammenfassung ist mit $I^2=72\%$ und $p_Q=0,0604$ heterogen. Es liegen allerdings deutliche gleichgerichtete Effekte vor, deshalb kann die Metaanalyse dennoch für den indirekten Vergleich in Betracht gezogen werden.

Abbildung 4-39: Forest-Plot zur Metaanalyse 1 (von oben nach unten: TEMSO, TOWER, Ergebnis der Metaanalyse) – Risk Ratio



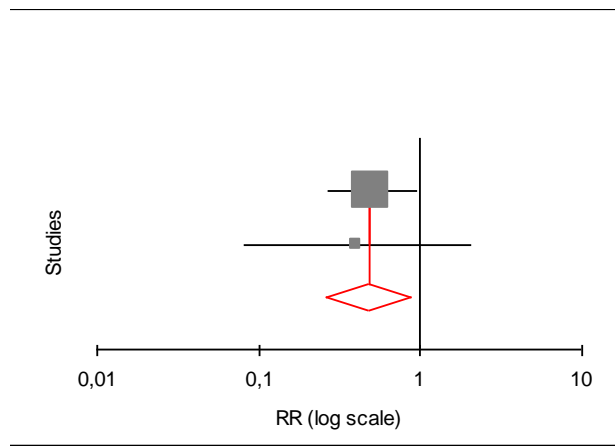
Indirekter Vergleich

Der indirekte Vergleich zeigt mit einem Risk Ratio von 0,403 einen deutlichen Vorteil für 14 mg Teriflunomid gegenüber Rebif® 44 µg, das Ergebnis war jedoch nicht statistisch signifikant.

Metaanalyse 2

Die Ergebnisse des indirekten Vergleichs passen gut mit den Ergebnissen der direkt vergleichenden TENERE-Studie zusammen, die Werte sind ausgesprochen homogen ($I^2=0\%$, $p_Q=0,8086$). Mit einem Risk Ratio von 0,486 war das Risiko, die Studie aufgrund von Nebenwirkungen abzubrechen, unter 14 mg Teriflunomid gegenüber Rebif® 44 µg um die Hälfte reduziert. Durch den Gewinn an Präzision wird in der Metaanalyse 2 ein p-Wert von 0,0192 und damit - wie bereits in TENERE ($p = 0,0369$) - ein signifikantes Ergebnis erreicht.

Abbildung 4-40: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risk Ratio



Unerwünschte Ereignisse, die zu einem Behandlungsabbruch führten - Fazit

Aus dem Risk Ratio von 0,486 und dem zugehörigen Konfidenzintervall von [0,266, 0,889] lässt sich für den Endpunkt „Unerwünschte Ereignisse, die zu einem Abbruch der Therapie geführt haben“ ein Zusatznutzen von 14 mg Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie ableiten. Wie in Modul 3 dargelegt ist dieser Endpunkt für das spezielle Anwendungsgebiet des zu bewertenden Arzneimittels (krankheitsmodifizierende Therapie der RRMS) von besonderer Relevanz, da die Therapieadhärenz für die langfristige Wirksamkeit der Dauertherapie ein kritischer Punkt ist.

4.3.2.1.4.5 Unerwünschte Ereignisse von besonderem Interesse – indirekte Vergleiche aus RCT

Für die indirekten Vergleiche soll zunächst für jeden Endpunkt eine Übersicht über die verfügbaren Vergleiche gegeben werden. Anschließend soll die Darstellung der Ergebnisse in 3 Schritten erfolgen: 1) Bewertung des Verzerrungspotenzials auf Endpunktebene pro Studie, 2) tabellarische Darstellung der Ergebnisse der einzelnen Studien, 3) Darstellung des indirekten Vergleichs. Für die Punkte 1 und 2 gelten die gleichen Anforderungen wie für die Darstellung der Ergebnisse der direkten Vergleiche in Abschnitt 4.3.1.3.

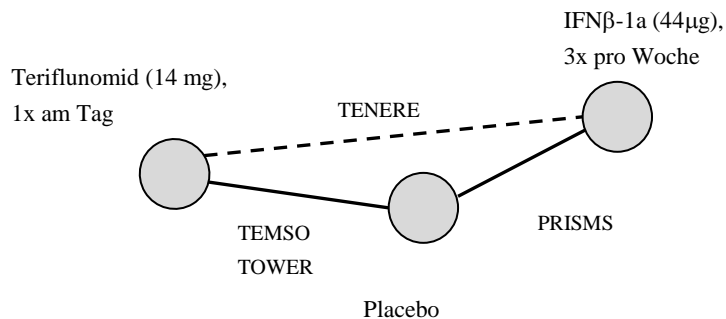
Geben Sie für den im vorliegenden Abschnitt präsentierten Endpunkt einen Überblick über die in den Studien verfügbaren Vergleiche.

Tabelle 4-169: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden – für die gesamte Studienpopulation

Anzahl Studien	Referenzen Studien	Intervention (14 mg Teriflunomid)	Placebo	Rebif® 44 µg (3 x wöchentlich)
2	TEMPO TOWER	• •	• •	
1	PRISMS		•	•

Stellen Sie zusätzlich die Netzwerkstruktur des indirekten Vergleichs grafisch dar.

Abbildung 4-41: Netzwerk für den indirekten Vergleich



Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-170: Operationalisierung von „Unerwünschte Ereignisse von besonderem Interesse“

Studie	Operationalisierung
Allgemeine Angaben	<p>Allgemeine Anmerkung: Unerwünschte Ereignisse, einschließlich abnormer Laborwerte, wurden in den Teriflunomid- Studien nach dem MedDRA® (Version 13.0 oder höher) Kodiersystem kodiert. TEMSO: Version 13.0; TOWER: Version 15.0; TENERE: Version 14.0</p> <hr/> <p>TEMSO: Termine für „adverse event reporting“ waren – abgesehen vom Datum der Randomisierung- angesetzt in den Wochen 4, 8, 12, 16, 20, 24 und 30 und dann alle 6 Wochen bis zum Studienende.</p> <p>TOWER und TENERE: Entsprechende Reportings erfolgten in Woche 2 und Woche 6 und dann alle 6 Wochen bis Woche 24, dann in Woche 36 und ab diesem Termin alle 12 Wochen bis zum Studienende.</p> <p>PRISMS: Hämatologische und biochemische Tests (inklusive Leberfunktionstests) wurden in den ersten 8 Wochen der Studie alle zwei Wochen durchgeführt, anschließend alle 3 Monate. Schilddrüsenfunktionstests wurden alle 6 Monate durchgeführt.</p>
TEMSO	<p>Folgende unerwünschte Ereignisse waren im Studienprotokoll als AESI prädefiniert:</p> <ul style="list-style-type: none"> • Leberfunktionsstörungen • Erkrankungen des Knochenmarks (jegliche Neutrophilenzahl von unter 1 Giga/L) • Erkrankungen des Pankreas (bestätigter Serumspiegel der Amylase oder Lipase auf >2fache des oberen Normwertes, abdominaler Ultraschall oder CT oder MRT Scans mit dem Befund einer pankreatischen Auffälligkeit) • Malignität • Überempfindlichkeit und möglicherweise damit zusammenhängende Störungen • Alopezie • Lungenfunktionsstörungen und Ereignisse, die potentiell mit einer interstitiellen Lungenerkrankung in Verbindung stehen

	<ul style="list-style-type: none"> • Bluthochdruck • Periphere Neuropathie
TOWER	<p>Folgende unerwünschte Ereignisse waren im Studienprotokoll als AESI prädefiniert:</p> <ul style="list-style-type: none"> • Gastrointestinale Störungen (Übelkeit und Durchfall) • Leberfunktionsstörungen • Erkrankungen des Pankreas • Erkrankungen des Knochenmarks • Infektionen und Infestationen • Überempfindlichkeit/Hauterkrankungen • Malignität • Bluthochdruck • Herzrhythmusstörungen • Lungenfunktionsstörungen und Ereignisse, die potentiell mit einer interstitiellen Lungenerkrankung in Verbindung stehen • Thrombose und Embolie • Hämorrhagie • Periphere Neuropathie und potentiell damit zusammenhängende Ereignisse • Konvulsionen • Alopezie
TENERE	<p>Folgende unerwünschte Ereignisse waren im Studienprotokoll als AESI prädefiniert:</p> <ul style="list-style-type: none"> • Gastrointestinale Störungen (Übelkeit und Durchfall) • Leberfunktionsstörungen • Erkrankungen des Pankreas • Immunsuppression (Erkrankungen des Knochenmarks) • Infektionen und Infestationen • Malignität • Bluthochdruck • Herzrhythmusstörungen • Lungenfunktionsstörungen und Ereignisse, die potentiell mit einer interstitiellen Lungenerkrankung in Verbindung stehen • Hämorrhagie • Periphere Neuropathie und potentiell damit zusammenhängende Ereignisse • Konvulsionen • AESI in anderen SOC (depression and suicide) • Alopezie
PRISMS	<p>Berichtete unerwünschte Ereignisse von besonderem Interesse (aufgetreten in den ersten 3 Monaten der Studie):</p> <ul style="list-style-type: none"> • Kopfschmerzen • Grippeähnliche Symptome • Reaktionen an der Injektionsstelle • Fatigue • Myalgie • Fieber • Lymphopenie

	<ul style="list-style-type: none"> • Erhöhung der Alanin-Aminotransferase-Werte • Leukopenie • Erhöhung der Aspartat-Aminotransferase-Werte • Granulocytopenie
AESI im vorliegenden indirekten Vergleich	<p>Unerwünschte Ereignisse von besonderem Interesse oder auch AESI (=adverse events of special interest) im Sinne des vorliegenden indirekten Vergleichs sind solche, die bei der Behandlung mit der Intervention (14 mg Teriflunomid) oder der Vergleichstherapie (Rebif® 44 µg) zu erwarten sind und bei diesen Wirkstoffen möglicherweise besondere Sicherheitsrisiken darstellen. Die Auswahl begründet sich einerseits auf Informationen der jeweiligen Fachinformationen, andererseits sollen die aufgeführten unerwünschten Ereignisse möglichst vorab im Studienprotokoll definiert sein.</p> <p>Es können nur solche AESI in den indirekten Vergleich aufgenommen werden, für die sich die entsprechenden Daten extrahieren lassen.</p> <p>Dementsprechend werden im indirekten Vergleich des vorliegenden Dossiers folgende unerwünschte Ereignisse von besonderem Interesse berücksichtigt:</p> <ul style="list-style-type: none"> • Grippeähnliche Symptome • Kopfschmerz • Erhöhung der Alanin-Aminotransferase-Werte • Leukopenie • Lymphopenie <p>Reaktionen an der Injektionsstelle können bei der Therapie mit 14 mg Teriflunomid nicht auftreten, weil bei einer oralen Darreichungsform dieses unerwünschte Ereignis nicht in Betracht kommen. Deswegen sind Reaktionen an der Injektionsstelle nicht im indirekten Vergleich ausgewertet, sondern lediglich beschreibend dargestellt.</p> <p>Bezogen auf Veränderungen des Blutbildes ist als besonderes Sicherheitsrisiko für Teriflunomid Neutropenie aufgeführt und im direkten Vergleich ausgewertet worden. Die entsprechende Auswertung liegt für Rebif® 44 µg nicht vor. Deswegen werden für den indirekten Vergleich Leukopenie und Lymphopenie, die in den Publikationen zu Studien mit der zweckmäßigen Vergleichstherapie aufgeführt sind, zusätzlich zu den AESI in Modul 4.3.1 ausgewertet.^a</p> <p>Eine Erhöhung der Leberenzyme ist sowohl für 14 mg Teriflunomid als auch für Rebif® 44 µg als besonderes Sicherheitsrisiko anzusehen. In allen fünf Studien waren die Werte der Alanin-Amino-Transferase häufiger erhöht als die Werte der Aspartat-Aminotransferase Erhöhung, so dass die Auswertung der ALT-Erhöhung als aussagekräftig angesehen wird, für die Auswertungen der anderen Leberenzyme sei auf die entsprechenden Studienberichte/Publikationen verwiesen [31, 32, 35, 78]. Als eine Erhöhung der ALT-Werte wird in den Teriflunomid-Studien ein Wert, der das zweifache des ULN (= upper limit of normal) übersteigt, definiert, wobei der Normbereich als ein Wert zwischen 6 und 43 IU/L angesehen wird. Eine genaue Definition findet sich in den Publikationen der Beta-Interferon-Studien nicht.</p>

^awie im Folgenden bei der Bewertung des Verzerrungspotentials näher erläutert, können die für 14 mg Teriflunomid als AESI eingestuft Ereignisse Gastrointestinale Beschwerden, Alopezie, Unverträglichkeiten und Neutropenie im indirekten Vergleich nicht berücksichtigt werden, weil die entsprechenden Angaben in den Rebif® 44 µg-Studienpublikationen fehlen. Das Risiko einer Neutropenie war in der direkt vergleichenden Studie TENERE für 14 mg Teriflunomid nicht signifikant unterschiedlich, aber tendenziell geringer als in der Rebif® 44 µg-Gruppe. Lymphopenie und Leukopenie sind im indirekten Vergleich anstelle der Neutropenie berücksichtigt, um ein möglicherweise bestehendes erhöhtes Sicherheitsrisiko in Bezug auf das blutbildende System, das im direkten Vergleich nicht erfasst wurde, zu berücksichtigen.

Tabelle 4-171: „Unerwünschte Ereignisse von besonderem Interesse“

Studie	Grippe-ähnliche Symptome	Reaktionen an der Injektionsstelle	Kopfschmerz	Erhöhung der ALT-Werte	Leukopenie	Lymphopenie
TEMPO	•	•	•	•	•	•
TOWER	•	•	•	•	•	•
TENERE	•	•	•	•	•	•
PRISMS	•	•	•	•	•	•

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-172: Bewertung des Verzerrungspotenzials für „Unerwünschte Ereignisse von besonderem Interesse“ in RCT für indirekte Vergleiche

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch / niedrig>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo						
TEMPO	niedrig	ja	ja	nein	nein	niedrig
TOWER	niedrig	ja	ja	nein	nein	niedrig
Studien: Interferon beta-1a 44 µg, s.c., dreimal pro Woche, versus Placebo						
PRISMS	niedrig	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Für die Bewertung des Verzerrungspotentials hinsichtlich des Endpunktes „Unerwünschte Ereignisse von besonderem Interesse“ in den Studien TEMPO und TOWER sei an dieser Stelle auf Modul 4.3.1.3.10 verwiesen.

Für die Studie PRISMS wird das Verzerrungspotential auf Studienebene ebenfalls als gering eingestuft. Die Endpunkterheber sind verblindet und das ITT-Prinzip adäquat umgesetzt. Es gibt keinen Hinweis auf ergebnisgesteuerte Berichterstattung oder sonstige das Verzerrungspotential erhöhende Faktoren. Das Verzerrungspotential wird für den Endpunkt „Unerwünschte Ereignisse von besonderem Interesse“ insgesamt als niedrig eingestuft.

An dieser Stelle soll bereits kritisch darauf hingewiesen werden, dass sich die Bewertung des Verzerrungspotentials, wie es für die einzelnen Studien erfolgt ist, lediglich auf die Qualität der in den jeweiligen Studien erhobenen Endpunkte und deren dort erfolgenden Auswertungen und Darstellungen bezieht. Die mögliche Verzerrung durch Auswahl einzelner Endpunkte im indirekten Vergleich ist damit noch nicht berücksichtigt. Im speziellen Fall der unerwünschten Ereignisse von besonderem Interesse schlägt diese Verzerrung durch eine selektive Auswahl besonders gravierend zu Buche. Dies ist der Tatsache geschuldet, dass unerwünschte Ereignisse, die ein besonderes Sicherheitsrisiko für 14 mg Teriflunomid darstellen, nicht aber für Rebif® 44 µg (gastrointestinale Beschwerden, verminderte Haardichte (HLT Alopecias)), in den Publikationen der placebokontrollierten Studien mit der zweckmäßigen Vergleichstherapie nicht berichtet sind. Ziel der Auswertung unerwünschter Ereignisse von besonderem Interesse ist deswegen primär eine Überprüfung der Verlässlichkeit der in TENERE erhobenen Werte, für eine Beurteilung des Gesamtbildes im Bereich Sicherheit ist die direkt vergleichende Studie als die beste Quelle anzusehen (s. Modul 4.3.1.3.10).

Für die Durchführung des indirekten Vergleiches gelten die folgenden Festlegungen:

Tabelle 4-173: Definitionen für den indirekten Vergleich

Analyse	Definition
Metaanalyse 1	Metaanalyse von TEMSO und TOWER
Indirekter Vergleich	Kombination von Metaanalyse 1 mit PRISMS (nach Bucher)
Metaanalyse 2	Metaanalyse von TENERE und dem indirekten Vergleich

Stellen Sie die Ergebnisse für den Endpunkt unerwünschte Ereignisse von besonderem Interesse für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

4.3.2.1.4.5.1 Unerwünschte Ereignisse von besonderem Interesse - Grippeähnliche Symptome

Grippeähnliche Symptome sind die häufigste Nebenwirkung der Therapie mit Rebif® 44 µg und treten bei ca. 70 % der Interferon-beta-Patienten auf [34]. Grippeähnliche Symptome sind kein unerwünschtes Ereignis, das sich typischerweise mit der Therapie mit 14 mg Teriflunomid in Verbindung bringen lässt, aus dem direkten Vergleich TENERE zeigt sich ein geringerer Schaden durch die Therapie mit 14 mg Teriflunomid im Vergleich zu seiner zweckmäßigen Vergleichstherapie Interferon beta-1a 44 µg, s. s. (Rebif®).

Tabelle 4-174: Ergebnisse für grippeähnliche Symptome aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation

Studie	Patientenzahl (Safety-Population ^a)	Anzahl Patienten mit dem entsprechenden Ereignis Prozentualer Anteil	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Grippeähnliche Symptome				
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo				
TEM SO	T14 (n=358)	8 2,2 %	0,894 [0,349; 2,291] p = 0,8152	0,891 [0,340; 2,337] p = 0,8152
	Pb (n=360)	9 2,5 %		
TOWER	T14 (n=371)	8 2,2 %	0,922 [0,360; 2,365] p = 0,8665	0,921 [0,351; 2,412] p = 0,8665
	Pb (n=385)	9 2,3 %		
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Interferon beta-1a 44 µg, s. c., dreimal pro Woche				
TENERE	T14 (n=110)	3 2,7 %	0,051 [0,017; 0,158] p < 0,0001	0,024 [0,007; 0,082] p < 0,0001
	Rb (n=101)	54 53,5 %		
Studien: Interferon beta-1a 44 µg, s. c., dreimal pro Woche, versus Placebo				
PRISMS	Rb (n = 184)	50 27,2 %	1,129 [0,798; 1,599] p = 0,4931	1,177 [0,738; 1,878] p = 0,4928
	Pb (n = 187)	45 24,1 %		

^as. Definition der Safety-Population für die Teriflunomid-Studien unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten), für PRISMS ist keine Abweichung der Safety-Population von der ITT-Population angegeben, so dass von der Annahme ausgegangen wird, dass zwischen den entsprechenden Populationen kein Unterschied besteht.

^bH₀: RR = 1 bzw. H₀: OR = 1

Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an die übliche Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.*

- *Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.*
- *Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere Widersprüche zwischen direkter und indirekter Evidenz.*

Grippeähnliche Symptome - Indirekter Vergleich

Tabelle 4-175: Ergebnisse für grippeähnliche Symptome aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse 1 (TEMSE & TOWER)	Teriflunomid 14 mg versus Placebo	0,908 [0,467; 1,767] p = 0,7763 I ² = 0%, p _Q = 0,9630	0,906 [0,458; 1,791] p = 0,7763 I ² = 0%, p _Q = 0,9629
Indirekter Vergleich (Metaanalyse 1 + PRISMS)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,804 [0,380; 1,704] p = 0,5693	0,769 [0,337; 1,757] p = 0,5339
Metaanalyse 2 (Indirekter Vergleich & TENERE)	Teriflunomid 14 mg versus Rebif [®] 44 µg	DL ^a : 0,210 [0,014; 3,122] p = 0,2568 I ² = 94 %, p _Q < 0,0001	DL ^a : 0,141 [0,005; 4,152] p = 0,2564 I ² = 95 %, p _Q < 0,0001

^aaufgrund der Heterogenität ist I² > 0, so dass die Metaanalyse mit Studie als zufälligem Effekt (nach DerSimonian und Laird (=DL)) andere Ergebnisse liefert als das Modell mit Studie als festem Effekt. Verwendet und interpretiert wurden in diesen Fällen ausschließlich die Ergebnisse nach DerSimonian und Laird

^bH₀: RR = 1 bzw. H₀: OR = 1

Metaanalyse 1

Sowohl in der TEMSO- als auch in der TOWER-Studie sind grippeähnliche Symptome relativ selten und in den Behandlungs- und Placebogruppen ungefähr gleich häufig und in sehr geringem Ausmaß aufgetreten. Die Metaanalyse der beiden Studien zeigt ein entsprechend homogenes Ergebnis (für relative Risiken I² = 0%, p_Q = 0,9630).

Indirekter Vergleich

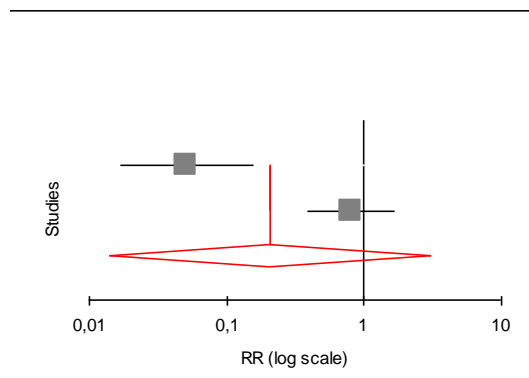
Im indirekten Vergleich konnte kein unterschiedliches Risiko für das Auftreten grippeähnlicher Symptome für die verschiedenen Interventionen 14 mg Teriflunomid und Rebif[®] 44 µg festgestellt werden.

Metaanalyse 2

Die Ergebnisse des indirekten Vergleichs und die Ergebnisse der direkt vergleichenden Studie TENERE unterscheiden sich hinsichtlich des AESI grippeähnliche Symptome erheblich. Dies zeigt sich in den entsprechenden Heterogenitätsmaßen I² = 94% und p_Q < 0,0001, es können

keine deutlichen, gleichgerichteten Effekte beobachtet werden. Der beträchtliche Zusatznutzen, der im direkten Vergleich gezeigt werden konnte, konnte in Bezug auf diesen Endpunkt im indirekten Vergleich nicht bestätigt werden. Die Inkonsistenz der Ergebnisse resultiert im Wesentlichen daraus, dass aus PRISMS keine nennenswerten Unterschiede zwischen Rebif® 44 µg und Placebo bezüglich der Raten von Patienten mit grippeähnlichen Symptomen berichtet sind.

Abbildung 4-42: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risk Ratio



Grippeähnliche Symptome - Fazit

Bezüglich der Häufigkeit von Grippe-ähnlichen Symptomen spiegeln die Ergebnisse des indirekten Vergleichs wieder, dass in der PRISMS und in der TENERE-Studie verschiedene Formulierungen von Rebif verwandt wurden (vgl. 4.3.2.1.2).

In der PRISMS-Studie waren auch unter Placebo Grippe-ähnliche Symptome häufig und die Unterschiede zwischen dem Placebo-Arm und dem Rebif® 44 µg-Arm gering. Offenbar war das Studiendesign der PRISMS-Studie mit Placeboinjektionen nicht geeignet, um diese für Interferon beta typischen unerwünschten Ereignisse nachzuweisen.

In der Fachinformation von Rebif® ist das Auftreten grippeähnlicher Symptome wie in Abschnitt 4.3.1.3.11.2 ausgeführt, mit bis zu 70 % angegeben [34], dies entspricht den Angaben in der zugehörigen Publikation für die neue Formulierung von Rebif® [86]. In der Rebif® New Formulation-Studie [86] wurde gezeigt, dass trotz allgemein besserer Verträglichkeit (insbesondere auf Reaktionen an der Injektionsstelle bezogen), grippeähnliche Symptome bei Anwendung der neuen Formulierung häufiger auftreten. Grippeähnliche Symptome gehören zu den häufigsten Gründen für mangelnde Therapieadhärenz der Patienten [20, 65], so dass trotz Verbesserungen in der Formulierung grippeähnliche Symptome ein noch ungelöstes Problem in der Interferon beta-Therapie darstellen.

Grippeähnliche Symptome sind bei Therapie mit 14 mg Teriflunomid ohne Relevanz. Dies wird in der direkt vergleichenden TENERE-Studie valide gezeigt.

4.3.2.1.4.5.2 Unerwünschte Ereignisse von besonderem Interesse – Reaktionen an der Injektionsstelle

Reaktionen an der Injektionsstelle gehören neben grippeähnlichen Symptomen zu den häufigsten Gründen für mangelnde Therapieadhärenz bei MS-Patienten [20, 65]. Reaktionen an der Injektionsstelle können per definitionem nur bei Wirkstoffen auftreten, die über eine Injektion verabreicht werden, nicht also bei dem zu bewertenden Arzneimittel Teriflunomid. Eine Berechnung relativer Risiken ist demgemäß wenig sinnvoll. Dennoch werden Reaktionen an der Injektionsstelle vom G-BA als valider Endpunkt anerkannt, über den der Vorteil einer oralen Darreichungsform dargestellt werden kann [10]. Dementsprechend werden an dieser Stelle die absoluten Ereignisraten in TEMSO und TOWER auf der einen Seite und PRISMS auf der anderen Seite vergleichend gegenübergestellt.

Tabelle 4-176: Ergebnisse für Reaktionen an der Injektionsstelle aus relevanten RCT – für die gesamte Studienpopulation

Studie	Patientenzahl (Safety-Population ^a)	Anzahl Patienten mit dem entsprechenden Ereignis (Prozentualer Anteil)
Reaktionen an der Injektionsstelle		
TEMSO	T14 (n=358)	1 (0,3 %)
	Pb (n=360)	1 (0,3 %)
TOWER	T14 (n=371)	0 (0 %)
	Pb (n=385)	0 (0%)
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Interferon beta-1a 44 µg, s. c., dreimal pro Woche		
TENERE	T14 (n=110)	0 (0 %)
	Rb (n=101)	22 (21,8 %)
Studien: Interferon beta-1a 44 µg, s. c., dreimal pro Woche, versus Placebo		
PRISMS	Rb (n = 184)	114 (62,0 % ^b)
	Pb (n = 187)	41 (21,9 %)

^as. Definition der Safety-Population für die Teriflunomid-Studien unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten), für PRISMS ist keine Abweichung der Safety-Population von der ITT-Population angegeben, so dass von der Annahme ausgegangen wird, dass zwischen den entsprechenden Populationen kein Unterschied besteht.

^bAngabe in der Publikation: Unterschied zu Placebo = 40, 95 % KI: [30,9; 49,2]; p<0,05

Wie erwartet, ist das Ereignis „Reaktionen an der Injektionsstelle“ bei der Behandlung mit Teriflunomid nicht aufgetreten (einzelne Ausnahmen müssen auf eine Begleitmedikation zurückzuführen sein), bei der Therapie mit der zweckmäßigen Vergleichstherapie jedoch sehr häufig. In der PRISMS-Studie sind entsprechende Ereignisse signifikant häufiger aufgetreten als unter Placebo. Dieses Ergebnis lässt sich jedoch nicht in einem indirekten Vergleich verwerten, da Reaktionen an der Injektionsstelle auch unter der Placebo-Injektion in der PRISMS-Studie sehr häufig aufgetreten sind. Diese Placebo-Injektion ist in Bezug auf den betrachteten Endpunkt nicht mit der Placebo-Filmtablette in den Studien TEMSO und TOWER gleichzusetzen und kann daher für den indirekten Vergleich nicht als Brückenkomparator dienen.

Die hohen Ereignisraten in der PRISMS-Studie unterstützen das Ergebnis der direkt vergleichenden TENERE-Studie. Die unterschiedlichen Ereignisraten in den Rebif 44 µg-Armen der beiden Studien können auch hier wahrscheinlich auf die verschiedenen Formulierungen von Rebif zurückzuführen sein (vgl. 4.3.2.1.2). Die niedrigere Rate an Reaktionen an der Injektionsstelle in der TENERE-Studie im Vergleich zur PRISMS-Studie ist in guter Übereinstimmung mit den Angaben aus der Rebif® New Formulation-Studie [86], die gezeigt hat, dass mit der neuen Formulierung gerade in Bezug auf diese Ereignisse eine verbesserte Verträglichkeit erreicht werden kann.

In Bezug auf Reaktionen an der Injektionsstelle hat Teriflunomid 14 mg einen beträchtlichen Vorteil gegenüber der zweckmäßigen Vergleichstherapie.

4.3.2.1.4.5.3 Unerwünschte Ereignisse von besonderem Interesse- Kopfschmerzen

Kopfschmerzen sind kein typisches Symptom der MS (s. Modul 3), aber es gibt Hinweise, dass Kopfschmerzen sowohl zu Beginn als auch im Verlauf der Erkrankung häufiger auftreten als in der nicht betroffenen Bevölkerung. Dabei sind Spannungskopfschmerzen und Migräne ohne Aura die häufigsten Formen, wobei auch Neuralgien und Clusterkopfschmerzen beschrieben worden sind (in diesen Fällen könnte auch die Lokalisation der MS-Läsionen eine maßgebliche Rolle spielen) [70]. Auch krankheitsmodifizierende Wirkstoffe wie die Beta-Interferone, können Kopfschmerzen auslösen oder verstärken [88]. Mit Hinblick auf das Ziel, den Patienten eine bessere Compliance zu ermöglichen, ist die Einschätzung des Kopfschmerz-auslösenden Potentials einer Dauertherapie ein wesentlicher Bestandteil der Bewertung.

Tabelle 4-177: Ergebnisse für Kopfschmerzen aus RCT des indirekten und direkten Vergleichs– für die gesamte Studienpopulation

Studie	Patientenzahl (Safety- Population ^a)	Anzahl Patienten mit dem entsprechenden Ereignis Prozentualer Anteil	Relatives Risiko 95 % KI [] p-Wert ^b	Odds Ratio 95 % KI [] p-Wert ^b
Kopfschmerzen				
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo				

TEMSO	T14 (n=358)	74 20,7 %	1,006 [0,755; 1,340] p = 0,9697	1,007 [0,701; 1,446] p = 0,9697
	Pb (n=360)	74 20,6 %		
TOWER	T14 (n=371)	54 14,6 %	1,057 [0,744; 1,502] p = 0,7557	1,067 [0,709; 1,606] p = 0,7557
	Pb (n=385)	53 13,8 %		
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Interferon beta-1a 44 µg, s. c., dreimal pro Woche				
TENERE	T14 (n=110)	18 16,4 %	0,570 [0,338; 0,961] p = 0,0349	0,486 [0,250; 0,944] p = 0,0331
	Rb (n=101)	29 28,7 %		
Studien: Interferon beta-1a 44 µg, s. c., dreimal pro Woche, versus Placebo				
PRISMS	Rb (n = 184)	83 45,1 %	1,029 [0,819; 1,291] p = 0,8073	1,052 [0,699; 1,585] p = 0,8073
	Pb (n = 187)	82 43,9 %		

^as. Definition der Safety-Population für die Teriflunomid-Studien unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten), für PRISMS ist keine Abweichung der Safety-Population von der ITT-Population angegeben, so dass von der Annahme ausgegangen wird, dass zwischen den entsprechenden Populationen kein Unterschied besteht.

^bH₀: RR = 1 bzw. H₀: OR = 1

Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an die übliche Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.*
- *Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.*
- *Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere Widersprüche zwischen direkter und indirekter Evidenz.*

Kopfschmerzen – Indirekter Vergleich

Tabelle 4-178: Ergebnisse für Kopfschmerzen aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI [] p-Wert ^b	Odds Ratio 95 % KI [] p-Wert ^b
Metaanalyse 1 (TEMSO & TOWER)	Teriflunomid 14 mg versus Placebo	1,026 [0,822; 1,281] p = 0,8209 I ² =0%, p _Q =0,8284	1,033 [0,788; 1,354] p = 0,8146 I ² =0%, p _Q =0,8353
Indirekter Vergleich (Metaanalyse 1 + PRISMS)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,997 [0,726; 1,371] p = 0,9871	0,982 [0,601; 1,604] p = 0,9410
Metaanalyse 2 (Indirekter Vergleich & TENERE)	Teriflunomid 14 mg versus Rebif [®] 44 µg	DL ^a : 0,785 [0,456; 1,350] p = 0,3813 I ² =69%, p _Q =0,0729	DL ^a : 0,717 [0,361; 1,422] p = 0,3407 I ² =64%, p _Q =0,0950

^aaufgrund der Heterogenität ist I²>0, so dass die Metaanalyse mit Studie als zufälligem Effekt (nach DerSimonian und Laird (=DL)) andere Ergebnisse liefert als das Modell mit Studie als festem Effekt. Verwendet und interpretiert wurden in diesen Fällen ausschließlich die Ergebnisse nach DerSimonian und Laird

^bH₀: RR = 1 bzw. H₀: OR = 1

Metaanalyse 1

In der TEMSO-Studie und der TOWER-Studie sind die relativen Risiken für Kopfschmerzen sehr ähnlich. Dementsprechend ist auch die Metaanalyse auf der Basis der relativen Risiken von guter Homogenität (I²=0%, p_Q=0,8284).

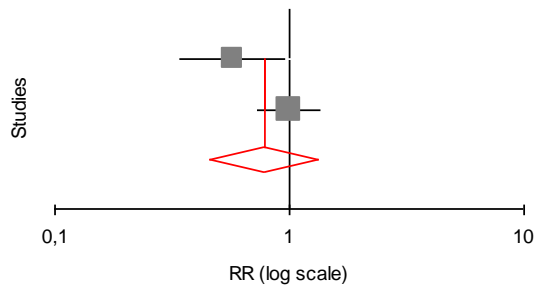
Indirekter Vergleich

Im indirekten Vergleich konnte in Bezug auf das unerwünschte Ereignis „Kopfschmerzen“ kein Unterschied zwischen den beiden Interventionen festgestellt werden (relatives Risiko 0,997; 95 % KI: [0,726; 1,371]).

Metaanalyse 2

Die Ergebnisse von TENERE und dem durchgeführten indirekten Vergleich passen nicht zusammen, es besteht deutliche Inkonsistenz zwischen den Ergebnissen (für relative Risiken: I²=69%, p_Q=0,0729), deutliche gleichgerichtete Effekte können nicht beobachtet werden.

Abbildung 4-43: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risk Ratio



Kopfschmerzen - Fazit

In keiner der Studien des indirekten Vergleichs konnten Unterschiede zwischen den Behandlungsarmen beobachtet werden. Dies steht im Widerspruch zu den Ergebnissen der TENERE-Studie, in der ein signifikanter Vorteil für 14 mg Teriflunomid gegenüber Rebif® 44 µg gezeigt werden konnte. Der signifikante Vorteil in der TENERE-Studie ist in guter Übereinstimmung mit den Angaben in der Fachinformation von Rebif® 44 µg [34], in der Kopfschmerzen als „sehr häufige“ unerwünschte Ereignisse aufgeführt werden, in der PRISMS-Studie sind die Ereignisraten unter Placebo sehr hoch, wodurch im indirekten Vergleich kein Vorteil gezeigt werden konnte.

4.3.2.1.4.5.4 Unerwünschte Ereignisse von besonderem Interesse- Erhöhte ALT-Werte

Tabelle 4-179: Ergebnisse für aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation

Studie	Patientenzahl (Safety- Population ^a)	Anzahl Patienten mit dem entsprechenden Ereignis Prozentualer Anteil	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Erhöhte ALT-Werte				
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo				
TEMSO	T14 (n=358)	51 14,2 %	2,137 [1,346; 3,394] p = 0,0013	2,326 [1,398; 3,870] p = 0,0012
	Pb (n=360)	24 6,7 %		
TOWER	T14 (n=371)	52 14,0 %	1,686 [1,112; 2,558] p = 0,0140	1,798 [1,129; 2,865] p = 0,0135
	Pb (n=385)	32 8,3 %		
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Interferon beta-1a 44 µg, s. c., dreimal pro Woche				
TENERE	T14 (n=110)	11 10,0 %	0,326 [0,173; 0,613] p = 0,0005	0,251 [0,118; 0,533] p = 0,0003
	Rb (n=101)	31 30,7 %		
Studien: Interferon beta-1a 44 µg, s. c., dreimal pro Woche versus Placebo				
PRISMS	Rb (n = 184)	12 6,5 %	6,098 [1,384; 26,87] p = 0,0169	6,454 [1,424; 29,25] p = 0,0156
	Pb (n = 187)	2 1,1 %		

^as. Definition der Safety-Population für die Teriflunomid-Studien unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten), für PRISMS ist keine Abweichung der Safety-Population von der ITT-Population angegeben, so dass von der Annahme ausgegangen wird, dass zwischen den entsprechenden Populationen kein Unterschied besteht.

^bH₀: RR = 1 bzw. H₀: OR = 1

Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an die übliche Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.*
- *Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.*
- *Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere Widersprüche zwischen direkter und indirekter Evidenz.*

Erhöhte ALT-Werte - Indirekter Vergleich

Tabelle 4-180: Ergebnisse für erhöhte ALT-Werte aus RCT mit dem zu bewertenden Arzneimittel- indirekter Vergleich – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse 1 (TEMZO & TOWER)	Teriflunomid 14 mg versus Placebo	1,875 [1,376; 2,555] p < 0,0001 I ² = 0%, p _Q =0,4560	2,022 [1,434; 2,851] p < 0,0001 I ² = 0%, p _Q =0,4650
Indirekter Vergleich (Metaanalyse 1 + PRISMS)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,308 [0,068; 1,399] p = 0,1271	0,313 [0,067; 1,476] p = 0,1422
Metaanalyse 2 (Indirekter Vergleich & TENERE)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,323 [0,180; 0,579] p = 0,0001 I ² = 0%, p _Q =0,9449	0,262 [0,133; 0,515] p = 0,0001 I ² = 0%, p _Q =0,8006

^bH₀: RR = 1 bzw. H₀: OR = 1

Metaanalyse 1

Die 14 mg Teriflunomid-Studien lassen sich ohne Hinweis auf Heterogenität metaanalytisch zusammenfassen. In Bezug auf die erhöhten ALT-Werte ist das Verzerrungspotential für alle Studien niedrig, Unterschiede in der Patientenpopulation haben keinen Einfluss auf die Leberwerte, da eine Erhöhung der Leberwerte mit der Intervention nicht aber mit der MS in Zusammenhang zu bringen ist. Die statistische Vergleichbarkeit zeigt sich in der guten Homogenität mit Werten von I²= 0% und p_Q=0,4560 für das relative Risiko als Abstandsmaß. Sowohl für 14 mg Teriflunomid als auch für Rebif[®] 44 µg zeigt sich ein erhöhtes Risiko für eine Erhöhung der ALT-Werte im Vergleich zu Placebo jedoch mit relativen Risiken von 1,875 (14 mg Teriflunomid) und 6,098 (Rebif[®] 44 µg) in deutlich unterschiedlichem Ausmaß.

Indirekter Vergleich

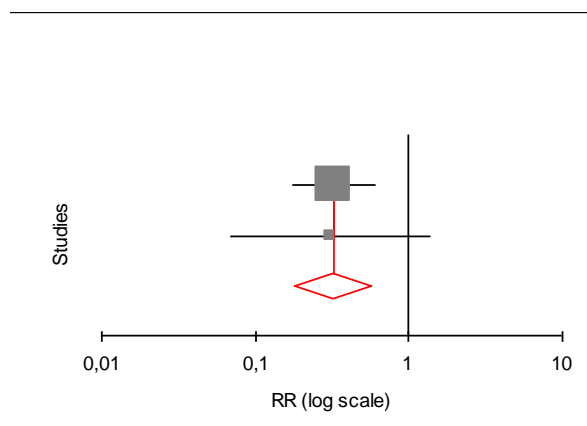
Im indirekten Vergleich zeigt sich in Bezug auf erhöhte ALT-Werte kein signifikanter Vorteil für 14 mg Teriflunomid im Vergleich zu der zweckmäßigen Vergleichstherapie Rebif[®] 44 µg (relatives Risiko 0,308; 95 % KI: [0,068; 1,399]), dies ist wahrscheinlich auf die geringe

Präzision des entsprechenden Schätzers in der PRISMS-Studie zurückzuführen. Der Wert von 0,308 zeigt jedoch eine klare Tendenz für einen Vorteil von 14 mg Teriflunomid gegenüber Rebif® 44 µg.

Metaanalyse 2

Der signifikante Vorteil für 14 mg Teriflunomid in Bezug auf erhöhte ALT-Werte im direkten Vergleich TENERE wird durch den indirekten Vergleich gestützt. Die Ergebnisse von direktem und indirektem Vergleich sind konsistent ($I^2=0\%$, $p_Q=0,9449$), mit einem relativen Risiko von 0,323 (95 % KI: [0,180; 0,579]) entspricht dies einer bedeutsamen Vermeidung nicht schwerwiegender Nebenwirkungen. Durch den Gewinn an Präzision ist das Ergebnis in der Metaanalyse 2 (wie bereits in TENERE allein, $p=0,0005$) statistisch signifikant ($p=0,0001$).

Abbildung 4-44: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risk Ratio



Erhöhte ALT-Werte - Fazit

Im Bezug auf eine Erhöhung der ALT-Werte konnte ein signifikanter Vorteil für 14 mg Teriflunomid im Vergleich zu Rebif® 44 µg gezeigt werden.

4.3.2.1.4.5.5 Unerwünschte Ereignisse von besonderem Interesse- Lymphopenie

Tabelle 4-181: Ergebnisse für Lymphopenie aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation

Studie	Patientenzahl (Safety- Population ^a)	Anzahl Patienten mit dem entsprechenden Ereignis Prozentualer Anteil	Relatives Risiko 95 % KI p-Wert ^c	Odds Ratio 95 % KI p-Wert ^c
Lymphopenie				
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo				
TEMSO	T14 (n=358)	3 0,8 %	1,508 [0,254; 8,973] p = 0,6514	1,504 ^b [0,259; 8,722] p = 0,6493
	Pb (n=360)	2 0,6 %		
TOWER	T14 (n=371)	1 0,3 %	3,113 ^d [0,127; 76,17] p = 0,4864	7,673 ^b [0,152; 387,0] p = 0,3083
	Pb (n=385)	0 0 %		
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Interferon beta-1a 44 µg, s. c., dreimal pro Woche				
TENERE	T14 (n=110)	0 0 %	0,306 ^d [0,013; 7,435] p = 0,4672	0,124 ^b [0,002; 6,261] p = 0,2967
	Rb (n=101)	1 1,0 %		
Studien: Interferon beta-1a 44 µg, s. c., dreimal pro Woche, versus Placebo				
PRISMS	Rb (n = 184)	24 13,0 %	3,485 [1,539; 7,888] p = 0,0027	3,857 [1,619; 9,192] p = 0,0023
	Pb (n = 187)	7 3,7 %		

^as. Definition der Safety-Population für die Teriflunomid-Studien unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten), für PRISMS ist keine Abweichung der Safety-Population von der ITT-Population angegeben, so dass von der Annahme ausgegangen wird, dass zwischen den entsprechenden Populationen kein Unterschied besteht.

^bAnalyse der Odds Ratios in TEMSO, TOWER und TENERE nach der Peto-Odds-Ratio-Methode

^cH₀: RR = 1 bzw. H₀: OR = 1

^dAddition von 0,5 zu jeder der absoluten Zellenhäufigkeiten in TOWER und TENERE

Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an die übliche Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- Homogenität der Ergebnisse: Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.
- Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.
- Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere Widersprüche zwischen direkter und indirekter Evidenz.

Lymphopenie - Indirekter Vergleich

Tabelle 4-182: Ergebnisse für Lymphopenie aus RCT mit dem zu bewertenden Arzneimittel-indirekter Vergleich – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^c	Odds Ratio 95 % KI p-Wert ^c
Metaanalyse 1 (TEMSO & TOWER)	Teriflunomid 14 mg versus Placebo	1,791 ^a [0,377; 8,502] p = 0,4632 I ² = 0%, p _Q = 0,6981	1,975 ^b [0,397; 9,824] p = 0,4056 I ² = 0%, p _Q = 0,4572
Indirekter Vergleich (Metaanalyse 1 + PRISMS)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,514 [0,089; 2,984] p = 0,4584	0,512 [0,083; 3,173] p = 0,4721
Metaanalyse 2 (Indirekter Vergleich & TENERE)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,456 ^a [0,098; 2,125] P = 0,3171 I ² = 0%, p _Q = 0,7805	0,398 [0,076; 2,080] p = 0,2748 I ² = 0%, p _Q = 0,5202

^aAddition von 0,5 zu jeder der absoluten Zellenhäufigkeiten in TOWER und TENERE

^bAnalyse der Odds Ratios in TEMSO, TOWER und TENERE nach der Peto-Odds-Ratio-Methode

^cH₀: RR = 1 bzw. H₀: OR = 1

Metaanalyse 1

Die absolute Ereignishäufigkeit in Bezug auf Lymphopenie war in den Teriflunomid-Studien insgesamt gering. Entsprechend wurden die Odds-Ratios in TEMSO, TOWER (und auch in TENERE) mit der Peto-Odds-Ratio-Methode analysiert [17].

Das Ergebnis der Metaanalyse auf Basis der relativen Risiken ergibt keinen signifikanten Unterschied zwischen der 14 mg Teriflunomid-Gruppe und der Placebo-Gruppe. Die Studien zeigen gute Homogenität (I² = 0%, p_Q = 0,6981).

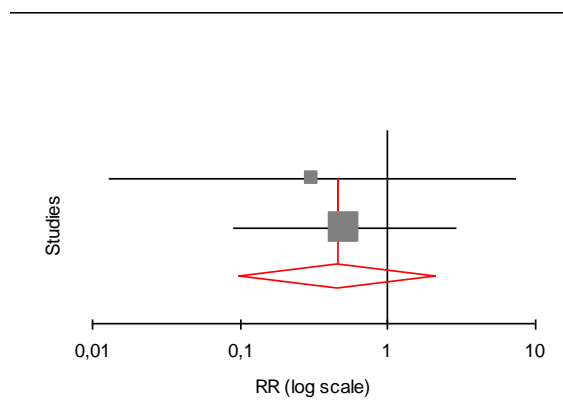
Indirekter Vergleich

Im indirekten Vergleich konnte mit einem relativen Risiko von 0,514 und einem zugehörigen 95 % KI von [0,089; 2,984] zwar eine Tendenz für einen Vorteil von 14 mg Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie, jedoch mit einem p-Wert von $p = 0,4584$ kein signifikanter Unterschied zwischen den Interventionen nachgewiesen werden.

Metaanalyse 2

Die Metaanalyse aus den Ergebnissen des direkten Vergleichs (TENERE) und dem indirekten Vergleich zeigt konsistente Ergebnisse ($I^2=0\%$, $p_Q=0,7805$).

Abbildung 4-45: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse 2) – Risk Ratio



Lymphopenie - Fazit

Lymphopenie ist in den Teriflunomid-Studien nur in Einzelfällen aufgetreten. Obwohl sie in der PRISMS-Studie signifikant häufiger aufgetreten ist als unter Placebo, konnte weder durch den direkten noch durch den indirekten Vergleich ein signifikanter Unterschied zwischen den Interventionen nachgewiesen werden.

4.3.2.1.4.5.6 Unerwünschte Ereignisse von besonderem Interesse- Leukopenie

Tabelle 4-183: Ergebnisse für Leukopenie aus RCT des indirekten und direkten Vergleichs – für die gesamte Studienpopulation

Studie	Patientenzahl (Safety-Population ^a)	Anzahl Patienten mit dem entsprechenden Ereignis Prozentualer Anteil	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Leukopenie				

Studien: Teriflunomid 14 mg, einmal pro Tag, versus Placebo				
TEMSO	T14 (n=358)	4 1,1 %	4,022 [0,452; 35,81] p = 0,2121	4,056 [0,451; 36,47] p = 0,2114
	Pb (n=360)	1 0,3 %		
TOWER	T14 (n=371)	5 1,3 %	5,189 [0,609; 44,20] p = 0,1320	5,246 [0,610; 45,12] p = 0,1311
	Pb (n=385)	1 0,3 %		
Studien: Teriflunomid 14 mg, einmal pro Tag, versus Interferon beta-1a 44 µg, s. c., dreimal pro Woche				
TENERE	T14 (n=110)	2 1,8 %	0,918 [0,132; 6,398] p = 0,9313	0,917 [0,127; 6,632] P = 0,9313
	Rb (n=101)	2 2,0 %		
Studien: Interferon beta-1a 44 µg, s. c., dreimal pro Woche, versus Placebo				
PRISMS	Rb (n = 184)	15 8,2 %	5,082 [1,496; 17,26] p = 0,0092	5,444 [1,549; 19,14] p = 0,0082
	Pb (n = 187)	3 1,6 %		

^as. Definition der Safety-Population für die Teriflunomid-Studien unter 4.3.1.3.9 Unerwünschte Ereignisse (Gesamtzahl, schwerwiegend, die zum Behandlungsabbruch führten), für PRISMS ist keine Abweichung der Safety-Population von der ITT-Population angegeben, so dass von der Annahme ausgegangen wird, dass zwischen den entsprechenden Populationen kein Unterschied besteht.

^bH₀: RR = 1 bzw. H₀: OR = 1

Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an die übliche Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.*
- *Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.*
- *Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere Widersprüche zwischen direkter und indirekter Evidenz.*

Indirekter Vergleich (Leukopenie)

Tabelle 4-184: Ergebnisse für Leukopenie aus RCT mit dem zu bewertenden Arzneimittel-indirekter Vergleich – für die gesamte Studienpopulation

Analyse	Komparatoren	Relatives Risiko 95 % KI p-Wert ^b	Odds Ratio 95 % KI p-Wert ^b
Metaanalyse 1 (TEMSO & TOWER)	Teriflunomid 14 mg versus Placebo	4,580 [0,992; 21,16] p = 0,0513 I ² =0%, p _Q = 0,8705	4,625 [0,995; 21,51] p = 0,0508 I ² =0%, p _Q = 0,8698
Indirekter Vergleich (Metaanalyse 1 + PRISMS)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,901 [0,127; 6,391] p = 0,9172	0,850 [0,117; 6,188] p = 0,8722
Metaanalyse 2 (Indirekter Vergleich & TENERE)	Teriflunomid 14 mg versus Rebif [®] 44 µg	0,910 [0,229; 3,612] p = 0,8931 I ² =0%, p _Q = 0,9895	0,883 [0,217; 3,585] p = 0,8614 I ² =0%, p _Q = 0,9577

^aaufgrund der Heterogenität ist I²>0, so dass die Metaanalyse mit Studie als zufälligem Effekt (nach DerSimonian und Laird (=DL)) andere Ergebnisse liefert als das Modell mit Studie als festem Effekt. Verwendet und interpretiert wurden in diesen Fällen ausschließlich die Ergebnisse nach DerSimonian und Laird

^bH₀: RR = 1 bzw. H₀: OR = 1

Metaanalyse 1

Leukopenie ist in der TEMSO-Studie und in der TOWER-Studie im 14 mg Teriflunomid-Arm häufiger aufgetreten als im Placebo-Arm. Mit relativen Risiken von 4,022 und 5,189 ist die Risikoerhöhung durch die Intervention in beiden Studien von einer ähnlichen Größenordnung. Dementsprechend zeigt sich in der Metaanalyse auch keine Heterogenität I²= 0%, p_Q=0,8705.

Indirekter Vergleich

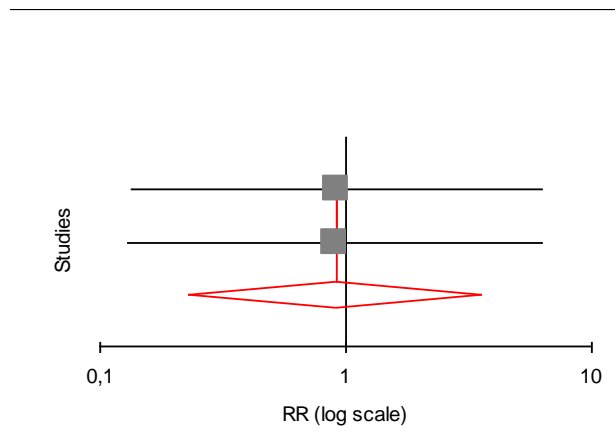
In der PRISMS-Studie war das Risiko einer Leukopenie der Rebif-Patienten im Vergleich zu Placebo auf das Fünffache erhöht (relatives Risiko von 5,082; 95 % KI: [1,496; 17,26], p=0,0092) – ähnlich wie das für Teriflunomid 14 mg in den Studien TEMSO und TOWER der Fall war.

Erwartungsgemäß ergibt sich im indirekten Vergleich kein Unterschied zwischen den Interventionen und mit einem relativen Risiko von 0,901 ein Wert nahe bei 1.

Metaanalyse 2

Das Ergebnis des indirekten Vergleichs ist mit dem Ergebnis des direkten Vergleichs konsistent (I²=0%, p_Q=0,9895), es konnte kein signifikanter Unterschied zwischen den Interventionen gezeigt werden.

Abbildung 4-46: Forest-Plot zur Metaanalyse 2 (von oben nach unten: TENERE, indirekter Vergleich, Ergebnis der Metaanalyse) – Risk Ratio



Leukopenie - Fazit

Bezüglich des Leukopenie-Risikos entsprechen die Ergebnisse des indirekten Vergleichs den Ergebnissen des direkten Vergleichs TENERE. Die verfügbaren Studien lassen in Bezug auf dieses unerwünschte Ereignis keinen Nachweis eines Unterschieds zwischen den Interventionen zu.

4.3.2.1.4.6 Zusammenfassung der Ergebnisse der indirekten Vergleiche

Insgesamt ergibt sich im indirekten Vergleich folgendes Bild:

- In Bezug auf die Wirksamkeitsendpunkte Schübe und Behinderungsprogression sind 14 mg Teriflunomid und Rebif® 44 µg vergleichbar wirksame Substanzen.

Die Präzision der Aussagen (im Sinne der Länge von Konfidenzintervallen) zur vergleichbaren Effektivität wird durch die indirekten Vergleiche deutlich gesteigert.

- In Bezug auf die Verträglichkeit zeigen sich Vorteile für 14 mg Teriflunomid. Dabei ist jedoch zu beachten, dass weder die grippeähnlichen Symptome noch spezielle Sicherheitsrisiken für 14 mg Teriflunomid berücksichtigt werden konnten. Für das Gesamtbild im Bereich der unerwünschten Ereignisse sei daher auf Modul 4.3.1 (direkte Vergleiche) verwiesen.

Im direkten Vergleich zeigten sich Vorteile für Teriflunomid in Bezug auf

- Patientenzufriedenheit
- Unerwünschte Ereignisse, die zum Therapieabbruch führten
- Mit der Leberfunktion assoziierte unerwünschte Ereignisse inklusive einer Erhöhung der ALT-Werte

- Reaktionen an der Injektionsstelle
- Grippeähnliche Symptome
- Kopfschmerzen.

Auch in indirekten Vergleichen unter Einbeziehung der randomisierten kontrollierten Studien TEMSO, TOWER und PRISMS zeigen sich die Vorteile von 14 mg Teriflunomid in Bezug auf Therapieabbrüche aufgrund von unerwünschten Ereignissen, erhöhte ALT-Werte und Reaktionen an der Injektionsstelle. Damit belegen die indirekten Vergleiche unterstützend die im direkten Vergleich bereits gezeigten Vorteile von Teriflunomid gegenüber Interferon beta-1a, 44 µg, s. c. (Rebif®).

Tabelle 4-185: Ergebnisse aus RCT mit dem zu bewertenden Arzneimittel- indirekte Vergleiche

Endpunkt	Abstandsmaß ^a Ergebnis ^c	Bewertung des Ergebnisses ^b
Mortalität		
Gesamt mortalität	Risikodifferenz 0,0061 95 % KI: [-0,0094; 0,0216] Kein nachweisbarer Unterschied zwischen den Interventionen	Geringe Ereignisrate; Vergleich nur auf Basis von Risikodifferenzen möglich, da in TEMSO und TENERE keine Todesfälle aufgetreten sind.
Morbidität (Effektivitätspunkte)		
Jährliche Schubrate (ARR)	Rate Ratio 0,961 95 % KI: [0,763; 1,211] Kein nachweisbarer Unterschied zwischen den Interventionen	Die Operationalisierung der Endpunkte ist vergleichbar. Es gibt keine wesentliche Heterogenität in den zugrunde liegenden Metaanalysen.
Zeit bis zur Behinderungsprogression	Hazard Ratio 1,122 95 % KI: [0,713; 1,766] Kein nachweisbarer Unterschied zwischen den Interventionen	Die Operationalisierung der Endpunkte ist vergleichbar. Es gibt keine wesentliche Heterogenität in den zugrunde liegenden Metaanalysen.
Unerwünschte Ereignisse (Verträglichkeitspunkte)		
Abbrüche wegen TEAE	Relatives Risiko 0,403 95 % KI: [0,079; 2,067] Trotz der Tendenz für einen Zusatznutzen besteht kein signifikanter Unterschied zwischen den Interventionen.	Die Operationalisierung der Endpunkte ist vergleichbar. Die Ergebnisse der placebokontrollierten Teriflunomid-Studien zeigen in Bezug auf diesen Endpunkt heterogene Ergebnisse, aber deutlich gleichgerichtete Effekte.
Grippeähnliche Symptome	Relatives Risiko 0,804 95 % KI: [0,380; 1,704] Kein nachweisbarer Unterschied zwischen den Interventionen.	Die Ergebnisse des indirekten Vergleichs sind aufgrund der hohen Ereignisrate im Placebo-Arm der PRISMS-Studie inkonsistent mit den Ergebnissen der TENERE-Studie.
Reaktionen an der	Indirekter Vergleich nicht möglich/sinnvoll; die absoluten Ereignisraten lassen jedoch	

Injektionsstelle	einen beträchtlichen Zusatznutzen in Bezug auf dieses Ereignis ableiten.	
Kopfschmerzen	Relatives Risiko 0,997 95 % KI: [0,726; 1,371] Kein nachweisbarer Unterschied zwischen den Interventionen.	Die Ergebnisse des indirekten Vergleichs (kein Unterschied zwischen den Interventionen) und dem direkten Vergleich TENERE (leichter Vorteil für 14 mg Teriflunomid) sind heterogen.
Erhöhung der ALT-Werte	Relatives Risiko 0,308 95 % KI: [0,068; 1,399] Tendenz für einen Zusatznutzen für 14 mg Teriflunomid. ^c	Genauere Definition einer Erhöhung der ALT-Werte bleibt für die Beta-Interferon-Studien unklar. Dennoch wird die Operationalisierung als hinreichend ähnlich angesehen.
Lymphopenie	Relatives Risiko 0,514 95 % KI: [0,089; 2,984] Trotz der Tendenz für einen Zusatznutzen besteht kein signifikanter Unterschied zwischen den Interventionen.	Geringe Ereignishäufigkeit in den Teriflunomid-Studien. Dadurch ist das Ergebnis wenig präzise.
Leukopenie	Relatives Risiko 0,901 95 % KI: [0,127; 6,391] Kein nachweisbarer Unterschied zwischen den Interventionen.	Geringe Ereignishäufigkeit in den Teriflunomid-Studien. Dadurch ist das Ergebnis wenig präzise.

^aIm Falle von Raten sind die Werte des indirekten Vergleichs basierend auf den relativen Risiken angegeben, auf Ebene der Odds Ratios führte der indirekte Vergleich zu keinen abweichenden Ergebnissen.

^bGrundlage für die Bewertung des Ergebnisses des indirekten Vergleichs sind die Verzerrungspotentiale der einzelnen Studien auf Studienebene, die Verzerrungspotentiale der einzelnen Endpunkte, die Vergleichbarkeit der Patientenpopulationen, die Homogenität der zugrunde liegenden Metaanalysen und Endpunkt-spezifische Punkte (seltene Ereignisse, nicht vergleichbare Placebo-Gruppen,...)

^cIm indirekten Vergleich konnte ein niedrigeres Risiko für die Erhöhung der ALT-Werte festgestellt werden. Das Ergebnis war jedoch aufgrund der geringen Präzision des entsprechenden Ergebnisses in der PRISMS-Studie nicht signifikant. Durch die Zusammenfassung mit der TENERE-Studie in der Metaanalyse 2 ergibt sich jedoch bei guter Konsistenz ein beträchtlicher Zusatznutzen, was auch dem Ergebnis von TENERE alleine entspricht.

4.3.2.1.4.7 Subgruppenanalysen – indirekte Vergleiche aus RCT

Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen auf Basis indirekter Vergleiche aus RCT. Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.12.

Für die indirekten Vergleiche wurden keine Subgruppenanalysen durchgeführt, da in der Publikation der PRISMS-Studie [78] keine den Auswertungen von TEMSO und TOWER entsprechenden Subgruppenanalysen dargestellt sind.

4.3.2.2 Nicht randomisierte vergleichende Studien

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn nicht randomisierte vergleichende Studien als Nachweis für einen Zusatznutzen herangezogen werden sollen.

4.3.2.2.1 Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien

Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung zu nicht randomisierten vergleichenden Studien. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen analog Abschnitt 4.3.1.1 zur Verfügung. Benennen Sie

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Literaturrecherche*
- *Studien aus der Suche in Studienregistern*
- *Resultierender Studienpool aus den einzelnen Suchschritten*

<< Angaben des pharmazeutischen Unternehmers >>

4.3.2.2.2 Charakteristika der nicht randomisierten vergleichenden Studien

Charakterisieren Sie nachfolgend die nicht randomisierten vergleichenden Studien. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2. und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.

Beschreiben Sie die Verzerrungsaspekte der nicht randomisierten vergleichenden Studie auf Studienebene mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Beschreibung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-186: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien

Studie	Zeitliche Parallelität der Gruppen	Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren	Verblindung		Hinweise auf ergebnisgesteuerte Berichterstattung	Andere Aspekte, die Verzerrungen verursachen können
			Patient	Behandler		
<Studie 1>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>

4.3.2.2.3 Ergebnisse aus nicht randomisierten vergleichenden Studien

4.3.2.2.3.1 <Endpunkt xxx> – nicht randomisierte vergleichende Studien

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-187: Operationalisierung von <Endpunkt xxx>

Studie	Operationalisierung
<Studie 1>	

Beschreiben Sie die Verzerrungsaspekte für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Bewertung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

Dokumentieren Sie die Einschätzung für jede Studie mit einem Bewertungsbogen in Anhang 4-G.

Tabelle 4-188: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien

Studie	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Ver- zerrungspotenzial be- einflussende Punkte
<Studie 1>	ja / nein / unklar	ja / nein / unklar	ja / nein / unklar	ja / nein

1) Stellen Sie die Ergebnisse der nicht randomisierten vergleichenden Studien gemäß den Anforderungen des TREND- bzw. des STROBE-Statements dar.

<< Angaben des pharmazeutischen Unternehmers >>

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus nicht randomisierten vergleichenden Studien fortlaufend in einem eigenen Abschnitt dar.

4.3.2.2.3.2 Subgruppenanalysen – nicht randomisierte vergleichende Studien

Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen aus nicht randomisierten vergleichenden Studien. Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.12.

<< Angaben des pharmazeutischen Unternehmers >>

4.3.2.3 Weitere Untersuchungen

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn über die in den Abschnitten 4.3.1, 4.3.2.1 und 4.3.2.2 genannten Studien hinausgehende Untersuchungen als Nachweis für einen Zusatznutzen herangezogen werden sollen.

4.3.2.3.1 Ergebnis der Informationsbeschaffung – weitere Untersuchungen

Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung nach Untersuchungen, die nicht in den Abschnitten 4.3.1, 4.3.2.1 und 4.3.2.2 aufgeführt sind.

Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen analog Abschnitt 4.3.1.1 zur Verfügung. Benennen Sie

- Studien des pharmazeutischen Unternehmers
- Studien aus der bibliografischen Literaturrecherche
- Studien aus der Suche in Studienregistern
- Resultierender Studienpool aus den einzelnen Suchschritten

<< Angaben des pharmazeutischen Unternehmers >>

4.3.2.3.2 Charakteristika der weiteren Untersuchungen

Charakterisieren Sie nachfolgend die weiteren Untersuchungen und bewerten Sie deren Verzerrungspotenzial.

Ergebnisse nicht randomisierter Studien, die keine kontrollierten Interventionsstudien sind, gelten aufgrund ihres Studiendesigns generell als potenziell hoch verzerrt. Trifft das auf die von Ihnen vorgelegten Studien nicht zu, begründen Sie Ihre Einschätzung.

Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2. und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.

<< Angaben des pharmazeutischen Unternehmers >>

4.3.2.3.3 Ergebnisse aus weiteren Untersuchungen

4.3.2.3.3.1 <Endpunkt xxx> – weitere Untersuchungen

Beschreiben Sie die Operationalisierung des Endpunkts für jede Studie in der folgenden Tabelle. Fügen Sie für jede Studie eine neue Zeile ein.

Tabelle 4-189: Operationalisierung von <Endpunkt xxx>

Studie	Operationalisierung
<Studie 1>	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt. Ergebnisse nichtrandomisierter Studien, die keine kontrollierten Interventionsstudien sind, gelten aufgrund ihres Studiendesigns generell als potenziell hoch verzerrt. Trifft das auf die von Ihnen vorgelegten Studien nicht zu, begründen Sie Ihre Einschätzung.

<< Angaben des pharmazeutischen Unternehmers >>

Stellen Sie die Ergebnisse der weiteren Untersuchungen gemäß den jeweils gültigen Standards für die Berichterstattung dar. Begründen Sie dabei die Auswahl des Standards für die Berichterstattung.

<< Angaben des pharmazeutischen Unternehmers >>

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus weiteren Untersuchungen fortlaufend in einem eigenen Abschnitt dar.

4.3.2.3.2 Subgruppenanalysen – weitere Untersuchungen

Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen aus weiteren Untersuchungen. Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.12.

<< Angaben des pharmazeutischen Unternehmers >>

4.3.2.4 Zusammenfassung der Ergebnisse aus weiteren Unterlagen

Der vorliegende Abschnitt soll einen Überblick über die Ergebnisse aus weiteren Unterlagen (Abschnitte 4.3.2.1, 4.3.2.2 und 4.3.2.3) geben. Die Zusammenfassung soll Aussagen zu allen in diesen Abschnitten präsentierten Endpunkten und Subgruppenanalysen enthalten. Dabei sollen, soweit verfügbar, numerische Ergebnisse aus Metaanalysen einschließlich Konfidenzintervallen dargestellt werden.

Fassen Sie die Ergebnisse aus weiteren Unterlagen zusammen.

Für die Nutzenbewertung von 14 mg Teriflunomid wurden keine Ergebnisse aus nicht randomisierten vergleichenden Studien und keine weiteren Untersuchungen herangezogen.

Als „weitere Unterlagen“ für die Nutzenbewertung des zu bewertenden Arzneimittels sind lediglich die indirekten Vergleiche von 14 mg Teriflunomid versus Rebif[®] 44 µg über Placebo als Brückenkompator zu rechnen. Die Ergebnisse dieser indirekten Vergleiche sind in Modul 4.3.2.1.4.6 zusammengefasst.

4.4 Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens

4.4.1 Beurteilung der Aussagekraft der Nachweise

Legen Sie für alle im Dossier eingereichten Unterlagen die Evidenzstufe dar. Beschreiben Sie zusammenfassend auf Basis der in den Abschnitten 4.3.1 und 4.3.2 präsentierten Ergebnisse die Aussagekraft der Nachweise für einen Zusatznutzen unter Berücksichtigung der Studienqualität, der Validität der herangezogenen Endpunkte sowie der Evidenzstufe.

4.4.1.1 Studienqualität

Zum Nachweis des Zusatznutzens von 14 mg Teriflunomid gegenüber Rebif® 44 µg sind entweder für den direkten oder die indirekten Vergleiche folgende Studien eingeflossen:

- TENERE (Nutzen: direkter Vergleich; Zusatznutzen: direkter Vergleich)
- TEMSO (Nutzen: direkter Vergleich; Zusatznutzen: indirekter Vergleich)
- TOWER (Nutzen: direkter Vergleich; Zusatznutzen: indirekter Vergleich)
- PRISMS (Zusatznutzen: indirekter Vergleich)

Die randomisierten, kontrollierten Zulassungsstudien TEMSO, TOWER und TENERE sind zur arzneimittelrechtlichen Zulassung konzipiert worden, um in dem Anwendungsgebiet krankheitsmodifizierende Therapie der Multiplen Sklerose eine maximale Aussagekraft bezüglich der Wirksamkeit und Verträglichkeit zu erreichen. Dies betrifft das allgemeine Design der Studien als auch die Auswahl der Endpunkte (s. 4.4.1.2)

Bei TEMSO und TOWER handelt es sich um doppelblinde Studien mit niedrigem Verzerrungspotential und hoher Qualität. Bei TENERE handelt es sich um eine einfachblinde Studie von guter Studienqualität. Details zum Design finden sich bei der Bewertung des Verzerrungspotentials (Modul 4.3.1.2.2). Bei der Endpunkterhebung wurden adäquate Maßnahmen getroffen, um das Verzerrungspotential niedrig zu halten. Die Aussagekraft der Studien TEMSO, TOWER und TENERE zum Nutzen und Zusatznutzen von 14 mg Teriflunomid zur Behandlung von erwachsenen Patienten mit RRMS wird als hoch eingestuft.

Auch die PRISMS-Studie ist als eine qualitativ hochwertige Studie mit hoher Aussagekraft einzustufen. Es handelt sich bei PRISMS um eine randomisierte, kontrollierte Studie, die doppelblind durchgeführt wurde. Das Studiendesign und die Auswahl der Endpunkte wird als für das Anwendungsgebiet relevant eingestuft (Details s. 4.3.2.1.2 bis 4.3.2.1.3).

Die Studienmethodik und eine mögliche Verzerrung der Ergebnisse sind in den Anhängen 4-F und 4-G im Detail dokumentiert.

In den Studien TENERE, TEMSO, TOWER, PRISMS erfüllten 90 - 100 % der Patienten die unter 4.2.2.2 vorab definierten Einschlusskriterien (erwachsene Patienten, RRMS, EDSS < 5,5), so dass die aus den Studiendaten abgeleiteten Aussagen geeignet sind, die Fragestellung zu beantworten.

Die vier eingeschlossenen Studien sind allesamt Studien mit parallelen Behandlungsgruppen.

Tabelle 4-190: Beurteilung der eingeschlossenen Studien

Studie	RCT	Verzerrungspotential auf Studienebene	Zur Beantwortung der Fragestellung passende Studienpopulation
TENERE	ja	niedrig	ja
TEMSO	ja	niedrig	ja
TOWER	ja	niedrig	ja
PRISMS	ja	niedrig	ja

Aus der Auswertung der Daten, die in den vorgelegten Unterlagen dargestellt sind, können daher Aussagen mit hoher Ergebnissicherheit abgeleitet werden.

4.4.1.2 Validität der Endpunkte

Die Auswahl der Endpunkte entspricht den Empfehlungen des CHMP [1, 2], der folgende primären und sekundären Endpunkte für die vorliegende Indikation (krankheitsmodifizierende Therapie der RRMS) als relevant einstuft:

- Primärer Endpunkt
 - Verhinderung oder Modifikation von Schüben oder
 - Verhinderung oder Verzögerung der Behinderungsprogression
- wesentlicher sekundärer Endpunkt
 - Verhinderung oder Modifikation von Schüben oder
 - Verhinderung oder Verzögerung der Behinderungsprogression
- weitere Endpunkte
 - aus bildgebenden Verfahren abgeleitete Parameter
 - die Behinderungsprogression ergänzende Endpunkte, z. B. Ergebnisse aus neurologischen Bewertungsskalen, Maße für die kognitive Einschränkung, Bewertungsskalen zur Erfassung der Fatigue oder ambulatorischer Index
 - Lebensqualität
 - Unerwünschte Ereignisse

Alle zur Nutzenbewertung herangezogenen Endpunkte sind damit für das vorliegende Krankheitsbild (RRMS) adäquat, die verwendeten Messinstrumente patientenberichteter Endpunkte (FIS, TSQM, SF-36 und EQ-5D) sind entsprechend validiert (s. Modul 4.2.5.2).

4.4.1.3 Evidenzstufe

Alle im vorliegenden Dossier eingeschlossenen Studien sind RCT mit einem niedrigen Verzerrungspotential. Gemäß der Klassifizierung der Unterlagen nach Verfahrensordnung des Gemeinsamen Bundesausschusses [11] entsprechen die vorgelegten Unterlagen der Evidenzstufe 1b (Randomisierte kontrollierte Studien).

Im Rahmen der Nutzenbewertung sind vorrangig „randomisierte, kontrollierte, klinische Studien, insbesondere direkte Vergleichsstudien mit anderen Arzneimitteln oder Behandlungsformen, zu berücksichtigen“ [11]. Vorrangig soll es sich um direkt vergleichende Studien mit der zweckmäßigen Vergleichstherapie handeln, so dass für den Nachweis des Zusatznutzens von Teriflunomid 14 mg gegenüber Rebif® 44 µg die TENERE-Studie die höchste Aussagekraft besitzt.

4.4.1.4 Fazit

Aus der Auswertung der Daten, die in den vorgelegten Unterlagen dargestellt sind und die sich auf die Studien TEMSO, TOWER, TENERE und PRISMS stützen, können Aussagen mit hoher Ergebnissicherheit und Aussagekraft abgeleitet werden.

4.4.2 Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß

Führen Sie die in den Abschnitten 4.3.1 und 4.3.1.3.13 beschriebenen Ergebnisse zum Zusatznutzen auf Ebene einzelner Endpunkte zusammen und leiten Sie ab, ob sich aus der Zusammenschau der Ergebnisse zu den einzelnen Endpunkten insgesamt ein Zusatznutzen des zu bewertenden Arzneimittels im Vergleich zur zweckmäßigen Vergleichstherapie ergibt. Liegt ein Zusatznutzen vor, beschreiben Sie, worin der Zusatznutzen besteht.

Stellen Sie die Wahrscheinlichkeit des Zusatznutzens dar, d. h., beschreiben und begründen Sie unter Berücksichtigung der in Abschnitt 4.4.1 dargelegten Aussagekraft der Nachweise die Ergebnissicherheit der Aussage zum Zusatznutzen.

Beschreiben Sie außerdem das Ausmaß des Zusatznutzens unter Verwendung folgender Kategorisierung (in der Definition gemäß AM-NutzenV):

- *erheblicher Zusatznutzen*
- *beträchtlicher Zusatznutzen*
- *geringer Zusatznutzen*
- *nicht quantifizierbarer Zusatznutzen*
- *kein Zusatznutzen belegbar*

- *der Nutzen des zu bewertenden Arzneimittels ist geringer als der Nutzen der zweckmäßigen Vergleichstherapie*

Berücksichtigen Sie bei den Aussagen zum Zusatznutzen ggf. nachgewiesene Unterschiede zwischen verschiedenen Patientengruppen.

Der neue Wirkstoff Teriflunomid (14 mg) hat als orale Behandlungsoption einen beträchtlichen Zusatznutzen in der krankheitsmodifizierenden Therapie der schubförmig-remittierenden Multiplen Sklerose. Dieser beträchtliche Zusatznutzen wird in den klinischen Studien abgebildet in statistisch signifikanten Ergebnissen für

- weniger unerwünschte Ereignisse, die zu einem Therapieabbruch führten,
- Vermeidung spezifischer unerwünschter Ereignisse und
- eine größeren Patientenzufriedenheit (s. Abschnitt 4.1.6).

4.4.2.1 Beschreibung des Zusatznutzens auf Endpunktebene

Die Ergebnisse für die einzelnen Endpunkte und der resultierende Zusatznutzen, der beansprucht wird, sind in Tabelle 4-191 und in den folgenden Abschnitten 4.4.2.1.1 bis 4.4.2.1.5 ausführlich dargestellt.

Tabelle 4-191: Einzelergebnisse aus denen sich der Zusatznutzen ableitet (aufgeschlüsselt nach Endpunkten)- direkter Vergleich (TENERE)

Endpunkt	Effektschätzer (TENERE) 95 % KI [] p-Wert	Beanspruchter Zusatznutzen	Beleglage
Mortalität			
Gesamt mortalität		Kein Zusatznutzen	-
Morbidität			
Schübe (jährliche Schubrate)	RR 1,197 [0,623; 2,299] p = 0,5896	Kein Zusatznutzen	-
Behinderungs- progression (Zeit bis zur Behinderungsprogressio n)	HR 0,982 [0,399; 2,420] p = 0,9690		

Endpunkt	Effektschätzer (TENERE) 95 % KI [] p-Wert	Beanspruchter Zusatznutzen	Beleglage
Therapieversagen (Time to Failure)	HR 0,861 [0,564; 1,314] p = 0,5953		
Fatigue	Mittelwertsdifferenz -5,003 [-12,31; 2,307] p = 0,1789		
Lebensqualität (SF-36 oder EQ-5D)			
Lebensqualität	entfällt, da in TENERE die entsprechende Erhebung fehlt	Kein Zusatznutzen	-
Patientenzufriedenheit (TSQM)			
Allgemeine Zufriedenheit	Mittelwertsdifferenz 7,843 [1,458; 14,23] p = 0,0162	Beträchtlicher Zusatznutzen	Hinweis
Zufriedenheit in Bezug auf die Wirksamkeit	Mittelwertsdifferenz 3,835 [-3,082; 10,75] p = 0,2759		
Zufriedenheit in Bezug auf die Verträglichkeit	Mittelwertsdifferenz 21,77 [16,48; 27,07] p < 0,0001		
Zufriedenheit in Bezug auf die Anwender- freundlichkeit	Mittelwertsdifferenz 27,96 [23,35; 32,56] p < 0,0001		
Unerwünschte Ereignisse			
Gesamtrate unerwünschter Ereignisse	RR 0,966 [0,904; 1,031] p = 0,2946	Kein Zusatznutzen	-

Endpunkt	Effektschätzer (TENERE) 95 % KI [] p-Wert	Beanspruchter Zusatznutzen	Beleglage
Gesamtrate schwerwiegender unerwünschter Ereignisse	RR 0,787 [0,274; 2,264] p = 0,6568	Kein Zusatznutzen	-
Gesamtrate unerwünschter Ereignisse, die zu einem Behandlungsabbruch führten	RR 0,501 [0,262; 0,959] p = 0,0369	Geringer Zusatznutzen	Beleg (Ergebnisse aus TENERE und indirekt aus TEMSO, TOWER und PRISMS)
Unerwünschte Ereignisse von besonderem Interesse			
Reaktionen an der Injektionsstelle	RR nicht berechenbar, da das Ereignis unter 14 mg Teriflunomid nicht auftritt	Beträchtlicher Zusatznutzen	Beleg (Ergebnisse aus TENERE und indirekt aus TEMSO, TOWER und PRISMS)
Grippeähnliche Symptome	RR 0,051 [0,017; 0,158] p < 0,0001		
Verminderte Haardichte (HLT Alopecias)	RR 20,2 [2,773; 147,1] p = 0,0030		
Mit der Leberfunktion assoziierte unerwünschte Ereignisse	RR 0,25 [0,13; 0,45] p < 0,0001		
Erhöhung der ALT-Werte	RR 0,326 [0,173; 0,613] p = 0,0005		
Durchfall	RR 2,640 [1,237; 5,632] p = 0,0121		
Übelkeit und Erbrechen	RR 2,295 [0,743; 7,089] p = 0,1487		

Endpunkt	Effektschätzer (TENERE) 95 % KI [] p-Wert	Beanspruchter Zusatznutzen	Beleglage
Neutropenie	RR 0,612 [0,178; 2,107] p = 0,4364		
Infektionen und Infestationen	RR 1,055 [0,795; 1,400] p = 0,7108		
Bluthochdruck	RR 1,148 [0,317; 4,156] p = 0,8338		
Periphere Neuropathie	RR 4,591 [0,546; 38,63] p = 0,1608		
Kopfschmerzen	RR 0,570 [0,338; 0,961] p = 0,0349		

In Tabelle 4-191 sind die Studienergebnisse der einzelnen Endpunkte für die TENERE-Studie dargestellt, die wie in 4.4.1.3 ausgeführt, die höchste Aussagekraft für die Bewertung des Zusatznutzens besitzt. Für die einzelnen Ergebnisse der placebokontrollierten Studien TEMSO und TOWER und der indirekten Vergleiche sei auf die entsprechenden Kapitel oder die Zusammenfassungen in 4.3.1.3.13 und 4.3.2.1.4.6 verwiesen.

Die indirekten Vergleiche (14 mg Teriflunomid aus TEMSO/TOWER vs. Interferon beta-1a, s.c, 44 µg, 3 x pro Woche aus PRISMS) führen zu keinen von den Ergebnissen des direkten Vergleichs (TENERE) abweichenden Resultaten. Für die wichtigsten Wirksamkeitsendpunkte, die jährliche Schubrate und die Zeit bis zur Behinderungsprogression, wird die Präzision der Aussage (kein signifikanter Unterschied zwischen den beiden Interventionen) deutlich erhöht, die Vorteile bezüglich einzelner unerwünschter Ereignisse von besonderem Interesse und der geringeren Therapieabbruchrate aufgrund von unerwünschten Ereignissen konnten durch die indirekten Vergleiche belegt werden.

4.4.2.1.1 Mortalität

Im Bereich Mortalität konnte für keine der Interventionen ein patientenrelevanter Vorteil oder ein größeres Risiko festgestellt werden.

4.4.2.1.2 Morbidität

Aus der vorhanden Evidenz ergibt sich eine vergleichbare Effektivität des zu bewertenden Arzneimittels 14 mg Teriflunomid (AUBAGIO®) und der zweckmäßigen Vergleichstherapie Interferon beta-1a (Rebif® 44 µg). In einer direkt vergleichenden Studie (TENERE) konnte kein Wirksamkeitsunterschied zwischen den beiden Interventionen nachgewiesen werden. Die vergleichbare Effektivität leitet sich ebenfalls aus indirekten Vergleichen mit Placebo als Brückenkomparator ab.

In Bezug auf die Effektivität wird für das zu bewertende Arzneimittel kein Zusatznutzen beansprucht.

4.4.2.1.3 Lebensqualität

Für den Zusatznutzen von Teriflunomid 14 mg im Bereich Lebensqualität gibt es keinen Beleg, weil sowohl in der direkt vergleichenden TENERE-Studie als auch in der PRISMS-Studie entsprechende Erhebungen fehlen.

In den placebokontrollierten Studien TEMSO und TOWER konnte für Teriflunomid 14 mg kein Zusatznutzen im Bereich Lebensqualität gezeigt werden. Dies ist jedoch kein Hinweis darauf, dass das zu bewertende Arzneimittel gegenüber der zweckmäßigen Vergleichstherapie keinen Zusatznutzen im Bereich Lebensqualität hat. Da die Lebensqualität der Patienten durch die Reaktionen an der Injektionsstelle und die grippeähnlichen Symptome, die unter Interferon beta-1a-Therapie auftreten, erheblich eingeschränkt ist, wird ein Zusatznutzen für Teriflunomid im Bereich Lebensqualität angenommen. Dieser kann jedoch mit den vorliegenden Studiendaten nicht belegt werden.

4.4.2.1.4 Patientenzufriedenheit

In Bezug auf die Patientenzufriedenheit ergibt sich ein beträchtlicher Zusatznutzen. Für MS-Patienten, die einer Dauertherapie mit krankheitsmodifizierenden Wirkstoffen zum Teil über Jahrzehnte ausgesetzt sind, ist eine bessere Verträglichkeit und Handhabung ihres angewendeten Arzneimittels von zentraler Bedeutung. Auch aus Sicht der Therapeuten ist es wichtig, dass Compliance und Therapieadhärenz sichergestellt sind, da die Wirkstoffe nur so effektiv greifen und die Therapieziele erreicht werden können. Diese Compliance und Therapieadhärenz sind auf Dauer nur mit Wirkstoffen zu erreichen, die von dem Patienten selbst als gut verträglich empfunden werden. Im Anwendungsgebiet A (krankheitsmodifizierende Therapie der schubförmig-remittierenden Multiplen Sklerose) ist die größere Zufriedenheit der Patienten resultierend aus einer besseren Anwenderfreundlichkeit und der Vermeidung von unerwünschten Ereignissen ausgesprochen bedeutsam. Damit liegt eine „bedeutsame Vermeidung anderer [d. h. nicht schwerwiegender] Nebenwirkungen“ [9] vor. Nach §5 Absatz 7 Satz 2 AM-NutzenV [9] resultiert daraus ein beträchtlicher Zusatznutzen.

Dieser wird als Hinweis eingestuft, weil entsprechende Daten lediglich aus der TENERE-Studie vorliegen

4.4.2.1.5 Unerwünschte Ereignisse

Die Behandlung mit Teriflunomid führte zu signifikant weniger unerwünschten Ereignissen, die zu einem Therapieabbruch führten, als die Behandlung mit der zweckmäßigen Vergleichstherapie Rebif® 44 µg. Dies zeigt einen beträchtlichen Zusatznutzen des neuen Wirkstoffes in der Therapie von erwachsenen Patienten mit RRMS. Der Unterschied in Bezug auf diesen Endpunkt war statistisch signifikant. Im indirekten Vergleich ergab sich ein Effektschätzer (relatives Risiko) zugunsten von Teriflunomid von ähnlicher Größe, der jedoch nicht signifikant war. Dementsprechend kann der Zusatznutzen in Bezug auf die unerwünschten Ereignisse, die zu einem Therapieabbruch führten, aus der vorhandenen Evidenz lediglich als gering eingestuft werden.

Im Anwendungsgebiet von Teriflunomid (AUBAGIO®) ist die Vermeidung von unerwünschten Ereignissen nicht schwerwiegenden Ausmaßes von großer Bedeutung für den Therapieerfolg, da es sich sowohl bei Teriflunomid als auch bei der zweckmäßigen Vergleichstherapie um lebenslang und dauerhaft anzuwendende Wirkstoffe handelt. Aus einer bedeutsamen Vermeidung nicht-schwerwiegender unerwünschter Ereignisse resultiert nach §5 Absatz 7 Satz 2 AM-NutzenV [9] ein beträchtlicher Zusatznutzen.

Die unerwünschten Ereignisse, die unter 14 mg Teriflunomid häufiger auftraten als unter Rebif® 44 µg –Therapie, waren von milder oder moderater Natur und eine auftretende Verminderung der Haardichte bildet sich in der Regel auch unter Behandlung mit Teriflunomid auch ohne Therapie wieder zurück. Dementsprechend führt neben der oralen Darreichungsform auch die langfristige Verträglichkeit von Teriflunomid über die Jahre zu einer für die Patienten bedeutend angenehmeren Anwendung als die Behandlung mit der zweckmäßigen Vergleichstherapie, die nicht nur mit den Injektionen selbst, sondern auch mit Reaktionen an der Injektionsstelle und grippeähnlichen Symptomen verbunden ist.

Der patientenrelevante Zusatznutzen bezüglich der unerwünschten Ereignisse ist durch die zusätzlichen indirekten Vergleiche belegt.

Als Resultat wird in Bezug auf die bedeutsame Vermeidung unerwünschter Ereignisse ein Beleg für einen beträchtlichen Zusatznutzen beansprucht.

4.4.2.2 Ausmaß des Zusatznutzens - Fazit

Aus den oben zusammengefassten Ergebnissen ergibt sich ein **Beleg für einen beträchtlichen Zusatznutzen** für 14 mg Teriflunomid gegenüber der zweckmäßigen Vergleichstherapie Interferon beta-1a, s. c., 44 µg (Rebif®).

4.4.3 Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht

Geben Sie auf Basis der in den Abschnitten 4.3.1 und 4.3.1.3.13 beschriebenen Ergebnisse und unter Berücksichtigung des in Abschnitt 4.4.2 dargelegten Zusatznutzens sowie dessen Wahrscheinlichkeit und Ausmaß in der nachfolgenden Tabelle an, für welche Patientengruppen ein therapeutisch bedeutsamer Zusatznutzen besteht. Benennen Sie das

Ausmaß des Zusatznutzens in Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen. Fügen Sie für jede Patientengruppe mit therapeutisch bedeutsamem Zusatznutzen eine neue Zeile ein.

Tabelle 4-192: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens

Bezeichnung der Patientengruppen	Ausmaß des Zusatznutzens
Erwachsene Patienten mit schubförmig-remittierender MS (RRMS)	beträchtlich

4.4.4 Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht – Arzneimittel zur Behandlung eines seltenen Leidens

Für Arzneimittel, die zur Behandlung eines seltenen Leidens zugelassen sind, müssen keine Nachweise zum medizinischen Nutzen und medizinischen Zusatznutzen vorgelegt werden (Abschnitt 4.3), solange der Umsatz des Arzneimittels mit der gesetzlichen Krankenversicherung zu Apothekenverkaufspreisen einschließlich Umsatzsteuer in den letzten 12 Monaten einen Betrag von 50 Millionen Euro nicht übersteigt.

Für diese Arzneimittel sind Angaben zu Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, im vorliegenden Abschnitt einzureichen. Dabei soll zunächst das Ausmaß des Zusatznutzens (in der Definition gemäß AM-NutzenV) begründet beschrieben werden. Unter Berücksichtigung des Ausmaßes des Zusatznutzens sollen Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, benannt werden. Weitere Abschnitte von Modul 4 können dabei nach Bedarf bearbeitet werden.

Beschreiben Sie begründet das Ausmaß des Zusatznutzens unter Verwendung der Kategorisierung gemäß AM-NutzenV. Berücksichtigen Sie bei den Aussagen zum Zusatznutzen nachgewiesene Unterschiede zwischen verschiedenen Patientengruppen. Verweisen Sie, wo notwendig, auf weitere Abschnitte von Modul 4.

<< Angaben des pharmazeutischen Unternehmers >>

Geben Sie auf Basis des Ausmaßes des Zusatznutzens in der nachfolgenden Tabelle an, für welche Patientengruppen ein therapeutisch bedeutsamer Zusatznutzen besteht. Benennen Sie das Ausmaß des Zusatznutzens in Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen. Fügen Sie für jede Patientengruppe mit therapeutisch bedeutsamem Zusatznutzen eine neue Zeile ein.

Tabelle 4-193: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens – Arzneimittel zur Behandlung eines seltenen Leidens

Bezeichnung der Patientengruppen	Ausmaß des Zusatznutzens

4.5 Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte

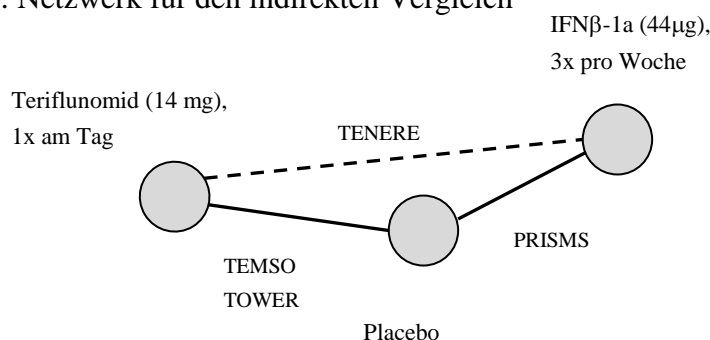
4.5.1 Begründung für die Vorlage indirekter Vergleiche

Sofern mit dem Dossier indirekte Vergleiche (Abschnitt 4.3.2.1) eingereicht wurden, begründen Sie dies. Begründen Sie dabei auch, warum sich die ausgewählten Studien jeweils für einen indirekten Vergleich gegenüber dem zu bewertenden Arzneimittel und damit für den Nachweis eines Zusatznutzens durch indirekten Vergleich eignen.

Mit der Studie TENERE ist ein direkter Vergleich des zu bewertenden Wirkstoffes Teriflunomid mit der zweckmäßigen Vergleichstherapie Interferon beta-1a, s. c. (Rebif[®]), vorhanden. Dieser direkte Vergleich hat das folgende Ergebnis: für Teriflunomid konnte keine bessere Wirksamkeit als für Interferon beta-1a, s. c. (Rebif[®]), gezeigt werden. In Bezug auf die Patientenzufriedenheit, Therapieabbrüche aufgrund von unerwünschten Ereignissen und relevante unerwünschte Ereignisse von besonderem Interesse (grippeähnliche Symptome, Reaktionen an der Injektionsstelle, Kopfschmerzen und mit der Leberfunktion assoziierte Ereignisse) zeigt Teriflunomid signifikante Vorteile. Damit hat Teriflunomid einen beträchtlichen Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie.

In TENERE war kein vorgeschalteter Nachweis der Nicht-Unterlegenheit bzgl. der Zeit bis zum Therapieabbruch geplant. In den sekundären Wirksamkeitsendpunkten gibt die TENERE-Studie einen Hinweis darauf, dass die Effektivität von Teriflunomid und Interferon beta-1a, 44 µg, s. c. (Rebif[®] 44 µg), vergleichbar ist, war aber zu klein angelegt, um die Abstände in diesen Wirksamkeitsendpunkten mit hoher Präzision zu schätzen. Damit lässt die Studie TENERE keine für einen Beleg ausreichenden Aussagen über den Zusatznutzen zu und das Ergebnis wurde durch einen indirekten Vergleich unterstützt.

Abbildung 4-47: Netzwerk für den indirekten Vergleich



Der Brückenkomparator für den indirekten Vergleich nach Bucher war Placebo. Es wurden dementsprechend Studien eingeschlossen, die das zu bewertende Arzneimittel 14 mg Teriflunomid gegen Placebo verglichen (TEMZO, TOWER) und des Weiteren Studien, die die zweckmäßige Vergleichstherapie Rebif® 44 µg gegen Placebo verglichen (PRISMS). Die Konsistenz des Ergebnisses des indirekten Vergleichs wurde mit Hilfe der Daten aus der direkt vergleichenden Studie TENERE geprüft.

Eignung der Studien für den indirekten Vergleich:

- TEMZO und TOWER:

Die beiden Studien sind Zulassungsstudien des zu bewertenden Arzneimittels, die Patientenpopulation besteht aus erwachsenen RMS-Patienten mit einem EDSS $\leq 5,5$. Die Teilkollektive der RRMS-Patienten (>90 % der Studienpopulationen) entsprechen der Zielpopulation von 14 mg Teriflunomid. Die jährliche Schubrate und die Zeit bis zur Behinderungsprogression sind jeweils der primäre und der wesentliche sekundäre Endpunkt, womit die zwei wesentlichen Wirksamkeitsendpunkte für die Beurteilung krankheitsmodifizierender Wirkstoffe in der MS-Therapie berücksichtigt sind.

- PRISMS:

Die Patientenpopulation besteht aus erwachsenen RRMS-Patienten mit einem EDSS von 0 - 5 und entspricht damit ebenfalls der Zielpopulation von 14 mg Teriflunomid. Die Anzahl der Schübe ist der primäre Endpunkt der Studie (die jährliche Schubrate ist daraus ableitbar) und die Zeit bis zur Behinderungsprogression ein sekundärer Endpunkt. Ein Vergleich mit den Ergebnissen der Teriflunomid-Studien ist damit prinzipiell möglich.

4.5.2 Begründung für die Vorlage nichtrandomisierter vergleichender Studien und weiterer Untersuchungen

Sofern mit dem Dossier nichtrandomisierte vergleichende Studien (Abschnitt 4.3.2.2) oder weitere Untersuchungen (Abschnitt 4.3.2.3) eingereicht wurden, nennen Sie die Gründe, nach denen es unmöglich oder unangemessen ist, zu den in diesen Studien bzw. Untersuchungen behandelten Fragestellungen Studien höchster Evidenzstufe (randomisierte klinische Studien) durchzuführen oder zu fordern.

Nichtrandomisierte vergleichende Studien und weitere Untersuchungen niedrigerer Evidenzklassen werden für den Nachweis des Nutzens und des Zusatznutzens von 14 mg Teriflunomid nicht herangezogen.

4.5.3 Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen

Falls aus Ihrer Sicht valide Daten zu patientenrelevanten Endpunkten zum Zeitpunkt der Bewertung noch nicht vorliegen können, begründen Sie dies.

Es liegen valide Daten für patientenrelevante Endpunkte vor.

4.5.4 Verwendung von Surrogatendpunkten

Die Verwendung von Surrogatendpunkten bedarf einer Begründung (siehe Abschnitt 4.5.3). Zusätzlich soll dargelegt werden, ob und warum die verwendeten Surrogatendpunkte im betrachteten Kontext valide Surrogatendpunkte darstellen bzw. Aussagen zu patientenrelevanten Endpunkten zulassen.

Eine Validierung von Surrogatendpunkten bedarf in der Regel einer Metaanalyse von Studien, in denen sowohl Effekte auf den Surrogatendpunkt als auch Effekte auf den interessierenden patientenrelevanten Endpunkt untersucht wurden (Burzykowski 2005¹³, Molenberghs 2010¹⁴). Diese Studien müssen bei Patientenkollektiven und Interventionen durchgeführt worden sein, die Aussagen für das dem vorliegenden Antrag zugrunde liegende Anwendungsgebiet und das zu bewertende Arzneimittel sowie die Vergleichstherapie erlauben.

Eine Möglichkeit der Verwendung von Surrogatendpunkten ohne abschließende Validierung stellt die Anwendung des Konzepts eines sogenannten Surrogate-Threshold-Effekts (STE) (Burzykowski 2006¹⁵) dar. Daneben besteht die Möglichkeit einer Surrogatvalidierung in der quantitativen Betrachtung geeigneter Korrelationsmaße von Surrogatendpunkt und interessierendem patientenrelevanten Endpunkt („individuelle Ebene“) sowie von Effekten auf den Surrogatendpunkt und Effekten auf den interessierenden patientenrelevanten Endpunkt („Studienebene“). Dabei ist dann zu zeigen, dass die unteren Grenzen der entsprechenden 95%-Konfidenzintervalle für solche Korrelationsmaße ausreichend hoch sind. Die Anwendung alternativer Methoden zur Surrogatvalidierung (siehe Weir 2006¹⁶) soll ausreichend begründet werden, insbesondere dann, wenn als Datengrundlage nur eine einzige Studie verwendet werden soll.

Berichten Sie zu den Studien zur Validierung oder zur Begründung für die Verwendung von Surrogatendpunkten mindestens folgende Informationen:

- Patientenpopulation
- Intervention
- Kontrolle
- Datenherkunft
- verwendete Methodik
- entsprechende Ergebnisse (abhängig von der Methode)

¹³ Burzykowski T (Ed.: *The evaluation of surrogate endpoints*. New York: Springer; 2005.

¹⁴ Molenberghs G, Burzykowski T, Alonso A, Assam P, Tilahun A, Buyse M: *A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials*. *Stat Methods Med Res* 2010; 19(3): 205-236.

¹⁵ Burzykowski T, Buyse M. *Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation*. *Pharm Stat* 2006; 5(3): 173-186.

¹⁶ Weir CJ, Walley RJ. *Statistical evaluation of biomarkers as surrogate endpoints: a literature review*. *Stat Med* 2006; 25(2): 183-203.

- Untersuchungen zur Robustheit
- ggf. Untersuchungen zur Übertragbarkeit

Sofern Sie im Dossier Ergebnisse zu Surrogatendpunkten eingereicht haben, benennen Sie die Gründe für die Verwendung von Surrogatendpunkten. Beschreiben Sie, ob und warum die verwendeten Surrogatendpunkte im betrachteten Kontext valide Surrogatendpunkte darstellen bzw. Aussagen zu patientenrelevanten Endpunkten zulassen.

Im vorliegenden Dossier werden mit der Darstellung der Krankheitslast und spezifischer Läsionen im MRT-Bild bei den Ergebnissen der TEMSO-Studie auch Surrogatparameter aufgeführt (Modul 4.3.1.3.5).

Die verschiedenen MRT-Parameter haben einen hohen Stellenwert in Phase II- Studien [1, 2] und in der Diagnostik der MS [15, 62]. Sie geben, wie oben ausgeführt, erste Hinweise auf die Wirksamkeit einer Therapie und es ist anzunehmen, dass die Bedeutung der MRT-Parameter in zukünftigen Studien zunehmen wird [63]. Daher ist sinnvoll, zusätzlich zu den klinischen Endpunkten Effekte auf Schübe und Behinderungsprogression auch die MRT-Parameter übersichtlich darzustellen. Hier konnte ein deutlicher Vorteil für Teriflunomid gezeigt werden. Dieses Ergebnis stützt den positiven Einfluss auf die klinischen Endpunkte, der für Teriflunomid gezeigt werden konnte (s. Kapitel 4.3.1.3.1 und 4.3.1.3.2).

4.6 Liste der eingeschlossenen Studien

Listen Sie alle für die Nutzenbewertung berücksichtigten Studien und Untersuchungen unter Angabe der im Dossier verwendeten Studienbezeichnung und der zugehörigen Datenquellen auf.

Tabelle 4-194: Liste der eingeschlossenen Studien und der zugehörigen Datenquellen

Studie	Genaue Benennung der Quelle	Kürzel
TEMSO	Studienbericht inklusive Studienprotokoll [32]	EFC6049 (TEMSO)
	Publikation der Studienergebnisse: O'Connor P., Wolinsky J. S., Confavreux C., Comi G., Kappos L., Olsson T. P., et al. Randomized trial of oral teriflunomide for relapsing multiple sclerosis. The New England journal of medicine; 365(14):1293-303. [35]	O'Connor 2011
	Publikation der Subgruppenanalysen: Miller A. E., O'Connor P., Wolinsky J. S., Confavreux C., Kappos L., Olsson T. P., et al. (2012)	Miller 2012

	<p>Pre-specified subgroup analyses of a placebo-controlled phase III trial (TEMPO) of oral teriflunomide in relapsing multiple sclerosis.</p> <p>Multiple sclerosis (Houndmills, Basingstoke, England); 18(11):1625-32 [43]</p>	
	<p>Publikation zusätzlicher Analysen:</p> <p>O'Connor P. W., Lublin F. D., Wolinsky J. S., Confavreux C., Comi G., Freedman M. S., et al. (2013):</p> <p>Teriflunomide reduces relapse-related neurological sequelae, hospitalizations and steroid use.</p> <p>Journal of neurology 2013 [41]</p>	O'Connor 2013
	<p>Publikation von MRT-Ergebnissen:</p> <p>Wolinsky J. S., Narayana P. A., Nelson F., Datta S., O'Connor P., Confavreux C., et al.</p> <p>Magnetic resonance imaging outcomes from a phase III trial of teriflunomide.</p> <p>Multiple sclerosis (Houndmills, Basingstoke, England) [42]</p>	Wolinsky 2013
TOWER	Studienbericht inklusive Studienprotokoll [31]	EFC10531 (TOWER)
TENERE	Studienbericht inklusive Studienprotokoll [22]	EFC10891 (TENERE)
PRISMS	<p>Publikation der Studienergebnisse:</p> <p>PRISMS (Prevention of Relapses and Disability by Interferon β-1a Subcutaneously in Multiple Sclerosis) Study Group</p> <p>Randomised double-blind placebo-controlled study of interferon β-1a in relapsing/remitting multiple sclerosis</p> <p>Lancet. 1998 Nov 7;352:1498-504 [78]</p>	Ebers et al.

4.7 Referenzliste

Benennen Sie nachfolgend alle Quellen (z. B. Publikationen, Studienberichte, Studienregister-einträge), die Sie in Modul 4 angegeben haben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard).

1. European Medicines Agency (2006): Guideline on clinical investigation of medicinal products for the treatment of multiple sclerosis. [Zugriff: 03/2012]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003485.pdf.
2. European Medicines Agency (2012): Guideline on clinical investigation of medicinal products for the treatment of Multiple Sclerosis/ Draft. [Zugriff: 10.10.2012]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/10/WC500133438.pdf.
3. Gemeinsamer Bundesausschuss (2013): Verfahrensordnung des Gemeinsamen Bundesausschusses. Bundesanzeiger Nr. 84a (Beilage) vom 10.06.2009, zuletzt geändert 18.04.2013, in Kraft getreten 25.07.2013;
4. Atkinson M. J., Kumar R., Cappelleri J. C., Hass S. L. (2005): Hierarchical construct validity of the treatment satisfaction questionnaire for medication (TSQM version II) among outpatient pharmacy consumers. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research; 8 Suppl 1:S9-S24.
5. Atkinson M. J., Sinha A., Hass S. L., Colman S. S., Kumar R. N., Brod M., et al. (2004): Validation of a general measure of treatment satisfaction, the Treatment Satisfaction Questionnaire for Medication (TSQM), using a national panel study of chronic disease. Health and quality of life outcomes; 2:12.
6. Bucher H. C., Guyatt G. H., Griffith L. E., Walter S. D. (1997): The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. Journal of clinical epidemiology; 50(6):683-91.
7. Fisk J. D., Pontefract A., Ritvo P. G., Archibald C. J., Murray T. J. (1994): The impact of fatigue on patients with multiple sclerosis. The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques; 21(1):9-14.
8. Fisk J. D., Ritvo P. G., Ross L., Haase D. A., Marrie T. J., Schlech W. F. (1994): Measuring the functional impact of fatigue: initial validation of the fatigue impact scale. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America; 18 Suppl 1:S79-83.
9. Bundesministerium für Gesundheit (2010): Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung-AM-Nutzen V). Bundesgesetzblatt; Teil 1 (68); 2324-8.
10. Gemeinsamer Bundesausschuss (2012): Niederschrift zum Beratungsgespräch gemäß § 8 Abs. 1 AM-NutzenV; Beratungsanforderung 2012-B-023.
11. Gemeinsamer Bundesausschuss (2012): Niederschrift zum Beratungsgespräch gemäß §8 AM-NutzenV; Beratungsanforderung 2012-B-023, Teriflunomid, zweckmäßige Vergleichstherapie.
12. Deutsches Cochrane-Zentrum Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften- Institut für Medizinisches Wissensmanagement, Ärztliches Zentrum für Qualität in der Medizin, (2013): „Manual Systematische

- Literaturrecherche für die Erstellung von Leitlinien“. [Zugriff: 18.06.2013]. URL: <http://www.cochrane.de/de/webliographie-litsuche>
<http://www.awmf.org/leitlinien/awmf-regelwerk/II-entwicklung.html>
<http://www.aezq.de/aezq/publikationen/kooperation> <http://www.freidok.uni-freiburg.de/volltexte/9020/>.
13. Poser C. M., Paty D. W., Scheinberg L., McDonald W. I., Davis F. A., Ebers G. C., et al. (1983): New diagnostic criteria for multiple sclerosis: guidelines for research protocols. *Annals of neurology*; 13(3):227-31.
 14. Polman C. H., Reingold S. C., Edan G., Filippi M., Hartung H. P., Kappos L., et al. (2005): Diagnostic criteria for multiple sclerosis: 2005 revisions to the "McDonald Criteria". *Annals of neurology*; 58(6):840-6.
 15. Polman C. H., Reingold S. C., Banwell B., Clanet M., Cohen J. A., Filippi M., et al. (2011): Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria. *Annals of neurology*; 69(2):292-302.
 16. McDonald W. I., Compston A., Edan G., Goodkin D., Hartung H. P., Lublin F. D., et al. (2001): Recommended diagnostic criteria for multiple sclerosis: guidelines from the International Panel on the diagnosis of multiple sclerosis. *Annals of neurology*; 50(1):121-7.
 17. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2011): Allgemeine Methoden. [Zugriff: 03/2012]. URL: https://www.iqwig.de/download/IQWiG_Methoden_Version_4_0.pdf.
 18. Treadaway K., Cutter G., Salter A., Lynch S., Simsarian J., Corboy J., et al. (2009): Factors that influence adherence with disease-modifying therapy in MS. *Journal of neurology*; 256(4):568-76.
 19. Steinberg S. C., Faris R. J., Chang C. F., Chan A., Tankersley M. A. (2010): Impact of adherence to interferons in the treatment of multiple sclerosis: a non-experimental, retrospective, cohort study. *Clinical drug investigation*; 30(2):89-100.
 20. Miller A. E., Rhoades R. W. (2012): Treatment of relapsing-remitting multiple sclerosis: current approaches and unmet needs. *Current opinion in neurology*; 25 Suppl:S4-S10.
 21. Cutter G. R., Baier M. L., Rudick R. A., Cookfair D. L., Fischer J. S., Petkau J., et al. (1999): Development of a multiple sclerosis functional composite as a clinical trial outcome measure. *Brain : a journal of neurology*; 122 (Pt 5):871-82.
 22. Sanofi-Aventis groupe (2011): Clinical Study Report; Study number: EFC10891, TENERE.
 23. Cordoba G., Schwartz L., Woloshin S., Bae H., Gotzsche P. C. (2010): Definition, reporting, and interpretation of composite outcomes in clinical trials: systematic review. *BMJ (Clinical research ed)*; 341:c3920.
 24. Schmidt R.M., Hoffmann F.A. (2012): Multiple Sklerose. München, Deutschland: Urban & Fischer (Elsevier).
 25. WHO (1946): Constitution of the World Health Organization. Stand: 2005 [Zugriff: 08.11.2012]. URL: <http://apps.who.int/gb/bd/PDF/bd47/EN/constitution-en.pdf>.
 26. Haupts M., Elias . G., Hardt C., Langenbahn H., Obert H., Pöhlau D., et al. (2003): Lebensqualität bei Patienten mit schubförmiger MS in Deutschland. *Nervenarzt*; (74):144-50.
 27. McHorney C. A., Ware J. E., Jr., Raczek A. E. (1993): The MOS 36-Item Short-Form Health Survey (SF-36): II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Medical care*; 31(3):247-63.

28. Freeman J. A., Hobart J. C., Langdon D. W., Thompson A. J. (2000): Clinical appropriateness: a key factor in outcome measure selection: the 36 item short form health survey in multiple sclerosis. *Journal of neurology, neurosurgery, and psychiatry*; 68(2):150-6.
29. Ware J. E., Jr., Sherbourne C. D. (1992): The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Medical care*; 30(6):473-83.
30. Fisk J. D., Brown M. G., Sketris I. S., Metz L. M., Murray T. J., Stadnyk K. J. (2005): A comparison of health utility measures for the evaluation of multiple sclerosis treatments. *Journal of neurology, neurosurgery, and psychiatry*; 76(1):58-63.
31. Sanofi-Aventis groupe (2012): Clinical Study Report; Study number: EFC10531, TOWER.
32. Sanofi-Aventis groupe (2011): Clinical Study Report; Study number: EFC6049, TEMSO.
33. Sanofi-Aventis groupe (2013): Fachinformation Aubagio.
34. Merck (2013): Rebif 44 Mikrogramm/0,5 ml Injektionslösung in einer Patrone; Fachinformation. Stand: 02/2013 [Zugriff: 05.08.2013]. URL: <http://www.fachinfo.de>.
35. O'Connor P., Wolinsky J. S., Confavreux C., Comi G., Kappos L., Olsson T. P., et al. (2011): Randomized trial of oral teriflunomide for relapsing multiple sclerosis. *The New England journal of medicine*; 365(14):1293-303.
36. Sanofi-Aventis groupe (2012): 2.7.3 SUMMARY OF CLINICAL EFFICACY - Teil der Zulassungsunterlagen in Kanada.
37. Bax L., Yu L. M., Ikeda N., Moons K. G. (2007): A systematic comparison of software dedicated to meta-analysis of causal studies. *BMC medical research methodology*; 7:40.
38. Bax L., Yu L. M., Ikeda N., Tsuruta H., Moons K. G. (2006): Development and validation of MIX: comprehensive free software for meta-analysis of causal research data. *BMC medical research methodology*; 6:50.
39. Sanofi-Aventis groupe (2012): Study EFC6260 - TOPIC.
40. Sanofi-Aventis groupe (2012): Study EFC6058 - TERACLES.
41. O'Connor P. W., Lublin F. D., Wolinsky J. S., Confavreux C., Comi G., Freedman M. S., et al. (2013): Teriflunomide reduces relapse-related neurological sequelae, hospitalizations and steroid use. *Journal of neurology*;
42. Wolinsky J. S., Narayana P. A., Nelson F., Datta S., O'Connor P., Confavreux C., et al. (2013): Magnetic resonance imaging outcomes from a phase III trial of teriflunomide. *Multiple sclerosis (Houndmills, Basingstoke, England)*;
43. Miller A. E., O'Connor P., Wolinsky J. S., Confavreux C., Kappos L., Olsson T. P., et al. (2012): Pre-specified subgroup analyses of a placebo-controlled phase III trial (TEMSO) of oral teriflunomide in relapsing multiple sclerosis. *Multiple sclerosis (Houndmills, Basingstoke, England)*; 18(11):1625-32.
44. Sanofi-Aventis groupe (2005): NCT00134563 ClinicalTrials.gov Registereintrag (TEMSO). Stand: Oktober 2012 [Zugriff: 05122012]. URL: <http://www.clinicaltrials.gov/ct2/show/NCT00134563?term=NCT00134563&rank=1>.
45. Sanofi-Aventis groupe (2005): NCT00134563 ClinicalTrials.gov Registereintrag-Ergebnisse (TEMSO). Stand: Oktober 2012 [Zugriff: 05122012]. URL: <http://www.clinicaltrials.gov/ct2/show/NCT00134563?term=NCT00134563&rank=1>.
46. Sanofi-Aventis groupe (2004): EUCTR2004-000555-42-NO ICTRP Registereintrag (TEMSO). Stand: August 2012 [Zugriff: 05.12.2012]. URL: <http://apps.who.int/trialsearch/Trial.aspx?TrialID=EUCTR2004-000555-42-NO>.

47. Sanofi-Aventis groupe (2004): A randomized, double-blind, placebo-controlled, parallel group design study to evaluate the efficacy and safety of teriflunomide (HMR1726D) in reducing the frequency of relapses and delaying the accumulation of physical disability in subjects with multiple sclerosis with relapses; Studienregistereintrag TEMSO EU ClinicalTrialsRegister.
48. Sanofi-Aventis groupe (2008): NCT00751881 ClinicalTrials.gov Registereintrag (TOWER). Stand: Juli 2013 [Zugriff: 30.07.2013]. URL: <http://www.clinicaltrials.gov/ct2/show/NCT00751881?term=NCT00751881&rank=1>.
49. Sanofi-Aventis groupe (2008): NCT00751881 ClinicalTrials.gov Registereintrag Ergebnisse (TOWER). Stand: Juli 2013 [Zugriff: 25072013]. URL: <http://www.clinicaltrials.gov/ct2/show/results/NCT00751881?term=TOWER&rank=1>.
50. Sanofi-Aventis groupe (2008): An Efficacy Study of Teriflunomide in Patients With Relapsing Multiple Sclerosis - Studienregistereintrag ICTRP der TOWER-Studie. Stand: Juli 2013 [Zugriff: 30.07.2013]. URL: <http://apps.who.int/trialsearch/Trial.aspx?TrialID=NCT00751881>.
51. Sanofi-Aventis groupe (2008): 2007-004452-36 Studienregistereintrag EU Clinical Trials Register der Studie TOWER. [Zugriff: 30.07.2013]. URL: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2007-004452-36/DE>.
52. Sanofi-Aventis groupe (2008): 2007-004452-36 A multi-center double-blind parallel-group placebo-controlled study of the efficacy and safety of teriflunomide in patients with relapsing multiple sclerosis; Studienregistereintrag der TOWER-Studie bei PharmNet.Bund.
53. Sanofi-Aventis groupe (2009): NCT00883337 ClinicalTrials.gov Registereintrag (TENERE). Stand: Juli 2013 [Zugriff: 30.07.2013]. URL: <http://www.clinicaltrials.gov/ct2/show/NCT00883337?term=NCT00883337&rank=1>.
54. Sanofi-Aventis groupe (2009): NCT00883337 ClinicalTrials.gov Registereintrag - Ergebnisse (TENERE). Stand: Juli 2013 [Zugriff: 30.07.2013]. URL: <http://www.clinicaltrials.gov/ct2/show/NCT00883337?term=NCT00883337&rank=1>.
55. Sanofi-Aventis groupe (2009): EUCTR2008-006226-34-ES ICTRP Registereintrag (TENERE). Stand: September 2012 [Zugriff: 30.07.2013]. URL: <http://apps.who.int/trialsearch/Trial.aspx?TrialID=EUCTR2008-006226-34-ES>.
56. Sanofi-Aventis groupe (2009): 2008-006226-34 EU Clinical Trials Registereintrag TENERE. [Zugriff: 30.07.2013]. URL: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2008-006226-34/DE>.
57. Sanofi-Aventis groupe (2009): 2008-006226-34 A multi-center, randomized, parallel-group, rater-blinded study comparing the effectiveness and safety of teriflunomide and interferon beta-1a in patients with relapsing multiple sclerosis plus a long-term extension period; Studienregistereintrag TENERE bei PharmNet.Bund.
58. Sanofi-Aventis groupe (2011): Analysis of time to disability progression sustained for 12 weeks-ITT population (TENERE-Studie).
59. Sanofi-Aventis groupe (2011): Analysis of time to confirmed relapse - ITT Population (TENERE-Studie).
60. Sanofi-Aventis groupe (2011): Analysis of time to treatment discontinuation - Randomized Population (TENERE-Studie).
61. European Medicines Agency (2011): Guideline on similar biological medicinal products containing interferon beta. [Zugriff: 13.12.2012]. URL:

- http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2012/01/WC500120652.pdf.
62. Deutsche Gesellschaft für Neurologie (2012): DGN / KKNMS Leitlinie zur Diagnose und Therapie der MS. [Zugriff: 19.04.2012]. URL: http://www.dgn.org/images/stories/dgn/leitlinien/LL_MS_Neu/DGN-KKNMS_MS-LL_20120412_final.pdf.
63. Sormani M. P., Li D. K., Bruzzi P., Stubinski B., Cornelisse P., Rocak S., et al. (2011): Combined MRI lesions and relapses as a surrogate for disability in multiple sclerosis. *Neurology*; 77(18):1684-90.
64. Delestras S., Roustit M., Bedouch P., Minoves M., Dobremez V., Mazet R., et al. (2013): Comparison between two generic questionnaires to assess satisfaction with medication in chronic diseases. *PloS one*; 8(2):e56247.
65. Giovannoni G., Southam E., Waubant E. (2012): Systematic review of disease-modifying therapies to assess unmet needs in multiple sclerosis: tolerability and adherence. *Multiple sclerosis (Houndmills, Basingstoke, England)*; 18(7):932-46.
66. Sanofi (2012): Arava 10mg/20mg/100mg Filmtabletten; Fachinformation. Stand: 12/2012 [Zugriff: 17.01.2012]. URL: <http://www.fachinfo.de>.
67. Freedman M. S., Confavreux C., Comi G., Kappos L., Olsson T. P., Miller A., et al. (2012): Hair thinning associated with teriflunomide therapy is manageable.
68. Bar-Or A., Freedman M. S., Kremenchtzky M., Menguy-Vacheron F., Bauer D., Jodl S., et al. (2013): Teriflunomide effect on immune response to influenza vaccine in patients with multiple sclerosis. *Neurology*;
69. Wiendl Heinz, Kieseier Bernd C. (2010): Multiple Sklerose, Kapitel 1. Stuttgart, Deutschland: Verlag W. Kohlhammer.
70. La Mantia L. (2009): Headache and multiple sclerosis: clinical and therapeutic correlations. *Neurological sciences : official journal of the Italian Neurological Society and of the Italian Society of Clinical Neurophysiology*; 30 Suppl 1:S23-6.
71. Bender R., Lange S., Ziegler A. (2007): Multiples Testen - Artikel Nr. 12 der Statistik-Serie in der DMW -. *Deutsche Medizinische Wochenschrift*; (132):e26-e9.
72. Janatzek S., Arndt C., Windeler J. (2004): G-2 Gutachten: Aussagekraft von Subgruppenanalysen (Fachbereich Evidenz-basierte Medizin, Medizinischer Dienst der Spitzenverbände der Krankenkassen e. V.). [Zugriff: 22.05.2013]. URL: <http://www.mds-ev.de/media/pdf/subgruppen-gutachten.pdf>.
73. Sanofi-Aventis groupe (2012): Additional analyses for the benefit dossier ("Questions raised by Germany").
74. Center for Drug Evaluation and Research (2012): Application Number 202991Orig1s000 Medical Reviews. [Zugriff: 28.05.2013]. URL: http://www.accessdata.fda.gov/drugsatfda_docs/nda/2012/202992Orig1s000MedR.pdf.
75. Goodin D. S., Frohman E. M., Garmany G. P., Jr., Halper J., Likosky W. H., Lublin F. D., et al. (2002): Disease modifying therapies in multiple sclerosis: report of the Therapeutics and Technology Assessment Subcommittee of the American Academy of Neurology and the MS Council for Clinical Practice Guidelines. *Neurology*; 58(2):169-78.
76. EbM-Guidelines (2011): Multiple Sklerose, Encephalitis desseminata. [Zugriff: 03/2012]. URL: http://www.ebm-guidelines.com/ebmga/ltk.koti?p_artikkeli=.
77. Multiple Sclerosis Therapy Consensus Group, Wiendl H., Toyka K. V., Rieckmann P., Gold R., Hartung H. P., et al. (2008): Basic and escalating immunomodulatory

- treatments in multiple sclerosis: current therapeutic recommendations. *Journal of neurology*; 255(10):1449-63.
78. Ebers G. C. et al, (PRISMS-Study Group), (1998): Randomised double-blind placebo-controlled study of interferon beta-1a in relapsing/remitting multiple sclerosis. PRISMS (Prevention of Relapses and Disability by Interferon beta-1a Subcutaneously in Multiple Sclerosis) Study Group. *Lancet*; 352(9139):1498-504.
79. Liu C., Blumhardt L. D. (1999): Randomised, double blind, placebo controlled study of interferon beta-1a in relapsing-remitting multiple sclerosis analysed by area under disability/time curves. *Journal of neurology, neurosurgery, and psychiatry*; 67(4):451-6.
80. Liu C., Blumhardt L. D. (2002): Randomized, double-blind, placebo-controlled study of subcutaneous interferon beta-1a in relapsing-remitting multiple sclerosis: a categorical disability trend analysis. *Multiple sclerosis (Houndmills, Basingstoke, England)*; 8(1):10-4.
81. Patten S. B., Metz L. M. (2001): Interferon beta-1 a and depression in relapsing-remitting multiple sclerosis: an analysis of depression data from the PRISMS clinical trial. *Multiple sclerosis (Houndmills, Basingstoke, England)*; 7(4):243-8.
82. Li D. K., Paty D. W. (1999): Magnetic resonance imaging results of the PRISMS trial: a randomized, double-blind, placebo-controlled study of interferon-beta1a in relapsing-remitting multiple sclerosis. *Prevention of Relapses and Disability by Interferon-beta1a Subcutaneously in Multiple Sclerosis. Annals of neurology*; 46(2):197-206.
83. Rieckmann P., O'Connor P., Francis G. S., Wetherill G., Alteri E. (2004): Haematological effects of interferon-beta-1a (Rebif) therapy in multiple sclerosis. *Drug safety : an international journal of medical toxicology and drug experience*; 27(10):745-56.
84. Kieseier B. C., Stüve O. (2011): A critical appraisal of treatment decisions in multiple sclerosis--old versus new. *Nature reviews Neurology*; 7(5):255-62.
85. Uitdehaag B. M., Barkhof F., Coyle P. K., Gardner J. D., Jeffery D. R., Mikol D. D. (2011): The changing face of multiple sclerosis clinical trial populations. *Current medical research and opinion*; 27(8):1529-37.
86. Giovannoni G., Barbarash O., Casset-Semanaz F., King J., Metz L., Pardo G., et al. (2009): Safety and immunogenicity of a new formulation of interferon beta-1a (Rebif New Formulation) in a Phase IIIb study in patients with relapsing multiple sclerosis: 96-week results. *Multiple sclerosis (Houndmills, Basingstoke, England)*; 15(2):219-28.
87. Sanofi-Aventis groupe (2013): Durchgeführte indirekte Vergleiche TEMSO/TOWER vs. PRISMS inklusive Konsistenzprüfung anhand der Ergebnisse aus TENERE.
88. Patti F., Nicoletti A., Pappalardo A., Castiglione A., Lo Fermo S., Messina S., et al. (2012): Frequency and severity of headache is worsened by Interferon-beta therapy in patients with multiple sclerosis. *Acta neurologica Scandinavica*; 125(2):91-5.
89. Wong S. S., Wilczynski N. L., Haynes R. B. (2006): Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. *Journal of the Medical Library Association : JMLA*; 94(4):451-5.
90. Sanofi-Aventis groupe (2005): Long Term Safety and Efficacy of Teriflunomide (HMR1726) in Multiple Sclerosis with Relapses - ICTRP Studienregistereintrag.

- Stand: 17.10.2012 [Zugriff: 30.07.2013]. URL:
<http://apps.who.int/trialsearch/Trial.aspx?TrialID=NCT00228163>.
91. Sanofi-Aventis groupe (2005): Extension of Protocol HMR1726D/2001, A Phase II Study of the Safety and Efficacy of Teriflunomide (HMR1726) in Multiple Sclerosis with Relapses - Studienregistereintrag clinicaltrials.gov. Stand: 27.03.2012 [Zugriff: 30.07.2013]. URL: <http://www.clinicaltrials.gov/ct2/show/NCT00228163>.
 92. Confavreux C., Li D. K., Freedman M. S., Truffinet P., Benzerdjeb H., Wang D., et al. (2012): Long-term follow-up of a phase 2 study of oral teriflunomide in relapsing multiple sclerosis: safety and efficacy results up to 8.5 years. *Multiple sclerosis (Houndmills, Basingstoke, England)*; 18(9):1278-89.
 93. Sanofi-Aventis groupe (2006): Long Term Safety and Efficacy Study of Teriflunomide 7 mg or 14 mg in Patients With Relapsing-Remitting Multiple Sclerosis - ICTRP Studienregistereintrag TEMSO extension. [Zugriff: 30.07.2013]. URL: <http://apps.who.int/trialsearch/Trial.aspx?TrialID=EUCTR2006-003361-14-DE>.
 94. Sanofi-Aventis groupe (2007): 2006-001152-12 Studienregistereintrag EU Clinical Trials Register TOPIC. [Zugriff: 30.07.2013]. URL: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2006-001152-12/DE>.
 95. Sanofi-Aventis groupe (2007): 2006-03361-14 Studienregistereintrag EU Clinical Trials Register (TEMSO extension). [Zugriff: 30.07.2013]. URL: <https://www.clinicaltrialsregister.eu/ctr-search/trial/2006-003361-14/DE>.
 96. Sanofi-Aventis groupe (2007): Long Term Safety and Efficacy Study of Teriflunomide 7 mg or 14 mg in Patients With Relapsing-Remitting Multiple Sclerosis - Studienregistereintrag clinicaltrials.gov TEMSO extension. [Zugriff: 30.07.2013]. URL: <http://www.clinicaltrials.gov/ct2/show/NCT00803049?term=teriflunomide+AND+Long+Term&rank=2>.
 97. Sanofi-Aventis groupe (2007): Long-term extension of the multinational, double-blind, placebo controlled study EFC6049 (HMR1726D/3001) to document the safety of two doses of teriflunomide (7 and 14 mg) in patients with multiple sclerosis with relapses - Studienregistereintrag Pharmnet TEMSO extension.
 98. Sanofi-Aventis groupe (2007): 2006-001152-12 Studienregistereintrag bei Pharmnet der TOPIC-Studie. [Zugriff: 30.07.2013]. URL.
 99. Sanofi-Aventis groupe (2008): Phase III Study With Teriflunomide Versus Placebo in Patients With First Clinical Symptom of Multiple Sclerosis (TOPIC) - Studienregistereintrag clinicaltrials.gov. [Zugriff: 30.07.2013]. URL: <http://www.clinicaltrials.gov/ct2/show/NCT00622700>.
 100. Sanofi-Aventis groupe (2007): EUCR2006-001152-12-DE Studienregistereintrag ICTRP der TOPIC-Studie. [Zugriff: 30.07.2013]. URL: <http://apps.who.int/trialsearch/Trial.aspx?TrialID=EUCTR2006-001152-12-DE>.

Anhang 4-A: Suchstrategien – bibliografische Literaturrecherche

Geben Sie nachfolgend die Suchstrategien für die in Abschnitt 4.2.3.2 beschriebene bibliografische Literaturrecherche an, und zwar einzeln für jede Datenbank. Geben Sie dabei zunächst jeweils den Namen der durchsuchten Datenbank (z. B. EMBASE), die verwendete Suchoberfläche (z. B. DIMDI, Ovid etc.), das Datum der Suche, das Zeitsegment (z. B.: „1980 to 2010 week 50“) und die gegebenenfalls verwendeten Suchfilter (mit Angabe einer Quelle) an. Listen Sie danach die Suchstrategie einschließlich der resultierenden Trefferzahlen auf. Geben Sie zusätzlich an, für welchen Abschnitt von Modul 4 die Suche durchgeführt wurde.

Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel:

Datenbankname	EMBASE	
Suchoberfläche	Ovid	
Datum der Suche	08.12.2010	
Zeitsegment	1980 to 2010 week 50	
Suchfilter	Filter für randomisierte kontrollierte Studien nach Wong 2006 [Quelle ¹⁷] – Strategy minimizing difference between sensitivity and specificity	
#	Suchbegriffe	Ergebnis
1	Meglitinide/	848
2	Nateglinide/	1686
3	Repaglinide/	2118
4	(glinid* or meglitinid* or nateglinid* or repaglinid*).ab,ti.	1069
5	(starlix or novonorm or novo norm or prandin).ab,ti.	32
6	(105816-04-4 or 135062-02-1).rn.	2854
7	or/1-6	3467
8	Diabetes mellitus/	224164
9	Non Insulin dependent Diabetes mellitus/	91081
10	(diabet* or niddm or t2dm).ab,ti.	379777
11	or/8-10	454517
12	(random* or double-blind*).tw.	650136
13	placebo*.mp.	243550
14	or/12-13	773621
15	and/7,11,14	719

¹⁷ Das Zitat zu dem hier beispielhaft angegebenen Suchfilter lautet wie folgt: Wong SSL, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. J Med Libr Assoc 2006; 94(4): 451-455.

Suchstrategie - randomisierte, kontrollierte Studien mit Teriflunomid 14 mg als Intervention (Abschnitt 4.3.1 und 4.3.2)

Tabelle 4-195 (Anhang): Dokumentation der Suchstrategie in MEDLINE (1)

Datenbankname	Medline	
Suchoberfläche	Ovid	
Datum der Suche	08.08.2013	
Zeitsegment	1946 bis 08.08.2013	
Suchfilter	Filter für RCT nach Wong [89]	
#	Suchbegriffe	Ergebnis
1	randomized controlled trial.pt.	382.290
2	randomized.mp.	576.240
3	randomised.mp.	70.329
4	placebo.mp.	165.223
5	1 OR 2 OR 3 OR 4	643.359
6	teriflunomide.mp	116
7	HMR1726.mp.	3
8	Aubagio.mp.	1
9	6 OR 7 OR 8	117
10	5 AND 9	13

Bei Teriflunomid handelt es sich um einen neuen Wirkstoff in der Therapie der RRMS. Alle Studien zu diesem Wirkstoff wurden durch den pU durchgeführt (s. auch Dokumentation der Studienregistersuche). Dem pU sind die geeigneten Studien und zugehörige Studienpublikationen bekannt, eine Sichtung der Ergebnisse zeigte jedoch, dass die aktuellsten relevanten Publikationen zum Zeitpunkt der Recherche noch nicht in die über Ovid zugängliche Medline-Datenbank aufgenommen waren. Aus diesem Grund wurde diese Recherche über eine Medline-Suche über Pubmed ergänzt.

Tabelle 4-196 (Anhang): Dokumentation der Suchstrategie in MEDLINE (2)

Datenbankname	Medline	
Suchoberfläche	Pubmed	
Datum der Suche	26.07.2013	
Zeitsegment	Keine Einschränkung	
Suchfilter	Filter für RCT nach Wong [89] adaptiert für Pubmed	
#	Suchbegriffe	Ergebnis
1	randomized controlled trial [pt]	347.757
2	controlled clinical trial [pt]	85.843
3	randomized [tiab]	292.902
4	placebo [tiab]	150.314
5	clinical trials as topic [mesh: noexp]	163.924
6	randomly [tiab]	199.613
7	trial [ti]	116.751
8	1 OR 2 OR 3 OR 4 OR 5 OR 6 OR 7	856.800
9	animals [mh] NOT humans [mh]	3.801.975
10	8 NOT 9	791.841
11	teriflunomide	101
12	teriflunomid*	101
13	HMR1726	102
14	11 OR 12 OR 13	102
15	10 AND 14	23

Tabelle 4-197 (Anhang): Dokumentation der Suchstrategie in EMBASE

Datenbankname	EMBASE	
Suchoberfläche	Ovid	
Datum der Suche	19.08.2013	
Zeitsegment	1974 bis Woche 33 2013	
#	Suchbegriffe	Ergebnis
1	random:.tw.	850.279
2	placebo:.mp.	321.515
3	double-blind:.tw.	146.039
4	1 OR 2 OR 3	1.058.984
5	randomized controlled trial/	356.668
6	4 OR 5	1.126.328
7	controlled clinical trial/	404.592
8	randomi?ed.mp.	625.891
9	7 OR 8	751.094
10	6 OR 9	1.249.392
11	multiple sclerosis/	73.819
12	teriflunomide/	829
13	teriflunomid*.mp.	842
14	HMR1726/	829
15	Aubagio/	660
16	12 OR 13 OR 14 OR 15	842
17	10 AND 11 AND 16	210

Tabelle 4-198 (Anhang): Dokumentation der Suchstrategie in Cochrane

Datenbankname	Cochrane	
Suchoberfläche	Cochrane Library	
Datum der Suche	01.08.2013	
Zeitsegment	Keine Einschränkung	
Suchfilter	Kein Suchfilter	
#	Suchbegriffe	Ergebnis
1	teriflunomide	20
2	teriflunomid*	20
3	HMR1726	2
4	Aubagio*	1
5	1 OR 2 OR 3 OR 4	20

Von den 20 Treffern sind **15 Treffer** in der Kategorie „Trials“.

Suchstrategie - randomisierte, kontrollierte Studien, die Rebif 44 µg, dreimal die Woche, mit Placebo vergleichen (Abschnitt 4.3.2)

Tabelle 4-199 (Anhang): Dokumentation der Suchstrategie in MEDLINE

Datenbankname	Medline	
Suchoberfläche	Ovid	
Datum der Suche	19.08.2013	
Zeitsegment	Filter für RCT nach Wong [89]	
Suchfilter	1946 to 19.08.2013	
#	Suchbegriffe	Ergebnis
1	Multiple Sclerosis/	39.519
2	Multiple Sclerosis, Relapsing-Remitting/	3.405
3	1 OR 2	42.578
4	Interferon-beta/	7.644
5	rebif.mp.	197
6	4 OR 5	7.679
7	randomized controlled trial.pt.	382.610
8	randomized.mp.	576.951

9	placebo.mp.	165.434
10	7 OR 8 OR 9	627.244
11	3 AND 6 AND 10	498

Tabelle 4-200 (Anhang): Dokumentation der Suchstrategie in EMBASE

Datenbankname	EMBASE	
Suchoberfläche	Ovid	
Datum der Suche	05.08.2013	
Zeitsegment	1974 bis Woche 31 2013	
Suchfilter	Filter für RCT nach Wong	
#	Suchbegriffe	Ergebnis
1	beta 1a interferon/ OR exp beta 1 interferon/	5270
2	rebif*.mp.	1836
3	interferon beta 1a*.mp.	126
4	1 OR 2 OR 3	5307
5	random:.tw.	847607
6	placebo:.mp.	320688
7	double-blind:.tw.	145702
8	5 OR 6 OR 7	1055872
9	multiple sclerosis/ OR multiple sclerosis*.mp.	80702
10	4 AND 8 AND 9	1225

Tabelle 4-201 (Anhang): Dokumentation der Suchstrategie in Cochrane

Datenbankname	Cochrane	
Suchoberfläche	Cochrane Library	
Datum der Suche	01.08.2013	
Zeitsegment	Keine Einschränkung	
Suchfilter	Kein Suchfilter	
#	Suchbegriffe	Ergebnis
1	MeSH descriptor: [Multiple Sclerosis] explode all trees	1646
2	Multiple Sclerosis	3396

3	1 OR 2	3400
4	MeSH descriptor: [Placebos] explode all trees	21225
5	Placebo	142700
6	4 OR 5	142700
7	interferon	8948
8	interferon*	8969
9	Rebif	90
10	7 OR 8 OR 9	8978
11	3 AND 6 AND 10	335

Von den 335 Treffern sind **264 Treffer** in der Kategorie “Trials”

Anhang 4-B: Suchstrategien – Suche in Studienregistern

Geben Sie nachfolgend die Suchstrategien für die in Abschnitt 4.2.3.3 beschriebene Suche in Studienregistern an, und zwar einzeln für jedes Studienregister. Geben Sie dabei jeweils den Namen des durchsuchten Studienregisters (z. B. *clinicaltrials.gov*), die Internetadresse, unter der das Studienregister erreichbar ist (z. B. *http://www.clinicaltrials.gov*), das Datum der Suche, die verwendete Suchstrategie und die resultierenden Treffer an. Geben Sie zusätzlich an, für welchen Abschnitt von Modul 4 die Suche durchgeführt wurde.

Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel:

Studienregister	clinicaltrials.gov
Internetadresse	http://www.clinicaltrials.gov
Datum der Suche	08.12.2010
Suchstrategie	(Starlix OR Novonorm OR Prandin OR Nateglinid OR Repaglinid) [ALL-FIELDS] AND ("Phase II" OR "Phase III" OR "Phase IV") [PHASE]
Treffer	23

Für Abschnitt 4.3.1 und 4.3.2 (zu bewertendes Arzneimittel)

Studienregister	clinicaltrials.gov
Internetadresse	http://www.clinicaltrials.gov
Datum der Suche	29.07.2013
Suchstrategie	teriflunomide OR aubagio [ALL-FIELDS]
Treffer	16

Alle 16 Treffer sind Studien des pharmazeutischen Unternehmens.

Kommentar: eine alternative Suche nach Studien mit „HMR1726“ oder „Teriflunomid“ anstelle von „Teriflunomide“ brachte keine zusätzlichen Treffer.

Studienregister	International Clinical Trials Registry Platform
Internetadresse	http://apps.who.int/trialsearch/
Datum der Suche	29.07.2013
Suchstrategie	teriflunomide OR aubagio
Treffer	95 Datensätze (davon 16 Studien)

Alle 16 Treffer sind Studien des pharmazeutischen Unternehmens.

Kommentar: eine alternative Suche nach Studien mit „HMR1726“ anstelle von „Teriflunomide“ brachte keine ergänzenden Treffer.

Studienregister	EU Clinical Trials Register
Internetadresse	https://www.clinicaltrialsregister.eu/
Datum der Suche	29.07.2013
Suchstrategie	teriflunomide OR aubagio
Treffer	10

Alle 10 Treffer sind Studien des pharmazeutischen Unternehmers.

Kommentar: eine alternative Suche nach Studien mit „HMR1726“ anstelle von „Teriflunomide“ brachte keine ergänzenden Treffer.

Studienregister	PharmNet
Internetadresse	https://www.pharmnet-bund.de/dynamic/de/index.html
Datum der Suche	29.07.2013
Suchstrategie	teriflunomide (Suche im Titel ^a)
Treffer	8

Alle 8 Treffer sind Studien des pharmazeutischen Unternehmers.

^aEine Suche mit dem gleichen Stichwort in der Kategorie „Active Substance“ gab keine abweichenden Treffer.

Kommentar: eine alternative Suche nach Studien mit „HMR1726“ anstelle von „Teriflunomide“ brachte keine ergänzenden Treffer.

Für Abschnitt 4.3.2 (zweckmäßige Vergleichstherapie)

Studienregister	clinicaltrials.gov
Internetadresse	http://www.clinicaltrials.gov
Datum der Suche	29.07.2013
Suchstrategie	(interferon OR rebif) [ALL-FIELDS] AND multiple sclerosis [CONDITION]
Treffer	235

Keiner der insgesamt 235 Treffer konnte entsprechend den vorab definierten Ein- und Ausschlusskriterien als relevant eingestuft werden.

Studienregister	International Clinical Trials Registry Platform
Internetadresse	http://apps.who.int/trialsearch/
Datum der Suche	06.08.2013
Suchstrategie	multiple sclerosis [Condition] AND interferon* OR rebif* [Intervention]
Treffer	64 Datensätze (davon 46 Trials)

Keine der insgesamt 46 Studien konnte entsprechend den vorab definierten Ein- und Ausschlusskriterien als relevant eingestuft werden.

Es wurde der Bereich „Advanced Search“ verwendet.

Studienregister	PharmNet
Internetadresse	https://www.pharmnet-bund.de/dynamic/de/index.html
Datum der Suche	29.07.2013
Suchstrategie	interferon? [Active substance] AND multiple sclerosis [Medical condition]
Treffer	20

Keiner der 20 Treffer erfüllte die vorab definierten Einschlusskriterien.

Studienregister	EU Clinical Trials Register
Internetadresse	https://www.clinicaltrialsregister.eu/
Datum der Suche	06.08.2013
Suchstrategie	multiple sclerosis AND (interferon* OR rebif*)
Treffer	100

Keiner der 100 Treffer erfüllte die vorab definierten Einschlusskriterien.

Anhang 4-C: Liste der im Volltext gesichteten und ausgeschlossenen Studien mit Ausschlussgrund

Listen Sie die im Volltext gesichteten und ausgeschlossenen Studien unter Nennung des Ausschlussgrunds auf.

Tabelle 4-202 (Anhang): Dokumentation des Volltext-Screenings (Recherche nach randomisierten, kontrollierten Studien, die Teriflunomid 14 mg, einmal pro Tag, mit Placebo vergleichen), ausgeschlossene Studien

Nr.	Ausschlussgrund	Referenz
1	A6	Confavreux C., Li D. K., Freedman M. S., Truffinet P., Benzerdjeb H., Wang D., et al. (2012): Long-term follow-up of a phase 2 study of oral teriflunomide in relapsing multiple sclerosis: safety and efficacy results up to 8.5 years. <i>Multiple sclerosis</i> (Houndmills, Basingstoke, England); 18(9):1278-89
2	A5	Wood H. (2011): Multiple sclerosis: Teriflunomide shows promise for MS treatment in phase III trial. <i>Nature reviews Neurology</i> ; 7(12):657.

Tabelle 4-203 (Anhang): Dokumentation des Volltext-Screenings (Recherche nach randomisierten, kontrollierten Studien, die Rebif 44 µg, dreimal die Woche, mit Placebo vergleichen), ausgeschlossene Studien

Nr.	Ausschlussgrund	Referenz
1	A4	Ann Neurol. 1999 Aug;46(2):197-206. Magnetic resonance imaging results of the PRISMS trial: a randomized, double-blind, placebo-controlled study of interferon-beta1a in relapsing-remitting multiple sclerosis. Prevention of Relapses and Disability by Interferon-beta1a Subcutaneously in Multiple Sclerosis. Li DK, Paty DW.
2	A4	J Neurol Neurosurg Psychiatry. 1999 Oct;67(4):451-6. Randomised, double blind, placebo controlled study of interferon beta-1a in relapsing-remitting multiple sclerosis analysed by area under disability/time curves. Liu C, Blumhardt LD.
3	A4	Mult Scler. 2001 Aug;7(4):243-8. Interferon beta-1 a and depression in relapsing-remitting multiple sclerosis: an analysis of depression data from the PRISMS clinical trial. Patten SB, Metz LM.
4	A4	Mult Scler. 2002 Feb;8(1):10-4. Randomized, double-blind, placebo-controlled study of subcutaneous interferon beta-1a in relapsing-remitting multiple sclerosis: a categorical disability trend analysis. Liu C, Blumhardt LD.
5	A7	Evidence-Based Medicine. 1999; 4 (3): 74-75 Interferon-beta 1a reduced relapses at 2 years in relapsing-remitting multiple sclerosis

Nr.	Ausschluss- grund	Referenz
6	A5	O'Connor P Drug Saf. 2004;27(10):745-56. Haematological effects of interferon-beta-1a (Rebif) therapy in multiple sclerosis. Rieckmann P, O'Connor P, Francis GS, Wetherill G, Alteri E.

Anhang 4-D: Liste der abgebrochenen Studien

Benennen Sie in der nachfolgenden Tabelle die abgebrochenen Studien aus dem durch die verschiedenen Suchschritte in Abschnitt 4.3.1.1 identifizierten Studienpool.

Tabelle 4-204 (Anhang): Studienpool, abgebrochene Studien, zu bewertendes Arzneimittel

Studie	Studienkategorie			verfügbare Datenquellen ^a		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	gesponserte Studie ^b (ja/nein)	Studie Dritter (ja/nein)	Studienbericht (ja/nein [Zitat])	Registereintrag ^c (ja/nein [Zitat])	Publikation (ja/nein [Zitat])
ggf. Zwischenüberschrift zur Strukturierung des Studienpools						
placebokontrolliert						
trifft nicht zu						
aktivkontrolliert, zweckmäßige Vergleichstherapie(n)						
trifft nicht zu						
a: Bei Angabe „ja“ sind jeweils die Zitate der Datenquelle(n) mit anzugeben.						
b: Studie, für die der Unternehmer Sponsor war oder an der der Unternehmer anderweitig finanziell beteiligt war.						
c: Zitat der Studienregistereinträge sowie, falls vorhanden, der in den Studienregistern aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.						

Anhang 4-E: Liste der laufenden Studien

Benennen Sie in der nachfolgenden Tabelle die laufenden Studien aus dem durch die verschiedenen Suchschritte in Abschnitt 4.3.1.1 identifizierten Studienpool.

Tabelle 4-205 (Anhang): Studienpool, laufende Studien, zu bewertendes Arzneimittel

Studie	Studienkategorie			verfügbare Datenquellen ^a		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	gesponserte Studie ^b (ja/nein)	Studie Dritter (ja/nein)	Studienbericht (ja/nein [Zitat])	Registereintrag ^c (ja/nein [Zitat])	Publikation (ja/nein [Zitat])
ggf. Zwischenüberschrift zur Strukturierung des Studienpools						
placebokontrolliert						
HMR1726D/2002 (LTS6048, Verlängerung der Phase II Monotherapie-Studie)	nein	ja	nein	nein	ja [90, 91]	ja [92]
Verlängerung der TEMSO-Studie (HMR1726D/3004, LTS6050)	nein	ja	nein	nein	ja [93-97]	nein
TOPIC (HMR1726D-3005, EFC6260)	nein	ja	nein	nein	ja [94, 98-100]	nein
Verlängerung der TOWER-Studie	nein	ja	nein	nein	ja [49, 50, 52]	nein
Verlängerung der TENERE-Studie	nein	ja	nein	nein	ja [54, 55, 57]	nein
aktivkontrolliert, zweckmäßige Vergleichstherapie(n)						
<p>a: Bei Angabe „ja“ sind jeweils die Zitate der Datenquelle(n) mit anzugeben. b: Studie, für die der Unternehmer Sponsor ist oder an der der Unternehmer anderweitig finanziell beteiligt ist. c: Zitat der Studienregistereinträge sowie, falls vorhanden, der in den Studienregistern aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.</p>						

Studie	Studienkategorie			verfügbare Datenquellen ^a		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	gesponserte Studie ^b (ja/nein)	Studie Dritter (ja/nein)	Studienbericht (ja/nein [Zitat])	Registereintrag ^c (ja/nein [Zitat])	Publikation (ja/nein [Zitat])
d: Es handelt sich bei beiden Studien, um Verlängerungen placebokontrollierter Studien, in denen die Placebo-Patienten nach Beendigung der Hauptstudie in die Gruppen 7 mg und 14 mg Teriflunomid randomisiert wurden.						

Anhang 4-F: Methodik der eingeschlossenen Studien – RCT

Beschreiben Sie nachfolgend die Methodik jeder eingeschlossenen, in Abschnitt 4.3.1.1.4 genannten Studie. Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle 4-206 inklusive eines Flow-Charts für den Patientenfluss.

Sollten Sie im Dossier indirekte Vergleiche präsentieren, beschreiben Sie ebenfalls die Methodik jeder zusätzlich in den indirekten Vergleich eingeschlossenen Studie (Abschnitt 4.3.2.1). Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle 4-206 inklusive eines Flow-Charts für den Patientenfluss.

Tabelle 4-206 (Anhang): Studiendesign und -methodik für Studie <Studienbezeichnung>

Item ^a	Charakteristikum	Studieninformation
-	Studienziel	
2 b	Genauere Ziele, Fragestellung und Hypothesen	
-	Methoden	
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	
4	Probanden / Patienten	
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	
4b	Studienorganisation und Ort der Studiendurchführung	
5	Interventionen Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	
6	Zielkriterien	
6a	Klar definierte primäre und sekundäre Zielkriterien, Erhebungszeitpunkte, ggf. alle zur Optimierung der Ergebnisqualität verwendeten Erhebungsmethoden (z. B. Mehrfachbeobachtungen, Training der Prüfer) und ggf. Angaben zur Validierung von Erhebungsinstrumenten	
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	
8	Randomisierung, Erzeugung der Behandlungsfolge	
8a	Methode zur Generierung der zufälligen Zuteilung	
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	
9	Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment) Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war	
10	Randomisierung, Durchführung Wer führte die Zuteilung durch, wer nahm die	

Item ^a	Charakteristikum	Studieninformation
	Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	
11	Verblindung	
11a	Waren a) die Probanden / Patienten und / oder b) diejenigen, die die Intervention / Behandlung durchführten, und / oder c) diejenigen, die die Zielgrößen beurteilten, verblindet oder nicht verblindet, wie wurde die Verblindung vorgenommen?	
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	
-	Resultate	
13	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	
13a	Anzahl der Studienteilnehmer für jede durch Randomisierung gebildete Behandlungsgruppe, die a) randomisiert wurden, b) tatsächlich die geplante Behandlung/Intervention erhalten haben, c) in der Analyse des primären Zielkriteriums berücksichtigt wurden	
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	
14	Aufnahme / Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	
14b	Informationen, warum die Studie endete oder beendet wurde	
a: nach CONSORT 2010.		

Stellen Sie den Patientenfluss in einem Flow-Chart je Studie dar.

Tabelle 4-207 (Anhang): Studiendesign und -methodik für die Studie TEMSO

Item ^a	Charakteristikum	Studieninformation
-	Studienziel	
2 b	Genaue Ziele, Fragestellung und Hypothesen	Ziel: Beurteilung der Effektivität von 2 Dosierungen Teriflunomid im Vergleich mit Placebo in Bezug auf die Schubrate und die Behinderungsprogression in Patienten mit schubförmiger MS.
-	Methoden	
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	Bei TEMSO handelt es sich um eine randomisierte, multizentrische, doppelblinde, placebokontrollierte Parallelgruppen-Studie. Die Patienten wurden in einem Verhältnis 1:1:1 den Gruppen Teriflunomid 14 mg, Teriflunomid 7 mg und Placebo zugeteilt.
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	Zum Studienprotokoll wurden 9 Amendments erstellt. Keines dieser Amendments beinhaltet Änderungen, die einen relevanten Einfluss auf die Interpretation der Studien haben. Folgende wesentliche Änderungen ergaben sich mit den Amendments (für weitere Änderungen s. Studienbericht [32]): Amendment 1: <ul style="list-style-type: none"> • Erhöhung der Anzahl der teilnehmenden Zentren (von 60 auf 120) • zusätzliche Sicherheitsmaßnahmen (Washout-Prozedur für gebärfähige Patientinnen und Männer mit dem Potential ein Kind zu zeugen) - gefordert von der FDA Amendment 2: <ul style="list-style-type: none"> • Erfassung der MS-Subtyp-Diagnose zum Studieneintrittsdatum Amendment 3: <ul style="list-style-type: none"> • Erhöhung der Probengröße von 870 auf 1080 Patienten • Klarstellung von dem Effekt von Teriflunomid auf die Behinderungsprogression als sekundäres Zielkriterium – erfasst mit EDSS • Definition der MRT-Variablen • Entfernung des „Physical Self-Maintenance Scale Questionnaire“ aufgrund von Schwierigkeiten in der Durchführung Amendment 4: <ul style="list-style-type: none"> • Hinzufügen einer klinischen Follow-up-Phase aller Patienten, die die Studie vorzeitig und permanent abgebrochen haben (um eine Dokumentation unerwünschter Ereignisse zu ermöglichen).

		<p>Amendment 7:</p> <ul style="list-style-type: none"> • Klarstellung von statistischen Methoden (primäre Analyse (Poisson Regression mit robuster Fehlervarianz, was dem GEE (generalized estimating equation)-Modell entspricht) und Handhabung von vorzeitigem Ausscheiden aus der Studie; Änderung bei dem wichtigsten Zielkriterium „Zeit bis zur Behinderungsprogression“ und Aktualisierung der sekundären Analyse; Modifizierung der Vorgehensweise zur Kontrolle von Multiplizität; Änderung der modifizierten ITT (mITT)-Population („primary efficacy population“; Stratifizierung nach Zentrum und EDSS Score) <p>Amendment 8:</p> <ul style="list-style-type: none"> • Klarstellungen, die zur Harmonisierung verschiedener Protokolle innerhalb des Teriflunomid-Entwicklungsprogramms notwendig waren (Handhabung von vorzeitigem Ausscheiden aus der Studie; TEAE-Zeitraum abgeändert auf bis zu 16 Wochen nach Absetzen der Studienmedikation; detaillierte Instruktionen für Schwangerschaften; Fehlerkorrektur bei EDSS-basierter Behinderungsprogression (1 Punkt, wenn der Baseline-Wert $\leq 5,5$ ist); BOD als wichtigster MRT-Parameter, alle anderen lediglich explorativ; MS Schub-Daten erfasst als Effektivitätsparameter und nicht als unerwünschtes Ereignis)
4	Probanden / Patienten	
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	<p>Einschlusskriterien:</p> <ul style="list-style-type: none"> • MS-Patienten zwischen 18 und 55 Jahren • EDSS $\leq 5,5$ • Schubförmiger Verlauf der Erkrankung, mit oder ohne Progression (RRMS, SPMS oder PRMS) • Erfüllung der MS-Diagnose-Kriterien nach McDonald • Mindestens 1 Schub in dem Jahr vor Beginn der Studie oder mindestens 2 Schübe innerhalb der letzten 2 Jahre vor der Studie • Keinen Schub in den 60 Tagen vor Randomisierung • Innerhalb der letzten 4 Wochen vor Randomisierung mussten die Patienten klinisch stabil sein, ohne ACTH oder systemischer Steroidbehandlung • Schriftliche Zustimmung sowohl zur Studienteilnahme als auch zur HIV-Testierung liegt vor <p>Ausschlusskriterien:</p> <ul style="list-style-type: none"> • Patienten mit signifikant eingeschränkter Knochenmarksfunktion oder Anämie, Leukopenie oder Thrombozytopenie: <ul style="list-style-type: none"> – Hämatokrit $< 24\%$ und/oder

		<ul style="list-style-type: none"> – Absolute Zahl der weißen Blutzellen <4000 Zellen/mm³ und/oder – Plättchenzahl <150000 Zellen/mm³ und/oder – Absolute Neutrophilenzahl ≤1500 Zellen/mm³. <ul style="list-style-type: none"> • Patienten mit angeborener oder erworbener schwerer Immunschwäche, einer Krebsvorgeschichte (mit Ausnahme von Basalzell- und Plattenepithelkarzinom, die operativ entfernt wurden und für die keine Metastase nachgewiesen werden konnte), lymphoproliferative Erkrankungen oder ein Patient, der eine Strahlentherapie des lymphatischen Systems erhalten hat. • HIV-positive Patienten • Bekannte Vorgeschichte von aktiver Tuberkulose, die nicht angemessen behandelt wurde • Persistierende signifikante oder schwere Infektion • Schwangere oder stillende Frauen • Patienten, die im Verlauf der Studiendauer Eltern werden möchten • Nicht erlaubte Therapien (mindestens 4 Wochen vor Randomisierung): Phenytoin, Warfarin, Tolbutamid, Johanniskraut, Cholestyramin • Patienten, die in den letzten 4 Wochen vor Randomisierung ACTH oder systemische Kortikosteroide eingenommen haben • Vorherige oder gleichzeitige Einnahme von Cladribin, Mitoxantron oder anderer immunsuppressiver Agentien, sowie Azathioprin, Cyclophosphamid, Cyclosporin, Methotrexat oder Mycophenolat • Vorherige Verwendung von Interferonen oder eine Zytokintherapie in den vorhergehenden 4 Monaten • Vorherige Verwendung von <ul style="list-style-type: none"> – Glatirameracetat in den vorhergehenden 4 Monaten – Intravenösen Immunglobulinen in den vorhergehenden 6 Monaten • Vorherige Einnahme irgendeines Prüfpräparates innerhalb der vorhergehenden 6 Monate • Kontraindikation gegen MRT, z. B. Patienten mit Herzschrittmacher, metallischen Implantaten (z. B. künstlichen Herzklappen, Aneurysma/Gefäßklammer), Anwesenheit metallischen Materials (z. B. Granatsplitter) in empfindlichen Regionen, bekannte Vorgeschichte von Allergie gegen irgendein Kontrastmittel, bekannte Vorgeschichte von Klaustrophobie, die ein vollständige Aufnahme aller geplanten MRTs verhindern würde. Hüftimplantate
--	--	--

		<p>waren keine Kontraindikation.</p> <ul style="list-style-type: none"> • Leberfunktionsstörung oder persistierende Erhöhungen der Serum-Glutamat-Pyruvat-Transferase/Alanin-Aminotransferase (SGPT/ALT), Serum-Glutamat-Oxalacetat-Transaminase/Aspartat-Aminotransferase (SGOT/AST) oder direktes Bilirubin mit einem Wert, der mehr um das 1,5fache des normalen Wertes (ULN) erhöht ist. • Bekannte Hepatitis-Vorgeschichte • Hyperproteinämie (z. B. im Fall von schweren Lebererkrankungen oder eines nephrotischen Syndroms) mit einem Serumalbuminwert <3,0 g/dL • Moderate bis schwere Einschränkung der renalen Funktion, gezeigt durch einen Serumkreatininwert >133 µmol/L (oder >1.5 mg/dL) • Vorherige Behandlung mit Teriflunomid oder Leflunomid (ARAVA®) • Wahrscheinlichkeit während der Studiendauer nicht erlaubte Medikation zu benötigen • Klinisch relevante kardiovaskuläre, hepatische, neurologische, endokrine oder andere schwere systemische Erkrankungen, die eine Anwendung des Protokolls oder eine Interpretation der Studienresultate schwierig machen oder die bedeuten, dass eine Teilnahme an der Studie für die Patienten ein Risiko ist. • Alkohol- oder Drogenmissbrauch in der Vorgeschichte • Geisteszustand, der es dem Patienten nicht erlaubt, die Art, den Umfang und mögliche Konsequenzen der Studie zu verstehen • Patienten, die sich wahrscheinlich nicht an das Studienprotokoll halten werden (unkooperative Einstellung, fehlende Möglichkeit Folgeuntersuchungen wahrzunehmen, bekannte Wahrscheinlichkeit, dass die Studie nicht beendet wird) • Patient gehört zum Personal oder ist verwandt mit Personal (z. B. Studienleiter, Forschungsassistent, Apotheker, ...), das direkt in die Durchführung der Studie eingebunden ist • Vorherige Verwendung von Natalizumab (Tysabri®) • Persistierende Anhebungen der Serumamylase oder Lipase auf einen Wert der mehr als das Zweifache des Normwertes beträgt • Bekannte Vorgeschichte einer pankreatischen Erkrankung oder Pankreatitis
4b	Studienorganisation und Ort der Studiendurchführung	Die Studie wurde in 126 Zentren in insgesamt 21 Ländern durchgeführt (Österreich, Kanada, Chile, Tschechien,

		Dänemark, Estland, Finnland, Frankreich, Deutschland, Italien, Niederlande, Norwegen, Polen, Portugal, Russische Föderation, Schweden, Schweiz, Türkei, Ukraine, Großbritannien, USA)
5	Interventionen Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	Teriflunomid 7 mg: Filmtabletten, 1x täglich morgens, oral (mit Wasser) Teriflunomid 14 mg: Filmtabletten, 1x täglich morgens, oral (mit Wasser) Placebo: Filmtabletten, 1x täglich morgens, oral (mit Wasser)
6	Zielkriterien	
6a	Klar definierte primäre und sekundäre Zielkriterien, Erhebungszeitpunkte, ggf. alle zur Optimierung der Ergebnisqualität verwendeten Erhebungsmethoden (z. B. Mehrfachbeobachtungen, Training der Prüfer) und ggf. Angaben zur Validierung von Erhebungsinstrumenten	<p>Das primäre Ziel der Studie war:</p> <ul style="list-style-type: none"> die Erfassung des Effektes von Teriflunomid auf die Reduktion der Schubrate in Personen mit schubförmiger MS. <p>Die sekundären Ziele der Studie waren die Bewertung von:</p> <ul style="list-style-type: none"> dem Effekt von Teriflunomid auf die Hinauszögerung der Behinderungsprogression nach 2 Jahren (gemessen mit Kurtzkes Expanded Disability Status Scale (EDSS)). Erhebungszeitpunkte: Screening, Woche 0, dann alle 12 Wochen und im Fall eines Schubes. den Effekten von Teriflunomid auf MRT-Variablen (Krankheitslast (Volumen an abnormalem Gehirngewebe im MRT) und andere MRT-Variablen inklusive Anzahl und Volumen an Gadolinium-anreichernden T1-Läsionen, Volumen an T2-Läsionen, Volumen an T1 hypointenser Läsionen, Atrophie und z-Score. Erhebungszeitpunkte: Woche 0, Wochen 24, 48, 72 und 108. den Effekten von Teriflunomid auf patientenberichtete Fatigue (gemessen mit dem Fatigue Impact Scale (FIS)). Erhebungszeitpunkte: Screening, Woche 0, dann alle 12 Wochen und im Fall eines Schubes. der Sicherheit und Verträglichkeit von Teriflunomid durch Nebenwirkungsreports, physische Untersuchungen, Vitalparameter, Laboruntersuchungen. <p>Termine für „adverse event reporting“ waren - abgesehen vom Datum der Randomisierung- angesetzt in den Wochen 4, 8, 12, 16, 20, 24 und 30 und dann alle 6 Wochen bis zum Studienende</p> <p>Die tertiären Zielkriterien waren die Untersuchung von:</p> <ul style="list-style-type: none"> dem Einfluss von Teriflunomid auf den allgemeinen Gesundheitszustand mit Hilfe des SF-36 Fragebogens (Short Form 36 Health Survey). Erhebungszeitpunkte: Baseline-Wert, Wochen 20, 42, 66, 90 und 108. dem ökonomischen Einfluss von Teriflunomid durch

		<p>eine Bewertung der Ressourcenverwendung, Beschränkung von Produktivität und Aktivitäten (WPAI) und EuroQoL (EQ-5D)-Fragebogen</p> <p>Erhebungszeitpunkte (EQ-5D): Screening, Woche 0, dann alle 12 Wochen und im Fall eines Schubes.</p> <ul style="list-style-type: none"> • dem Einfluss von Teriflunomid auf das Fortschreiten der Krankheitsaktivität mit Hilfe des MSFC <p>Erhebungszeitpunkte: Screening, Wochen 0, 24, 48, 72 und 96</p> <ul style="list-style-type: none"> • der Beziehung zwischen Dosis, Konzentration, Effektivität, Sicherheit und Blutspiegel von Teriflunomid • dem Zusammenhang zwischen den wichtigsten Enzymen des Teriflunomid-Metabolismus und Leberschädigung, bzw. anderer potentieller Verbindungen zwischen Genvariationen und klinischem Endpunkt (optionale pharmakogenomische Substudie)
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	Die Zielkriterien wurden nach Studienbeginn nicht geändert.
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	<p>Die Abschätzung der Fallzahlen basiert auf einer 1:1:1-Randomisierung auf Teriflunomid 14 mg: Teriflunomid 7 mg: Placebo und die primären und wichtigen sekundären Wirksamkeitsvergleiche zwischen Teriflunomid und Placebo hinsichtlich jeweils jährlicher Schubrate und Zeit bis zur Behinderungsprogression.</p> <p>Ausgehend von einer jährlichen Schubrate von 0,74 unter Placebo mit 370 randomisierten Patienten per Teriflunomid- oder Placebo-Behandlungsarm, hat die Studie eine 94%-Power eine relative Risikoreduktion von 25 % zu detektieren (d. h. eine jährliche Schubrate von 0,55 für Teriflunomid) für den Endpunkt jährliche Schubrate mit einem zweiseitigen Signifikanzlevel von $\alpha=0,05$.</p> <p>Die Berechnung von Fallzahl und Power stützt sich auf das Simulationsergebnis einer Poissonverteilung, das Simulationsprogramm wird im Analysenplan zur Verfügung gestellt werden.</p> <p>Die Berechnung geht von der Annahme aus, dass die Schubrate ungefähr einer Poisson-Verteilung folgt mit einem Poisson-Parameter von 1,3, welcher durch aktuelle klinische Studien abgeschätzt wurde (Tysabri statistical review document and internal protocol 2001, LTS6048); die jährliche Schubrate von 0,74 für Placebo basiert auf aktuell verfügbaren klinischen Studien im Bereich MS, in diesen Studien ist die 2-Jahres-Schubrate auf 1,48 geschätzt; die Berechnung geht weiterhin von der Annahme aus, dass die Rekrutierung 1,5 Jahre dauert mit einer linearen Rekrutierungsrate, so ist die durchschnittliche Exposition für laufende Patienten 1,75 Jahre. Schließlich ist die Fallzahl entsprechend einer 20%-</p>

		<p>Dropout-Rate adjustiert. Zusätzlich wird die Studie eine 75%ige Power haben, um eine 37%ige Reduktion des Hazard Ratios bezüglich der Zeit bis zur Behinderungsprogression (unter Verwendung eines Log Rank Tests) nachzuweisen.</p> <p>Die Fallzahl und die Power-Abschätzung ist berechnet unter Verwendung des Programms EAST 4.0 mit den Annahmen einer Hazardrate von 0,1783 in der Placebogruppe und 0,1116 in der Teriflunomid-Gruppe (d. h. 30 % Wahrscheinlichkeit einer Behinderungsprogression für die Placebo-Patienten bis zum Ende der 2 Jahre, und eine entsprechende Wahrscheinlichkeit von 20 % für die Teriflunomid-Patienten). Die Fallzahl ist auch entsprechend einer 20%-Dropout-Rate adjustiert.</p>
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	Es wurden keine Zwischenanalysen durchgeführt und die Studie wurde auch nicht vorzeitig abgebrochen.
8	Randomisierung, Erzeugung der Behandlungsfolge	
8a	Methode zur Generierung der zufälligen Zuteilung	Die zufällige Zuteilung erfolgte über ein interaktives Spracherkennungssystem (interactive voice response system, IVRS).
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	Blockrandomisierung (Blockgröße: 6) Die Randomisierung war hinsichtlich Studienzentrum und EDSS-Score ($\leq 3,5$ oder $> 3,5$)
9	<p>Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment)</p> <p>Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war</p>	<p>Die Studienmedikation wurde in nummerierten Behältern (5-stellige Zahlen) ausgeteilt. Die Behälter wurden den Patienten über das IVRS zugeteilt.</p> <p>Die Geheimhaltung war bis zur Zuteilung gewährleistet.</p>
10	Randomisierung, Durchführung Wer führte die Zuteilung durch, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	IVRS
11	Verblindung	
11a	Waren a) die Probanden / Patienten und / oder b) diejenigen, die die Intervention / Behandlung durchführten, und / oder c) diejenigen, die die Zielgrößen beurteilten, verblindet oder nicht verblindet, wie wurde die Verblindung vorgenommen?	<p>a) ja, die Patienten waren verblindet</p> <p>b) ja, der Behandler (sowohl der behandelnde als auch der auswertende Neurologe) war verblindet</p>
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Die Studienmedikation Teriflunomid 7 bzw. 14 mg und Placebo wurden als identisch aussehende weiß bis hellgelbe Filmtabletten in kindersicheren Blisterverpackungen ausgegeben.
12	Statistische Methoden	

12a	<p>Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien</p>	<p>Bewertung des primären Zielkriteriums</p> <p>Jährliche Schubrate (ARR):</p> <p>Die primäre Analyse der jährlichen Schubrate wurde mittels eines Poisson-Regressions-Modells durchgeführt, mit robuster Fehlervarianz, um die potenziell übermäßige Verteilung der Daten angemessen einzubeziehen. Das Modell beinhaltete die absolute Anzahl bestätigter Schübe mit Start zwischen Randomisierung und letzter Dosierung als Response-Variable und 3-Behandlungsarme (Placebo, Teriflunomide 7 mg, Teriflunomid 14 mg), EDSS-Strate (Baseline EDSS $\leq 3,5$ versus $>3,5$), sowie Region als Kovariate. Um die unterschiedlichen Studiendauern der einzelnen Patienten auf angemessene Weise mit einzubeziehen, wurde die log-transformierte standardisierte Studiendauer in dem Modell als Offset-Variable für die entsprechende Berechnung der Schubrate integriert. Die robuste Fehlervarianz wurde durch Spezifizierung des „patient identifier in the repeated statement“ abgeschätzt- unter Verwendung des SAS PROC GENMOD, welches dem Generalized Estimating Equation (GEE) Modell, äquivalent ist. Zweiseitige 95%-Konfidenzintervalle der Rate Ratios werden für den Vergleich jeder aktiven Behandlung gegen Placebo geliefert. Die geschätzte Schubrate und ihr zweiseitiges 95%-Konfidenzintervall und die grobe Schätzung der jährlichen Schubrate sind für jede Behandlungsgruppe angegeben.</p> <p>Der primäre Endpunkt wurde mit dem GEE Modell anstelle des regulären Poisson Modells analysiert, da der GEE Schätzer robust gegen eine Verletzung der Korrelationsstruktur und der Verteilungsannahmen war.</p> <p>Zusätzlich wurden „Behandlung - EDSS“- und „Behandlung- Region“-Interaktionen am 5%-Signifikanzlevel getrennt bewertet, und zwar durch Hinzufügen der korrespondierenden Interaktionsterme zu dem oben angegebenen GEE-Modell. In dem Fall, dass eine statistisch signifikante Interaktion entdeckt werden konnte, wurden weitere Untersuchungen durchgeführt, um die Möglichkeit einer qualitativen Interaktion zu erkennen. Nelson-Aalen „mean cumulative“-Funktionsplots werden zur graphischen Darstellung angegeben.</p> <p>Bewertung der sekundären Zielkriterien</p> <p>Zeit bis zur Behinderungsprogression:</p> <p>Die Zeit bis zur Behinderungsprogression wurde mittels Log-Rank-Test analysiert und zwar mit Zeit bis zur Behinderungsprogression als abhängige Variable, Behandlungsgruppe als Testvariable, sowie Region und Baseline EDSS Strata als Stratifizierungsfaktoren.</p> <p>Die Kaplan-Meier Methode wurde verwendet, um die Zeit bis zur Behinderungsprogression abzuschätzen und zwar spezifisch für jede Gruppe, basierend auf der ITT Population. Kaplan-Meier-Graphen wurden erzeugt, sowie Quartile und Punktwahrscheinlichkeiten berechnet.</p>
-----	--	--

		<p>Schätzungen der Intervalle wurden mit punktwisen 95 % Konfidenzintervallen berechnet.</p> <p>Die Hauptanalyse war die Zeit bis zur Behinderungsprogression bestätigt für mindestens 12 Wochen. Die unterstützenden Analysen wurden auch für eine für mindestens 24 Wochen bestätigte Behinderungsprogression durchgeführt und für die Zeit bis zur Behinderungsprogression für die PP Population.</p> <p>Die Analyse zur Erfassung des Anteils an Patienten, die nach 6 Monaten, 1 Jahr und 2 Jahren keine Behinderungsprogression aufwiesen war die gleiche wie die zur Auswertung der Zeit bis zur Behinderungsprogression.</p> <p>MRT-Parameter:</p> <p>Die Veränderung der Krankheitslast (Burden of Disease = BOD) im Verhältnis zum Baseline-Wert wurde mit einem MMRM (mixed-effect model with repeated Measures)-Modell analysiert. Dafür wurden Kubikwurzel-transformierte Volumendaten eingesetzt. Das Modell beinhaltete Faktoren (fixe Effekte) für Behandlung, EDSS-Stratum, Region, Visite, BehandlungxVisite-Interaktion, Baseline-Wert (Kubikwurzel-transformiert) und Baseline-Visite-Interaktion. Eine unstrukturierte Korrelationsmatrix wurde zur Modellierung der Intra-Patient-Fehler verwendet. Die Parameter wurden über die „restricted maximum likelihood“-Methode mit dem Newton-Raphson Algorithmus geschätzt und der gemeinsame Nenner der Freiheitsgrade wurden über die Satterthwaite Approximation geschätzt.</p> <p>Die Baseline-adjustierten LS-Means Schätzer in Woche 108 nach Behandlungsgruppe sind angegeben, so wie auch die jeweiligen Unterschiede gegenüber Placebo (inklusive Standardfehler und zugehöriger 95 % Konfidenzintervalle). Die statistische Signifikanz der Vergleiche jeder aktiven Behandlung gegenüber Placebo wurde über Student-t-Tests ermittelt.</p> <p>Die Anzahl Gd-anreichernder Läsionen und die Anzahl einzelner aktiver Läsionen per Scan wurden mit dem gleichen Modell analysiert wie für die jährliche Schubrate verwendet. Das Modell beinhaltete die Anzahl Gd-anreichernder T1-Läsionen als Responsevariable und Behandlungsgruppe, EDSS-Strata, Region und Baseline-Anzahl an Gd-anreichernden T1-Läsionen als Kovariate. Um die unterschiedliche Anzahl an Scans, die pro Patient durchgeführt worden sind, adäquat zu berücksichtigen, wurden die log-transformierten Anzahlen der Scans in das Modell als Offset-Variable einbezogen.</p> <p>Das Gesamtvolumen Gd-anreichernder T1-Läsionen pro MRT-Scan wurde mit einem Log-Rank-Test analysiert. Adjustierung auf Kovarianz in Rang und Strata wurde mit einem ANCOVA-Modell durchgeführt. Das Modell beinhaltete klassifizierte Volumina der Gd-anreichernden T1-Läsionen per Scan als Response-Variable and</p>
--	--	---

	<p>klassifizierte Baseline-Volumina der Gd-anreichernden T1-Läsionen, EDSS-Strata und Region als Kovariate.</p> <p>Die Veränderung gegenüber dem Baselinewert der transformierten Volumina der hypointensen post-Gd T1-Läsionen und den T2-Läsionen wurde mit dem gleichen Modell wie die Krankheitslast analysiert.</p> <p>Die Veränderung gegenüber dem Baselinewert anderer MRT Variablen (Atrophie, Volumen der weißen Substanz, Volumen der grauen Substanz und z4 Score) in Woche 108 wurde mit dem gleichen Modell analysiert als die Krankheitslast, jedoch ohne Datentransformation.</p> <p>Fatigue (Fatigue Impact Scale): Die Veränderungen im Vergleich zum Baseline-Wert in Woche 108 wurden im Gesamtwert und in den drei Subskalen analysiert. Dafür wurde das gleiche Modell verwendet wie für die Krankheitslast, jedoch ohne Datentransformation.</p> <p>Lebensqualität (SF-36): Die Veränderungen im Vergleich zum Baseline-Wert für die 8 Skalen (mit einem Range zwischen 0 und 100) und der zwei zusammenfassenden Scores (physische und psychische Dimension) wurden mit dem gleichen Modell analysiert wie die Krankheitslast (nicht transformierte Daten).</p> <p>Lebensqualität (EQ-5D): Für die quantitativen Variablen des EQ-5d (visuelle Analogskala und “Single index utility”) wurde die Veränderung in Bezug zum Baseline-Wert mit dem gleichen Modell analysiert wie die Krankheitslast (nicht transformierte Daten). Die qualitativen Variablen des EQ-5D (5-Dimensionen, 3-Level-System) wurden zusammengefasst nach Arztbesuch, Baseline-Status und Behandlungsgruppe, unter Verwendung der Anzahl und des prozentualen Anteils an Patienten. Der Nenner war die Anzahl an ITT-Patienten für die Behandlungsgruppe, für die eine Erhebung sowohl eines Baseline-Wertes als auch zum jeweiligen Zeitpunkt stattgefunden hatte.</p> <p>Unerwünschte Ereignisse: Die Anzahl und der prozentuale Anteil an Patienten mit unerwünschten Ereignissen während der Behandlungszeit (TEAEs) werden entsprechend den Kategorien system organ class (SOC), high level group term (HLGT), high level term (HLT) und preferred term (PT) präsentiert.</p>
--	---

		<p>Multiple Vergleiche/Multiplizität für das primäre und das wichtigste sekundäre Zielkriterium:</p> <p>Multiplizität kommt ins Spiel, wenn auf eine der folgenden Hypothesen getestet wird:</p> <ul style="list-style-type: none"> • H1: kein Behandlungsunterschied zwischen Teriflunomid 14 mg und Placebo in jährlicher Schubrate • H2: kein Behandlungsunterschied zwischen Teriflunomid 7 mg und Placebo in jährlicher Schubrate • S1: kein Behandlungsunterschied zwischen Teriflunomid 14 mg und Placebo in Behinderungsprogression • S2: kein Behandlungsunterschied zwischen Teriflunomid 7 mg und Placebo in Behinderungsprogression <p>Um diese Multiplizität zu adressieren und die Typ-I-Fehlerrate für diese Familie stark zu kontrollieren, wurde ein „stepdown“ Testverfahren angewendet, um die Hypothesen in der oben spezifizierten Reihenfolge zu testen. Jede Hypothese wurde nur in dem Fall getestet, wenn die vorhergehende am 5% Level abgelehnt wurde.</p> <p>Umgang mit fehlenden Werten:</p> <p>Bei kategorialen Variablen wurden die Patienten mit fehlenden Werten in die Berechnungen von prozentualen Anteilen einbezogen, es sei denn es ist anders ausgewiesen.</p>
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	<p>Primärer Endpunkt (jährliche Schubrate):</p> <p>Subgruppenanalysen wurden durchgeführt nach:</p> <ul style="list-style-type: none"> • Geschlecht • Rasse • Alterskategorie (<38, ≥38; 38 war nah an dem Median des Alters aller randomisierten Patienten) • Region (Osteuropa, Westeuropa, Amerika) • Baseline EDSS-Strata • Anzahl Schübe in den letzten 2 Jahren (0, 1, 2, 3 und ≥4) • MS-Subtyp (RRMS, SPMS, PRMS), so wie von der Studienleitung beschrieben • Vorbehandlung mit DMT (ja, nein) • Anzahl der Baseline Gd-anreichernder Läsionen (0, ≥1) • Baseline Krankheitslast (<13, ≥13, BOD von 13 mL ist nahe an dem Median der Baseline-Krankheitslast aller randomisierten Patienten)

		<p>Um eine Aussage zu treffen, ob ein Beweis für eine Inkonsistenz des Behandlungseffektes zwischen den Subgruppenfaktoren besteht, wurde die GEE Methode mit einem „Behandlung nach Subgruppe“-Interaktionstest angewendet. Jeder Subgruppenfaktor wurde getrennt bewertet. Das Modell beinhaltete Behandlung, EDSS Strata, Region, Subgruppe und „Behandlung- Subgruppe“-Interaktion als fixe Effekte und log-transformierte standardisierte Studiendauern als Offset. In Fällen, in denen der p-Wert kleiner als 0,05 war, wurden weitere Untersuchungen durchgeführt, um eine mögliche qualitative Interaktion festzustellen.</p> <p>Zusätzlich werden zusammenfassende Statistiken der Schubraten für jede Subgruppe geliefert.</p> <p>Sekundärer Endpunkt („Zeit bis zur Behinderungsprogression“):</p> <p>Subgruppenanalysen wurden für das wesentliche sekundäre Zielkriterium für die gleichen Subgruppen durchgeführt wie für das primäre Zielkriterium.</p>
-	Resultate	
13	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	
13a	Anzahl der Studienteilnehmer für jede durch Randomisierung gebildete Behandlungsgruppe, die a) randomisiert wurden, b) tatsächlich die geplante Behandlung/Intervention erhalten haben, c) in der Analyse des primären Zielkriteriums berücksichtigt wurden	<p>Teriflunomid 7 mg</p> <p>a) randomisiert waren 366 Patienten b) 365 Patienten erhielten die geplante Intervention c) 365 Patienten wurden in der Analyse des primären Zielkriterium berücksichtigt</p> <p>Teriflunomid 14 mg</p> <p>a) randomisiert waren 359 Patienten b) 358 Patienten erhielten die geplante Intervention c) 358 Patienten wurden in der Analyse des primären Zielkriterium berücksichtigt</p> <p>Placebo</p> <p>a) randomisiert waren 363 Patienten b) 363 Patienten erhielten die geplante Intervention c) 363 Patienten wurden in der Analyse des primären Zielkriterium berücksichtigt</p>
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	<p>Teriflunomid 7 mg</p> <p>91 Patienten beendeten die Studienbehandlung: 37 aufgrund von Nebenwirkungen 14 aufgrund Wahrnehmung geringer Wirksamkeit 4 aufgrund von Krankheitsprogression 2 aufgrund von Protokollverletzung 4 waren „lost to follow-up“</p>

		<p>32 widerriefen ihr Einverständnis 2 hatten andere Gründe</p> <p>Teriflunomid 14 mg 95 Patienten beendeten die Studienbehandlung: 38 aufgrund von Nebenwirkungen 17 aufgrund Wahrnehmung geringer Wirksamkeit 2 aufgrund von Krankheitsprogression 5 aufgrund von Protokollverletzung 2 waren „lost to follow-up“ 26 widerriefen ihr Einverständnis 6 hatten andere Gründe</p> <p>Placebo 104 Patienten beendeten die Studienbehandlung: 29 aufgrund von Nebenwirkungen 24 aufgrund Wahrnehmung geringer Wirksamkeit 11 aufgrund von Krankheitsprogression 3 aufgrund von Protokollverletzung 4 waren „lost to follow-up“ 33 widerriefen ihr Einverständnis</p>
14	Aufnahme / Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	<p>Aufnahme des ersten Patienten: 24 September 2004 Studienende des letzten Patienten: 08 Juli 2010 (inclusive 16 Wochen Wash-out-Periode) 4 Wochen Screeningphase wurden gefolgt von 108 Wochen Studienbehandlung, abschließend erfolgte ein 16 Wochen Wash-out-Verfahren, so dass die Gesamtbeobachtungsdauer 128 Wochen umfasste.</p>
14b	Informationen, warum die Studie endete oder beendet wurde	Die Studie wurde planmäßig beendet.
a: nach CONSORT 2010.		

Stellen Sie den Patientenfluss in einem Flow-Chart je Studie dar.
Flowchart für die Studie TEMSO

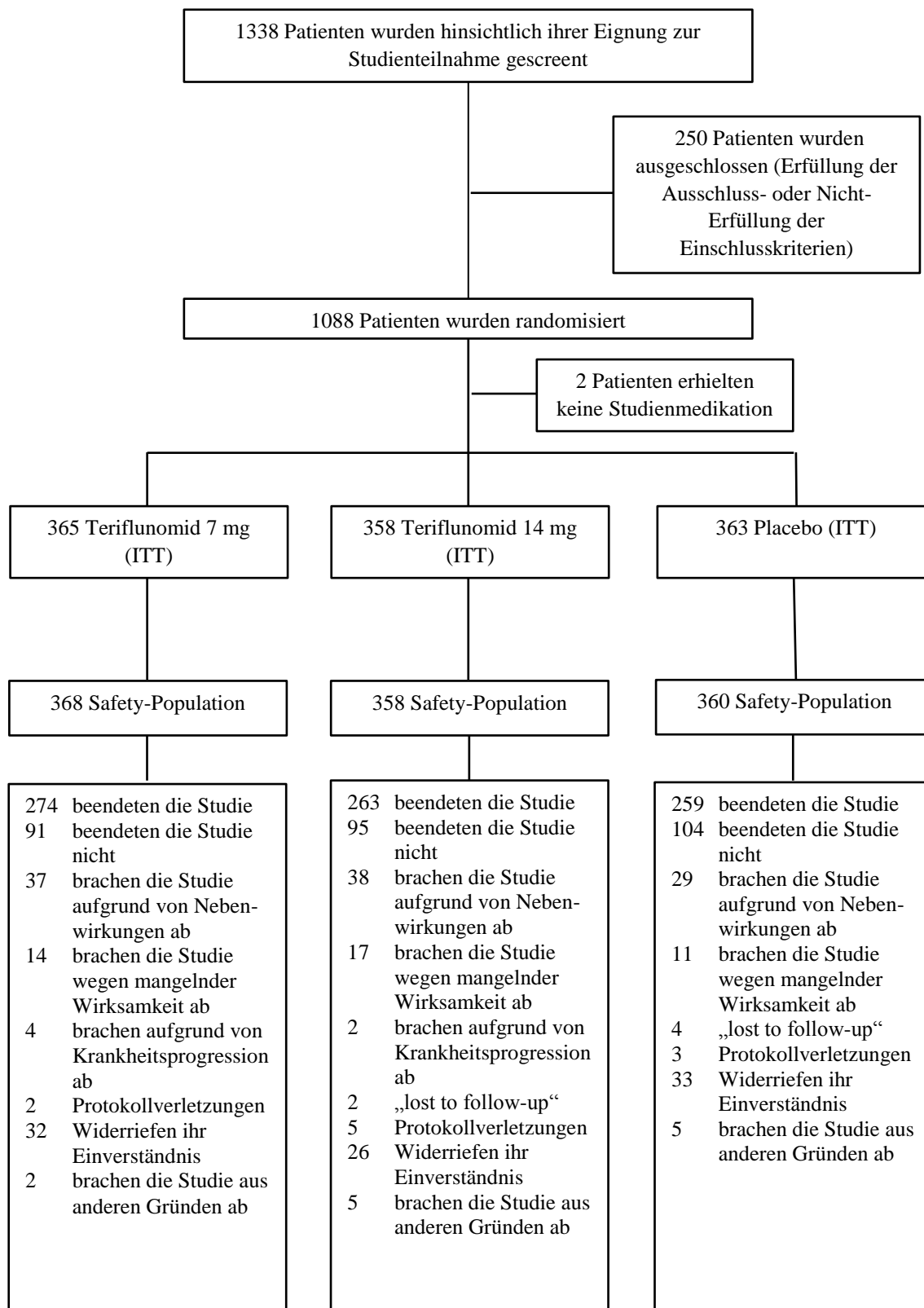


Tabelle 4-208 (Anhang): Studiendesign und -methodik für Studie TOWER

Item ^a	Charakteristikum	Studieninformation
-	Studienziel	
2 b	Genaue Ziele, Fragestellung und Hypothesen	Ziel der Studie war die Erfassung der Effektivität von 2 Dosen Teriflunomid (7 mg und 14 mg) in der Behandlung der schubförmigen MS, hauptsächlich in Bezug auf das gut etablierte Maß zur Messung der Krankheitsaktivität, der Schubrate.
-	Methoden	
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	Es handelt sich um eine multizentrische, multinationale, randomisierte, doppelblinde placebo-kontrollierte Parallelgruppenstudie (3 Gruppen). Es handelt sich um eine Phase 3 Studie. Patienten wurden zufällig einer der drei Gruppen 7 mg Teriflunomid, 14 mg Teriflunomid und Placebo im Verhältnis 1:1:1 zugeteilt.
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	Das Originalprotokoll beinhaltete eine vierte Behandlungsgruppe (open-label Interferon-beta) als Kalibratorarm, mit einer vorgesehenen 2:2:2:1 Randomisierung. Diese Behandlungsgruppe wurde vor Einschluss von Patienten in die Studie wieder gelöscht. Zum Studienprotokoll wurden 5 Amendments erstellt. Folgende wesentliche Änderungen ergaben sich mit den Amendments (für weitere Details, s. Studienbericht): Amendment 1: <ul style="list-style-type: none"> • Löschen des Interferon-Kalibratorarms aus dem Studiendesign • Entfernen folgender sekundärer Zielkriterien: Hospital Anxiety and Depression Scale und Suicidality Tracking Scale • Änderung der Kodierung der AEs in 4 Level: Preferred Term (PT), High Level Group Term (HLGT), HighLevel Term (HLT) und primary System Organ Class (SOC) anstatt in 2 Level: PT and SOC Amendment 2: <ul style="list-style-type: none"> • Implementierung eines Lungenfunktionstests für ein Subset an Patienten, um die Dokumentation des Sicherheitsprofils in Bezug auf Lungenerkrankungen zu unterstützen Amendment 4: <ul style="list-style-type: none"> • Änderung des Zeitpunktes für die Bestätigung der Behinderungsprogression von 24 auf 12 Wochen, um die Analyse der Studie EFC6049

Item ^a	Charakteristikum	Studieninformation
		<p>anpassen zu können</p> <ul style="list-style-type: none"> • Hinzufügen anderer sekundärer Zielkriterien: Zeit bis zum ersten bestätigten Schub; Anteil an Patienten ohne Schub; Anteil Patienten ohne Behinderungsprogression zu den Zeitpunkten 6 Monate, 1 Jahr, 2 Jahre; Change from Baseline in EDSS • Hinzufügen der unterstützenden Analysen der PP Population für die Endpunkte jährliche Schubrate und Zeit bis zur Behinderungsprogression • Hinzufügen einer Interimanalyse, um für regulatorische Zwecke mehr Daten zum Nutzen-Risiko-Verhältnis vorliegen zu haben • Hinzufügen von peripherer Neuropathie bestätigt durch elektrophysiologische Tests als einen „alert term“, um eine bessere Dokumentation dieser Fälle zu erreichen
4	Probanden / Patienten	
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	<p>Einschlusskriterien:</p> <ul style="list-style-type: none"> • Schubförmige Formen der MS (erfüllen der McDonald's Kriterien (2005 Revision)) • EDSS score ≤ 5.5 zum Screening-Zeitpunkt • mindestens 1 Schub in den 12 Monaten vor Randomisierung oder mindestens 2 Schübe in den 24 Monaten vor Randomisierung • Schriftliche Zustimmung zur Studienteilnahme muss vorliegen (nach ausführlicher Information über alternative Behandlungsmöglichkeiten) <p>Ausschlusskriterien:</p> <ul style="list-style-type: none"> • <18 Jahre oder ≥ 56 Jahre alt zum Randomisierungszeitpunkt • Kein Einverständnis zum HIV Test • Schub in den letzten 30 Tagen vor Randomisierung • Mentaler Zustand, der es dem Patienten unmöglich machte die Art, den Umfang und mögliche Konsequenzen der Studie zu verstehen • Patienten, die sich wahrscheinlich nicht an das Studienprotokoll halten werden (unkooperative Einstellung, fehlende Möglichkeit Folgeuntersuchungen wahrzunehmen, bekannte Wahrscheinlichkeit, dass die Studie nicht beendet wird) • Klinisch relevante kardiovaskuläre, neurologische, endokrine oder andere

Item ^a	Charakteristikum	Studieninformation
		<p data-bbox="871 277 1410 456">schwerwiegende systemische Erkrankung, die eine Anwendung des Studienprotokolls bzw. Auswertung der Studienergebnisse unmöglich gemacht hätte oder die den Patienten durch Studienteilnahme einem Risiko ausgesetzt hätten.</p> <ul data-bbox="842 479 1410 1982" style="list-style-type: none"> <li data-bbox="842 479 1410 748">• Patienten mit angeborener oder erworbener schwerer Immunschwäche, einer Krebsvorgeschichte (mit Ausnahme von Basalzell- und Plattenepithelkarzinom, die operativ entfernt wurden und für die keine Metastase nachgewiesen werden konnte), lymphoproliferative Erkrankungen oder ein Patient, der eine Strahlentherapie des lymphatischen Systems erhalten hat. <li data-bbox="842 770 1410 801">• HIV-positive Patienten <li data-bbox="842 824 1410 904">• Bekannte Vorgeschichte von aktiver Tuberkulose, die nicht angemessen behandelt wurde <li data-bbox="842 927 1410 1016">• Hypoproteinämie (z. B. im Fall von schwerer Lebererkrankung oder nephrotischem Syndrom) mit einem Serumalbuminwert <3,0 g/dL <li data-bbox="842 1039 1410 1128">• Moderate bis schwere Einschränkung der renalen Funktion, gezeigt durch einen Serum-Kreatinin-Wert von >133 µmol/L (oder >1.5 mg/dL) <li data-bbox="842 1151 1410 1509">• Patienten mit signifikant eingeschränkter Knochenmarksfunktion oder signifikanter Anämie, Leukopenie oder einem Hämatokritwert <24% und/oder <ul data-bbox="900 1285 1410 1509" style="list-style-type: none"> <li data-bbox="900 1285 1410 1352">– Absolute Zahl der weißen Blutzellen <4000 Zellen/mm³ und/oder <li data-bbox="900 1375 1410 1442">– Plättchenzahl <150000 Zellen/mm³ und/oder <li data-bbox="900 1464 1410 1509">– Absolute Neutrophilenzahl ≤1500 Zellen/mm³ <li data-bbox="842 1532 1410 1585">• Persistierende signifikante oder schwere Infektion <li data-bbox="842 1608 1410 1756">• Leberfunktionsstörungen oder persistierende Erhöhungen (durch wiederholten Test bestätigt) der Werte für ALT, AST oder direktem Bilirubin auf einen Wert, der das 1,5fache des Normalwertes übersteigt. <li data-bbox="842 1778 1410 1809">• Bekannte Hepatitis-Vorgeschichte <li data-bbox="842 1832 1410 1886">• Vorgeschichte von Drogen- oder Alkoholmissbrauch <li data-bbox="842 1908 1410 1982">• Patient gehört zum Personal oder ist verwandt mit Personal (z. B. Studienleiter, Forschungsassistent, Apotheker, ...), das direkt in die

Item ^a	Charakteristikum	Studieninformation
		<p>Durchführung der Studie eingebunden ist</p> <ul style="list-style-type: none"> • Anwendung von Adrenocorticotropem Hormon (ACTH) oder systemischer Kortikosteroide innerhalb der letzten zwei Wochen vor Randomisierung • Vorherige oder gleichzeitige Anwendung einer Zytokintherapie, Glatirameracetat oder intravenöser Immunglobuline in den letzten 3 Monaten vor Randomisierung • Vorherige oder gleichzeitige Anwendung von Natalizumab (Tysabri®) • Vorherige oder gleichzeitige Anwendung von Natalizumab • Vorherige oder gleichzeitige Anwendung von Cladribin, Mitoxantron oder anderer immunsuppressiver Agentien (z. B. Azathioprin, Cyclophosphamid, Cyclosporin, Methotrexat oder Mycophenolat) • Vorherige Behandlung mit Teriflunomid und vorherige oder gleichzeitige Anwendung von Leflunomide (ARAVA®) oder Überempfindlichkeit gegenüber einem der anderen Inhalts- oder Hilfsstoffe der Studienmedikamentation • Schwangere oder stillende Frauen • Gebärfähige Frauen oder Männer, die keine effektiven Verhütungsmethoden anwenden und/oder gebärfähige Frauen, die nicht auf eine Schwangerschaft getestet werden wollen oder können • Patienten, die im Laufe der Studiendauer Eltern werden wollen • Vorherige Einnahme irgendeiner Studienmedikamentation in den vorhergehenden 6 Monaten • Vorherige oder gleichzeitige Anwendung von Interferonen in den 3 Monaten vor Randomisierung • Bekannte Überempfindlichkeit gegenüber Teriflunomid oder Leflunomid in der Vorgeschichte • Persistierende Erhöhung (durch wiederholten Test bestätigt) der Serumamylase oder -lipase auf einen Wert der das 2fache des oberen Normalwertes übersteigt • Bekannte Vorgeschichte chronischer pankreatischer Erkrankung oder Pankreatitis. • Vorherige Anwendung (innerhalb der letzten 4

Item ^a	Charakteristikum	Studieninformation
		Wochen vor Randomisierung) von Cholestyramin
4b	Studienorganisation und Ort der Studiendurchführung	Es handelt sich um eine multizentrische Studie, die in 189 aktive Zentren in 26 Ländern durchgeführt wurde: Australien, Österreich, Weißrussland, Belgien, Kanada, Chile, China, Tschechien, Estland, Frankreich, Deutschland, Griechenland, Mexiko, Niederlande, Philippinen, Polen, Rumänien, Slowakei, Spanien, Schweden, Thailand, Tunesien, Türkei, Ukraine, Großbritannien, USA
5	Interventionen Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	Teriflunomid 7 mg: Filmtabletten, 1x täglich morgens, oral (mit Wasser) Teriflunomid 14 mg: Filmtabletten, 1x täglich morgens, oral (mit Wasser) Placebo: Filmtabletten, 1x täglich morgens, oral (mit Wasser)
6	Zielkriterien	
6a	Klar definierte primäre und sekundäre Zielkriterien, Erhebungszeitpunkte, ggf. alle zur Optimierung der Ergebnisqualität verwendeten Erhebungsmethoden (z. B. Mehrfachbeobachtungen, Training der Prüfer) und ggf. Angaben zur Validierung von Erhebungsinstrumenten	Primäres Zielkriterium (jährliche Schubrate, ARR): Das primäre Zielkriterium war die Erfassung des Effektes zweier Dosen Teriflunomid im Vergleich zu Placebo auf die Schubfrequenz bei Patienten mit schubförmigen Formen der MS. Sekundäre Zielkriterien: Zeit bis zur Behinderungsprogression: Das wichtigste sekundäre Zielkriterium war die Erfassung des Effektes zweier Dosen Teriflunomid im Vergleich zu Placebo auf die Behinderungsprogression bei Patienten mit schubförmigen Formen der Multiplen Sklerose. Der EDSS/FS wurde routinemäßig zum Randomisierungszeitpunkt, in Woche 12, 24, dann alle 12 Wochen und zum Behandlungsende bestimmt. Zusätzlich war eine EDSS/FS-Bewertung im Falle eines Schubes erforderlich. Andere Zielkriterien waren: Die Erfassung des Effektes zweier Dosen Teriflunomid im Vergleich zu Placebo auf: <ul style="list-style-type: none"> • Fatigue (FIS) Erhebungszeitpunkte waren der Zeitpunkt der Randomisierung, Woche 12, 24, dann alle 24 Wochen und zum Behandlungsende. • Gesundheitsbezogene Lebensqualität (SF-36) Erhebungszeitpunkte waren der Zeitpunkt der Randomisierung, Woche 12, 24, dann alle 24

Item ^a	Charakteristikum	Studieninformation
		<p>Wochen und zum Behandlungsende.</p> <ul style="list-style-type: none"> • Evaluierung der Sicherheit und Verträglichkeit von Teriflunomid. <p>Entsprechende „Adverse event reportings“ erfolgten in Woche 2 und Woche 6 und dann alle 6 Wochen bis Woche 24, dann in Woche 36 und ab diesem Termin alle 12 Wochen bis zum Studienende.</p> <ul style="list-style-type: none"> • Optional ein pharmakogenomischer Test mit dem Ziel mögliche Zusammenhänge zwischen dem enzymatischen System, das hauptsächlich für die Teriflunomid-Metabolisierung zuständig ist, und Genvariationen und klinische Folgen aufzudecken
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	<p>Amendment 4 (19. Januar 2011):</p> <ul style="list-style-type: none"> • Änderung des Zeitpunktes für die Bestätigung der Behinderungsprogression von 24 auf 12 Wochen, um die Analyse der Studie EFC6049 (TEMSSO) anpassen zu können • Hinzufügen anderer sekundärer Zielkriterien: Zeit bis zum ersten bestätigten Schub; Anteil an Patienten ohne Schub; Anteil Patienten ohne Behinderungsprogression zu den Zeitpunkten 6 Monate, 1 Jahr, 2 Jahre; Change from Baseline in EDSS <p>Die Änderungen der Zielkriterien erfolgten, um in besserer Übereinstimmung mit der anderen wesentlichen Phase 3 Studie EFC6049 (TEMSSO) zu sein, was eine spätere Metaanalyse der Studienergebnisse ermöglicht.</p>
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	<p>Die Berechnung der Fallzahl basierte auf einer 1:1:1 Randomisierung für 14 mg Teriflunomid, 7 mg Teriflunomid und Placebo sowie dem primären Zielkriterium ARR und dem wesentlichen sekundären Zielkriterium Zeit bis zur anhaltenden Behinderungsprogression, mit den folgenden Annahmen:</p> <ul style="list-style-type: none"> • Jährliche Schubrate für Placebo 0,74 (basierend auf aktuell erhältlichen MS-Daten, wo die Placebo 2-Jahres-Schubrate auf 1.48 geschätzt wurde) • Eine 25%ige Risikoreduktion in Bezug auf die jährliche Schubrate (d. h. eine ARR von 0,55 für Teriflunomid) • Anzahl der Schübe folgt ungefähr einer Poisson-Verteilung mit einem „over dispersion parameter“ von 1,3 (geschätzt anhand von aktuell erhältlichen Daten aus Tysabri® Studien)

Item ^a	Charakteristikum	Studieninformation
		<p>and Protokoll 2001, LTS6048)</p> <ul style="list-style-type: none"> • Eine 1,5 Jahre dauernde Rekrutierungsphase mit einer linearen Rekrutierungsrate, so dass die durchschnittliche Expositionsdauer für laufende Patienten 1,75 Jahre beträgt • Ein zweiseitiges 5%iges Signifikanzlevel • Erwartete Drop-out-Rate von 20 % <p>Basierend auf den genannten Annahmen, waren für die Studie insgesamt 1110 Patienten (370 pro Behandlungsgruppe) notwendig, um eine 94 % ige Power zu haben eine 25 % ige Reduktion des Risikos bezüglich der jährlichen Schubrate zu detektieren.</p> <p>Die Berechnungen wurden mit intern entwickelten statistischen Programmen durchgeführt.</p> <p>Zusätzlich hat die Studie eine Power von 75 %, um eine 37 %ige Reduktion des Hazards in Bezug auf die Zeit bis zur Behinderungsprogression zu erfassen (Log Rang Test).</p> <p>Die Fallzahl und Powerschätzung wurde mit dem Programm EAST 4.0 berechnet mit den Annahmen einer Hazard-Rate von 0,1783 in der Placebogruppe und 0,1116 in der Teriflunomid-Gruppe (d. h. einer 30 % igen Wahrscheinlichkeit einer Behinderungsprogression für Placebo-Patienten nah zwei Jahren und einer entsprechenden Wahrscheinlichkeit von 20 % für Teriflunomid-Patienten. Die Fallzahl war auch für eine 20% ige Drop-out-Quote adjustiert.</p>
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	<p>Um die Sicherheit der Patienten, die in der Studie EFC10531/TOWER Teriflunomid ausgesetzt sind, zu gewährleisten, wurden die erhobenen Sicherheitsdaten durch ein unabhängiges Data Monitoring Committee (DMC) durchgesehen und bewertet. Dieses DMC war verantwortlich dafür, angemessene Empfehlungen bezüglich der Studiendurchführung an den Sponsor abzugeben.</p> <p>Eine ursprünglich nicht geplante Interimanalyse zu allen randomisierten Patienten wurde zum 30 November 2010 durchgeführt (1080 or 360 per treatment group). Patientendaten bis zum 28. Februar 2011 wurden in der Analyse berücksichtigt, so dass der letzte Patient, der eingeschlossen wurde, als Minimum ca. 3 Monate Follow-up hatte, die erwartete durchschnittliche Behandlungsdauer war 1 Jahr. Eine Interimanalyse basierend auf 1096 Patienten folgte für den Zeitpunkt 3 Monate nachdem der letzte Patient randomisiert war, um zusätzliche Informationen über das Sicherheitsprofil von Teriflunomid zu erlangen und um unterstützende Daten zum vorteilhaften Effekt auf die Schubrate für regulatorische Zwecke vorliegen zu haben.</p> <p>Ein Studienstopp auf Basis der Interim-</p>

Item ^a	Charakteristikum	Studieninformation
		Effektivitätsdaten war nicht geplant. Die Analyse hatte lediglich beschreibenden Charakter und bestand aus Schätzungen (95 %iges Konfidenzintervall) der jährlichen Schubrate innerhalb jeder Behandlungsgruppe in der Interimanalyse-Population und innerhalb vordefinierter Subgruppen. Kein formales statistisches Testen wurde durchgeführt und so gab es auch keine Anpassung des Alpha für die finale Studienanalyse. Die Analyse wurde durch eine unabhängige Gruppe an Statistikern durchgeführt, die nicht in die Studie eingebunden waren und keine Kenntnis über das Teriflunomid Studienteam hatten bis zum „final lock“ der kompletten Studiendaten.
8	Randomisierung, Erzeugung der Behandlungsfolge	
8a	Methode zur Generierung der zufälligen Zuteilung	Die Patienten wurden zentral über ein Interactive Voice Response System (IVRS) randomisiert und zufällig 1:1:1 in die drei Behandlungsgruppen 7 mg Teriflunomid, 14 mg Teriflunomid und Placebo aufgeteilt.
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	Die Randomisierung wurde über eine permutierte Blockrandomisierung erreicht. Die Randomisierung wurde hinsichtlich Studienort und Baseline EDSS-Wert stratifiziert.
9	Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment) Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war	Die Studienmedikation Teriflunomid 7 bzw. 14 mg und Placebo wurden als identisch aussehende weiß bis hellgelbe Filmtabletten in kindersicheren Blisterverpackungen ausgegeben. Passende Placebo-Tabletten wurden ebenfalls vom Sponsor bereitgestellt. Die Studienmedikation war in nummerierten Behältern abgepackt (medication kits). Das IVRS ordnete eine Medication Kit Nummer einem Patienten zu, das dann von dem Koordinator des entsprechenden Studienortes dokumentiert und dem Patienten ausgehändigt wurde. Die Geheimhaltung war bis zur Zuteilung gewährleistet.
10	Randomisierung, Durchführung Wer führte die Zuteilung durch, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	Die zufällige Zuteilung erfolgte über ein interaktives Spracherkennungssystem (interactive voice response system, IVRS).
11	Verblindung	
11a	Waren a) die Probanden / Patienten und / oder b) diejenigen, die die Intervention / Behandlung durchführten, und / oder c) diejenigen, die die Zielgrößen beurteilten, verblindet oder nicht verblindet, wie wurde	a) ja, die Patienten waren verblindet b) ja, die Behandler waren auch verblindet

Item ^a	Charakteristikum	Studieninformation
	die Verblindung vorgenommen?	
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Die Studienmedikation Teriflunomid 7 bzw. 14 mg und Placebo wurden als identisch aussehende weiß bis hellgelbe Filmtabletten in kindersicheren Blisterverpackungen ausgegeben. Passende Placebo-Tabletten wurden ebenfalls vom Sponsor bereitgestellt.
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	<p>Analyse des primären Zielkriteriums (jährliche Schubrate, ARR):</p> <p>Die primäre Analyse der jährlichen Schubrate wurde mittels eines Poisson-Regressions-Modells durchgeführt, mit robuster Fehlervarianz, um die potenziell übermäßige Verteilung der Daten angemessen einzubeziehen. Das Modell beinhaltete die absolute Anzahl bestätigter Schübe mit Start zwischen Randomisierung und letzter Dosierung als Response-Variable und 3-Behandlungsarme (Placebo, Teriflunomide 7 mg, Teriflunomid 14 mg), EDSS-Strate (Baseline EDSS $\leq 3,5$ versus $> 3,5$), sowie Region als Kovariate. Um die unterschiedlichen Studiendauern der einzelnen Patienten auf angemessene Weise mit einzubeziehen, wurde die log-transformierte standardisierte Studiendauer in dem Modell als Offset-Variable für die entsprechende Berechnung der Schubrate integriert. Die robuste Fehlervarianz wurde durch Spezifizierung des „patient identifier in the repeated statement“ abgeschätzt- unter Verwendung des SAS PROC GENMOD, welches dem Generalized Estimating Equation (GEE) Modell, äquivalent ist.</p> <p>Zweiseitige 95%-Konfidenzintervalle der Rate Ratios werden für den Vergleich jeder aktiven Behandlung gegen Placebo geliefert. Die geschätzte Schubrate und ihr zweiseitiges 95%-Konfidenzintervall und die grobe Schätzung der jährlichen Schubrate sind für jede Behandlungsgruppe angegeben.</p> <p>Der primäre Endpunkt wurde mit dem GEE Modell anstelle des regulären Poisson Modells analysiert, da der GEE Schätzer gegen eine Verletzung der Korrelationsstruktur und der Verteilungsannahmen robust war.</p> <p>Zusätzlich wurden „Behandlung - EDSS“- und „Behandlung- Region“-Interaktionen am 5%-Signifikanzlevel getrennt bewertet, und zwar durch Hinzufügen der korrespondierenden Interaktionsterme zu dem oben angegebenen GEE-Modell. In dem Fall, dass eine statistisch signifikante Interaktion entdeckt werden konnte, wurden weitere Untersuchungen durchgeführt, um die Möglichkeit einer qualitativen Interaktion zu erkennen. Nelson-Aalen „mean cumulative“ Funktionsplots werden zur</p>

Item ^a	Charakteristikum	Studieninformation
		<p>graphischen Darstellung angeben.</p> <p>Analyse der sekundären Zielkriterien: Zeit bis zur Behinderungsprogression:</p> <p>Die Zeit bis zur Behinderungsprogression wurde mittels Log-Rang-Test analysiert und zwar mit Zeit bis zur Behinderungsprogression als abhängige Variable, Behandlungsgruppe als Testvariable, sowie Region und Baseline-EDSS Strata als Stratifizierungsfaktoren. Die Schätzungen der Hazard Ratios für jede Teriflunomid-Behandlungsgruppe gegenüber Placebo wurden mittels eines Cox-Regressionsmodells geschätzt, und zwar mit Region und Baseline EDSS Strata als Kovariate. Für die Schätzung der gruppenspezifischen Rate der Zeit bis zur Behinderung wurde die Kaplan-Meier-Methode verwendet, und zwar für die ITT Population. Kaplan-Meier-Graphen wurden erzeugt und die jeweiligen Quartile und Punktwahrscheinlichkeiten berechnet. Intervallschätzungen wurden unter Verwendung 95%iger punktweiser Konfidenzintervalle berechnet.</p> <p>Die hauptsächliche Analyse war die Zeit bis zur Behinderungsprogression, die für mindestens 12 Wochen bestätigt werden konnte. Die unterstützenden Analysen wurden auch für die Zeit bis zur Behinderungsprogression, die für 24 Wochen bestätigt werden konnte, durchgeführt, basierend auf der PP Population.</p> <p>Multiple Vergleiche/Multiplizität für das primäre und das wichtigste sekundäre Zielkriterium:</p> <p>Multiplizität kommt ins Spiel, wenn auf eine der folgenden Hypothesen getestet wird:</p> <ul style="list-style-type: none"> • H1: kein Behandlungsunterschied zwischen Teriflunomid 14 mg und Placebo in jährlicher Schubrate • H2: kein Behandlungsunterschied zwischen Teriflunomid 7 mg und Placebo in jährlicher Schubrate • S1: kein Behandlungsunterschied zwischen Teriflunomid 14 mg und Placebo in Behinderungsprogression • S2: kein Behandlungsunterschied zwischen Teriflunomid 7 mg und Placebo in Behinderungsprogression <p>Um diese Multiplizität zu adressieren und die Typ-I-Fehlerrate für diese Familie stark zu kontrollieren, wurde ein „stepdown“ Testverfahren angewendet, um die Hypothesen in der oben spezifizierten Reihenfolge zu testen. Jede Hypothese wurde nur in dem Fall getestet, wenn die vorhergehende am 5%</p>

Item ^a	Charakteristikum	Studieninformation
		<p>Level abgelehnt wurde.</p> <p>Andere sekundäre Zielkriterien:</p> <p>Fatigue Impact Scale: Gesamtscore und Bereichscores (kognitive, physische und psychosoziale Funktion) „Change from baseline“ in FIS wurde mit einem „mixed-effect model using repeated measures“ (MMRM) mit Faktoren (fixed effects) für Behandlung, EDSS Strata, Region, Arztbesuch, Behandlung-Arztbesuch-Interaktion, Baseline-Wert und Baseline-Arztbesuch-Interaktion. Statistische Interferenzen wurden für „change from baseline“ in Woche 48 abgeleitet. Zusätzlich wurde ein ANCOVA (analysis of covariance) Modell mit Faktoren für Behandlungsgruppe, Arztbesuchsnummer des letzten Arztbesuchs, Baseline EDSS Strata, Region und FIS Baseline-Score für Change from Baseline zur letzten Arztbesuch durchgeführt.</p> <p>Zeit bis zum ersten Schub: Die Zeit bis zum ersten Schub wurde mit winem Log-Rang-Test analysiert mit Zeit bis zum ersten Schub als abhängige Variable, Behandlungsgruppe als Testvariable und Region und Baseline EDSS Strata als Kovariate. Hazard Ratios wurden mit einem Cox-Regressions-Modell geschätzt und zwar mit Behandlungsgruppe, Region und Baseline EDSS Strata als Kovariate.</p> <p>Anteil an Patienten ohne Schub: Der Anteil an Patienten ohne Schub wurde mit der Kaplan-Meier-Methode abgeschätzt. Kaplan-Meier-Graphen wurden erzeugt, Quartile und Punktwahrscheinlichkeiten wurden berechnet. Intervallschätzungen wurden unter Verwendung 95%iger punktweiser Konfidenzintervalle berechnet.</p> <p>Anteil an Patienten ohne Behinderungsprogression: (zu den Zeitpunkten 6 Monate, 1 Jahr und 2 Jahre) Der Anteil an Patienten ohne Behinderungsprogression wurde basierend auf der Kaplan-Meier-Kurve (entsprechend dem wichtigsten sekundären Zielkriterium) geschätzt.</p> <p>Veränderungen gegenüber dem Baseline-Wert in EDSS-Scores: Veränderungen im Vergleich zum Baseline-Wert zum Zeitpunkt Woche 48 wurde mit dem gleichen Modell wie für FIS verwendet analysiert. Lediglich der EDSS Score von geplanten Untersuchungs-</p>

Item ^a	Charakteristikum	Studieninformation
		<p>zeitpunkten wurde für die Analyse verwendet, weil so die temporären Fluktuationen im klinischen Status, die mit einem Schub einhergehen können, ausgeschlossen werden können.</p> <p>Fatigue (FIS) und Lebensqualität (SF-36): Für kontinuierliche sekundäre Endpunkte so wie FIS, SF-36 und EDSS change from baseline, konzentrierte sich die statistische Auswertung auf Woche 48 so wie es dem Studiendesign entsprach. Der statistische Test wurde ebenfalls durchgeführt unter Verwendung der Bewertungen zur letzten Behandlung. Beschreibende Zusammenfassungen nach Arztbesuch wurden geliefert nach Woche 48, weil die Daten für Patienten, die früh in die Studie aufgenommen worden waren, vorhanden waren.</p> <p>Für FIS und SF-36 wurden die Werte für einige Patienten alle 12 Wochen nach den ersten 24 Wochen zusätzlich zu den Messungen alle 24 Wochen gesammelt- bedingt durch einen Datenerhebungsfehler. Sowohl in MMRM als auch in ANCOVA wurden nur die im Protokoll geplanten Arztbesuche eingeschlossen, d. h. Woche 12, 24 und alle 24 Wochen danach (48, 72, 96,...). In der beschreibenden Zusammenfassung der Statistik, wurden alle vorhandenen Daten eingeschlossen.</p>
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	<p>Weitere Analysen für den primären Endpunkt jährliche Schubrate: Die Analyse der jährlichen Schubrate wurde außerdem auch für die PP Population als unterstützende Analyse durchgeführt.</p> <p>Eine unterstützende Analyse für die jährliche Schubrate wurde mit einem negativen binomialen Modell durchgeführt, mit den gleichen Modellspezifizierungen wie in dem oben genannten Poisson-Regressions-Modell.</p> <p>Die primäre Analyse beinhaltet bestätigte Schübe bis zum Ende der Studienbehandlung. Das Zufügen eines klinischen Follow-ups aller Patienten, die vorzeitig und permanent die Studienmedikation abbrechen, wurde implementiert. Unter Verwendung der zusätzlichen Daten aus der Follow-up-Periode wurde eine Sensitivitätsanalyse für die jährliche Schubrate durchgeführt, um die Ergebnisse der primären Analyse zu untermauern. In dieser Sensitivitätsanalyse wurden bestätigte Schübe über die gesamte Studiendauer inklusive Follow-up Periode analysiert, und zwar mit dem gleichen statistischen Ansatz wie bereits in der primären Analyse. Für diese Sensitivitätsanalyse wurde die Gesamtanzahl bestätigter Schübe und die standardisierte absolute Studiendauer wie folgt</p>

Item ^a	Charakteristikum	Studieninformation
		<p>berechnet:</p> <p>Die bestätigten Schübe, die zwischen Randomisierung und letztem Follow-up Datum der Studie aufgetreten sind, sind als die Gesamtzahl bestätigter Schübe für die Sensitivitätsanalyse definiert. Wenn Patienten in die Langzeitverlängerung eingetreten sind, wurde der letzte Follow-up Zeitpunkt als Tag der Randomisierung für die Erweiterungsstudie definiert.</p> <p>Der letzte Follow-Up Zeitpunkt war der letzte der folgenden 3 Daten: Beginn des letzten Schubes, letzte Untersuchung oder letzter vorhandener Kontakt im Follow-Up Zeitraum.</p> <p>Die standardisierte absolute Studiendauer (in Jahren) für die Sensitivitätsanalyse wurde wie folgt berechnet:</p> $(\text{last follow-up date} - \text{randomization date} + 1) / 365.25$ <p>Subgruppenanalysen:</p> <p>Subgruppenanalysen für das primäre Zielkriterium wurden durchgeführt für:</p> <ul style="list-style-type: none"> • Geschlecht • Rasse • Altersgruppe (<38, ≥38; 38 war nahe an dem Median aller randomisierter Patienten) • Region (Osteuropa, Westeuropa, Afrika, Amerika, Asien und Australien) • Baseline EDSS Wert • Anzahl an Schüben während des letzten Jahres (≤1, 2, und ≥3) • Anzahl an Schüben während der letzten 2 Jahre (≤1, 2, 3, and ≥4) • MS Subtyp (relapsing remitting MS, secondary progressive MS, or progressive relapsing) wie durch den Investigator beschrieben • Vorbehandlung mit einem krankheitsmodifizierenden MS-Therapeutikum vor Studienbeginn (Ja, Nein) • Gebiet (US, nicht-US) <p>Um zu bewerten, ob ein Beweis für Inkonsistenz des Behandlungseffektes zwischen den Subgruppen vorlag, wurde die GEE Methode angewendet, unter Durchführung eines Behandlung-Subgruppe-Interaktionstests. Jede Subgruppe wurde getrennt voneinander bewertet. Das Modell beinhaltete Behandlung, EDSS Strata, Region, Subgruppe und</p>

Item ^a	Charakteristikum	Studieninformation
		<p>Behandlung-Subgruppe-Interaktion als fixe Effekte und log-transformierte standardisierte Studiendauer als Offset. In den Fällen in denen der p-Wert für eine Interaktion kleiner als 0,05 war, wurde wegen der Möglichkeit einer qualitativen Interaktion eine weitere Untersuchung durchgeführt. Zusätzlich, wurden zusammenfassende Statistiken für jede Subgruppe erstellt.</p> <p>Zeit bis zur Behinderungsprogression</p> <p>Die Subgruppenanalysen für das wichtigste sekundäre Zielkriterium „Zeit bis zur Behinderungsprogression“ wurden für die gleichen Subgruppen durchgeführt wie für das primäre Zielkriterium. Auf Konsistenz der Behandlungseffekte über die Subgruppenlevel wurde mit einem Cox-Regressions-Modell getestet, unter Verwendung eines Behandlung-Subgruppe-Interaktionstests. Jede Subgruppe wurde einzeln bewertet. Das Modell beinhaltete Behandlung, EDSS Stara, Region, Subgruppe und Behandlung-Subgruppe-Interaktion als Kovariate. In den Fällen in denen der p-Wert für eine Interaktion kleiner als 0,05 war, wurde wegen der Möglichkeit einer qualitativen Interaktion eine weitere Untersuchung durchgeführt. Zusätzlich, wurden zusammenfassende Statistiken für jede Subgruppe erstellt.</p>
-	Resultate	
13	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	
13a	Anzahl der Studienteilnehmer für jede durch Randomisierung gebildete Behandlungsgruppe, die a) randomisiert wurden, b) tatsächlich die geplante Behandlung/Intervention erhalten haben, c) in der Analyse des primären Zielkriteriums berücksichtigt wurden	<p>Teriflunomid 7 mg</p> <p>a) 408 Patienten waren randomisiert b) 407 Patienten haben die geplante Behandlung erhalten c) 407 Patienten wurden in der Analyse des primären Zielkriteriums berücksichtigt</p> <p>Teriflunomid 14 mg</p> <p>a) 372 Patienten waren randomisiert b) 370 Patienten haben die geplante Behandlung erhalten c) 370 Patienten wurden in der Analyse des primären Zielkriteriums berücksichtigt</p> <p>Placebo</p> <p>a) 389 Patienten waren randomisiert b) 388 Patienten haben die geplante Behandlung erhalten c) 388 Patienten wurden in der Analyse des primären</p>

Item ^a	Charakteristikum	Studieninformation
		Zielkriteriums berücksichtigt
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	<p>Teriflunomid 7 mg</p> <p>134 Patienten beendeten die Studienbehandlung: 54 aufgrund von Nebenwirkungen 30 aufgrund Wahrnehmung geringer Wirksamkeit 3 aufgrund von Protokollverletzung 4 waren „lost to follow-up“ 43 hatten andere Gründe</p> <p>Teriflunomid 14 mg</p> <p>126 Patienten beendeten die Studienbehandlung: 58 aufgrund von Nebenwirkungen 20 aufgrund Wahrnehmung geringer Wirksamkeit 4 aufgrund von Protokollverletzung 3 waren „lost to follow-up“ 41 hatten andere Gründe</p> <p>Placebo</p> <p>125 Patienten beendeten die Studienbehandlung: 26 aufgrund von Nebenwirkungen 37 aufgrund Wahrnehmung geringer Wirksamkeit 15 aufgrund von Protokollverletzung 6 waren „lost to follow-up“ 41 hatten andere Gründe</p>
14	Aufnahme / Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	<p>08/2008 – 04/2012</p> <p>Die Studiendauer beinhaltet eine Screeningphase von bis zu 4 Wochen, gefolgt von einer Behandlungsphase von einer Mindestdauer von 48 Wochen, gefolgt bei einer 4 Wochen andauernden Follow-up Auswaschphase .</p>
14b	Informationen, warum die Studie endete oder beendet wurde	Die Studie wurde plangemäß beendet.
a: nach CONSORT 2010.		

Stellen Sie den Patientenfluss in einem Flow-Chart je Studie dar.

Flow-Chart für die Studie TOWER

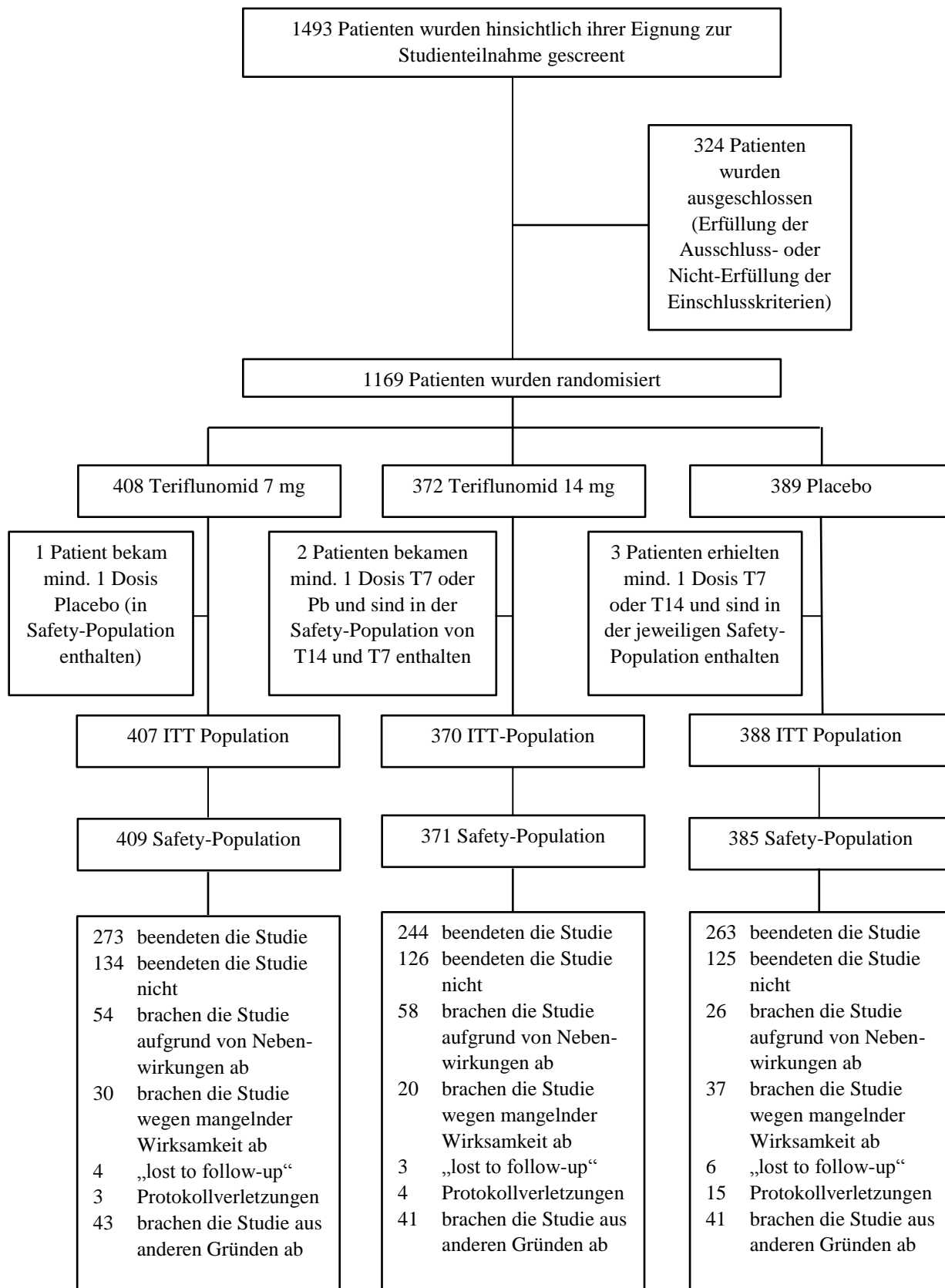


Tabelle 4-209 (Anhang): Studiendesign und -methodik für Studie TENERE

Item ^a	Charakteristikum	Studieninformation
-	Studienziel	
2 b	Genauere Ziele, Fragestellung und Hypothesen	<p>Die Studie ist Teil des klinischen Programms, das durchgeführt wird, um die Zulassung zu erhalten und Teriflunomid der weltweiten klinischen Anwendung zur Verfügung zu stellen. Die Studie wurde global durchgeführt.</p> <p>Primäres Ziel: Beurteilung der Effektivität von 2 Dosierungen Teriflunomid im Vergleich mit Interferon-β 1a bezogen auf das Therapieversagen („Time to Failure“).</p> <p>Sekundäre Ziele: Beurteilung der Effektivität von 2 Dosierungen Teriflunomid im Vergleich mit Interferon-β 1a hinsichtlich Schubfrequenz, Fatigue und Patientenzufriedenheit.</p> <p>Weitere Ziele:</p> <ul style="list-style-type: none"> • Beurteilung der Verträglichkeit und Sicherheit von 2 Dosierungen Teriflunomid im Vergleich mit Interferon-β 1a. • Durchführung einer optionalen pharmakogenomischen Substudie mit dem Ziel den Zusammenhang zwischen dem enzymatischen Teriflunomid-Metabolismus und möglichen Leberschäden, bzw. anderer potentieller Interaktionen zwischen Genvariationen und klinischen Endpunkten, zu beurteilen.
-	Methoden	
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	<p>Bei TENERE handelt es sich um eine multizentrische, multinationale, randomisierte aktivkontrollierte Parallelgruppenstudie.</p> <p>Die Patienten wurden in einem Verhältnis von 1:1:1 den Gruppen Teriflunomid 14 mg, Teriflunomid 7 mg und Rebif zugeteilt.</p> <p>Die beiden verschiedenen Dosierungen von Teriflunomid wurden doppelblind verabreicht, in Bezug auf Interferon-β 1a war die Studie nicht verblindet. Der auswertende Neurologe war in Bezug auf die Medikation verblindet.</p>
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	<p>Es gab 1 Amendment zum Studienprotokoll mit mehreren Detailänderungen. Keine dieser Änderungen hat einen relevanten Einfluss auf die Auswertung der Studienergebnisse.</p> <p>Nachtrag zum Studienprotokoll (25.03.2011) mit folgenden Änderungen:</p> <ul style="list-style-type: none"> • Ausarbeitung der Studierweiterung, die im Originalprotokoll nur erwähnt war • Verkürzung der Eliminationsdauer von 16 auf 4 Wochen („rapid elimination“)

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • Änderung der erlaubten Begleitmedikation: der Ausschluss von CYP2C9 Substraten wurde aufgehoben. Darüber hinaus wurden auch Johanniskrautpräparate erlaubt. • Aufnahme von peripherer Neuropathie (bestätigt durch elektrophysiologische Tests) als ein „alert term“ • Reduktion der Probennahmefrequenz für pharmakokinetische Daten • Thrombozytopenie: Antiplättchen-Antikörper wurden in dem örtlichen Labor anstelle des zentralen Labors getestet • Es wurde klargestellt, dass das Unterschreiben der Einverständniserklärung das Berichten von Nebenwirkungen mit einschließt • Definition von Rebif-Überdosierung, Haftung und Compliance, Entsorgung von Spritzen/ Kartuschen und andere redaktionelle Korrekturen bezüglich des Prüfpräparates • Korrektur einer Unklarheit im Protokoll: <ul style="list-style-type: none"> – Der Schwellenwert, der in Bezug auf eine bestätigte zum Studienabbruch führenden Neutropenie festgelegt ist, sind 1000 Zellen/μL, dementsprechend wurde der Algorithmus in Appendix E durch einen der Studie EFC10891/ TENERE angepassten ersetzt
4	Probanden / Patienten	
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	<p>Einschlusskriterien:</p> <ul style="list-style-type: none"> • Patienten mit schubförmigen Formen der Multiplen Sklerose • Gesicherte MS-Diagnose nach den McDonald's Kriterien • EDSS \leq5,5 • Schriftliche Zustimmung zur Studienteilnahme liegt vor <p>Ausschlusskriterien betreffend Studienbewertung und Methodik:</p> <ul style="list-style-type: none"> • <18 Jahre • HIV positive Patienten oder fehlendes Einverständnis zum HIV Test • Auftreten eines Schubes innerhalb der 30 Tage vor Randomisierung • Geisteszustand, der es dem Patienten nicht erlaubt, die Art, den Umfang und mögliche Konsequenzen der Studie zu verstehen • Patienten, die sich wahrscheinlich nicht an das

Item ^a	Charakteristikum	Studieninformation
		<p>Studienprotokoll halten werden (unkooperative Einstellung, fehlende Möglichkeit Folgeuntersuchungen wahrzunehmen, bekannte Wahrscheinlichkeit, dass die Studie nicht beendet wird)</p> <ul style="list-style-type: none"> • Klinisch relevante kardiovaskuläre, neurologische oder andere schwerwiegende systemische Erkrankungen, die die Umsetzung des Protokolls oder die Interpretation der Studienergebnisse schwierig machen oder den Patienten durch die Studienteilnahme einem Risiko aussetzen • Patienten mit angeborenen oder erworbenen schweren Immunschwächen, einer Krebsvorgeschichte (mit Ausnahme von Basalzell- und Plattenepithelkarzinom, die operativ entfernt wurden und für die keine Metastase nachgewiesen werden konnte), lymphoproliferative Erkrankungen oder ein Patient, der eine Strahlentherapie des lymphatischen Systems erhalten hat • Bekannte Vorgeschichte von aktiver Tuberkulose, die nicht angemessen behandelt wurde • Hypoproteinämie (z. B. im Fall von schwerer Leberschädigung oder des nephrotischen Syndroms) mit einem Serumalbuminwert <3,0 g/dL • Mäßigen bis schweren Beeinträchtigungen der Nierenfunktion, gezeigt durch einen Serumkreatininwert >133µmol/L (oder >1,5mg/dL) • Patienten mit signifikant beeinträchtigter Knochenmarksfunktion oder signifikanter Anämie, Leukopenie oder Thrombozytopenie <ul style="list-style-type: none"> – Hämatokrit <24% und/oder – Absolute Leukozytenzahl <4 000 Zellen/mm³ und/oder – Thrombozytenzahl <150 000 Zellen/mm³ und/oder – Absolute Neutrophile ≤1500 Zellen/mm³ • Persistierende signifikante oder schwere Infektion • Leberfunktionsstörung oder persistierende Leberwerterhöhungen (durch Tests bestätigt) der Serum-Glutamat-Pyruvat-Transferase/Alanin-Aminotransferase, Serum-Glutamat-Oxalacetat-Transferase/Aspartat-Aminotransferase, oder einen Bilirubinwert (direkt), der das 1,5 fache der Obergrenze des Normalwertes übersteigt • Bekannte Hepatitis-Vorgeschichte • Alkohol- oder Drogenmissbrauch in der Vorgeschichte

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • Patient gehört zum Personal oder ist verwandt mit Personal (z.B. Studienleiter oder Forschungsassistent), das direkt in die Durchführung der Studie eingebunden ist • Patient wurde in den 2 Wochen vor Randomisierung mit einer systemischen Kortikosteroidtherapie oder adrenokortikotropem Hormon behandelt • Vorherige oder gleichzeitige Durchführung einer Zytokin-, Glatirameracetat- oder IVIg-Therapie in den 3 Monaten vor Randomisierung • Vorherige oder gleichzeitige Einnahme von Natalizumab (Tysabri®) • Vorherige oder gleichzeitige Einnahme von Cladribin, Mitoxantron oder anderen Immunsuppressiva sowie Azathioprin, Cyclophosphamid, Cyclosporin, Methotrexat oder Mycophenolat • Vorherige Behandlung mit Teriflunomid und vorherige oder gleichzeitige Verwendung von Leflunomid (ARAVA®) oder Unverträglichkeit auf irgendeinen anderen Inhalts- oder Hilfsstoff des Prüfpräparates • Schwangere oder stillende Frauen • Gebärfähige Frauen oder Männer, die keine effektiven Verhütungsmethoden anwenden und/oder gebärfähige Frauen, die nicht auf eine Schwangerschaft getestet werden wollen oder können • Patienten, die im Verlauf der Studiendauer Eltern werden möchten • Vorherige Einnahme irgendeines Prüfpräparates innerhalb der letzten 6 Monate vor Randomisierung <p>Ausschlusskriterien, die sich auf Interferon beta-1a beziehen:</p> <ul style="list-style-type: none"> • Vorherige Anwendung von Rebif® • Vorherige oder gleichzeitige Anwendung anderer Interferone in den 3 Monaten vor Randomisierung oder Vorgeschichte einer Interferontherapie, die aufgrund von Nebenwirkungen, Unverträglichkeit oder Unwirksamkeit abgebrochen wurde • Bekannte Überempfindlichkeit gegen Interferon beta-1a • Vorgeschichte mit Depressionen und/oder Selbstmordgefährdung oder Krampfanfall • Bekannte Überempfindlichkeit gegenüber Humanalbumin, Mannitol (nach United States Pharmacopoeia [USP]) oder Natriumacetat

Item ^a	Charakteristikum	Studieninformation
		<p>Ausschlusskriterien, die sich auf Teriflunomid beziehen:</p> <ul style="list-style-type: none"> • Bekannte Überempfindlichkeit gegenüber Teriflunomid oder Leflunomid • Persistierende Erhöhung (durch wiederholte Testung bestätigt) der Serumamylase- oder Lipase-Werte auf einen Wert, der das 2fache des oberen Limits des Normalwertes übersteigt • Bekannte Vorgeschichte chronischer Pankreaserkrankung oder Pankreatitis • Vorherige Anwendung (innerhalb 4 Wochen vor Randomisierung) oder gleichzeitige Einnahme von Substanzen, die über Cytochrome P2C9 (CYP2C9) verstoffwechselt werden (dies sind u. a. Phenytoin, Warfarin, Tolbutamid, Cholestyramin); vorherige Anwendung von Johanniskrautprodukten oder Hyperforin-haltige Präparate mit einer unbekanntem Konzentration oder einer Konzentration >1% des Extraktes
4b	Studienorganisation und Ort der Studiendurchführung	Die Studie wurde in 53 Zentren in 13 verschiedenen Ländern durchgeführt (Belgien, Kanada, Tschechien, Frankreich, Deutschland, Griechenland, Ungarn, Italien, Polen, Spanien, Schweiz, Tunesien und das Vereinigte Königreich)
5	Interventionen Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	<ul style="list-style-type: none"> • Teriflunomid 7 mg: Filmtabletten, 1x täglich morgens, oral (mit Wasser) • Teriflunomid 14 mg: Filmtabletten, 1x täglich morgens, oral (mit Wasser) • Rebif® (neue Formulierung) Lösung zur Injektion (8,8 µg, 22 µg and 44 µg), einschleichende Dosierung (die ersten 2 Wochen 8,8 µg, die folgenden 2 Wochen 22 µg und ab dann 44 µg pro Applikation), 3x die Woche (immer die gleichen 3 Tage mit jeweils mindestens 48 h Abstand) zur gleichen Zeit (später Nachmittag oder Abend), subkutane Injektion <p>Vorbehandlung mit einem nicht-steroidalen Entzündungshemmer empfohlen</p>
6	Zielkriterien	
6a	Klar definierte primäre und sekundäre Zielkriterien, Erhebungszeitpunkte, ggf. alle zur Optimierung der Ergebnisqualität verwendeten Erhebungsmethoden (z. B. Mehrfachbeobachtungen, Training der Prüfer) und ggf. Angaben zur Validierung von Erhebungsinstrumenten	<p>Primäres Zielkriterium (Therapieversagen):</p> <ul style="list-style-type: none"> • Das primäre Zielkriterium war die Zeit bis zum Therapieversagen „Time to Failure“, wobei das Therapieversagen definiert war als das erste Auftreten eines Schubes oder als permanenter Therapieabbruch aus jedwedem Grund, was auch immer zuerst eintritt. • Ein Schub wurde definiert als das Auftreten eines neuen klinischen Symptoms oder klinische Verschlechterung eines alten Symptoms (stabil über mindestens 30 Tage), das für mindestens 24 persistiert und nicht mit Fieber einhergeht. Jede Episode eines

Item ^a	Charakteristikum	Studieninformation
		<p>Schubes wurde durch den behandelnden Neurologen bestätigt, basierend auf der objektiven Beurteilung durch einen unabhängigen Bewerter (der untersuchende Neurologe), der Folgendes dokumentiert:</p> <ul style="list-style-type: none"> – Eine Erhöhung um 1 Punkt in mindestens 2 FS-Scores oder eine Erhöhung um 2 Punkte in mindestens einem FS-Score (ausgenommen Blase/Mastdarm und Hirnstamm) im Vergleich mit der vorherigen klinisch stabilen Einstufung oder – Eine Erhöhung um mindestens 0,5 Punkte im EDSS-Score (es sei denn der EDSS ist 0, dann ist eine Erhöhung um einen ganzen Punkt erforderlich) im Vergleich mit der vorherigen klinisch stabilen Einstufung. <p>Die Patienten wurden instruiert, ihren Prüfarzt sofort zu kontaktieren, sollte ein Symptom auftreten, das auf einen Schub hinweisen kann. Die Patienten wurden nach dem folgenden Schema so früh wie möglich untersucht, jedoch spätestens nach 7 Tagen nach Kontaktaufnahme:</p> <ul style="list-style-type: none"> – Der behandelnde Neurologe bewertete die Notwendigkeit einer Überweisung an den untersuchenden Neurologen – War eine Beurteilung des untersuchenden Neurologen erforderlich, erhob dieser so früh wie möglich eine Erhebung des FS-Status und des EDSS-Scores durch (vorzugsweise am gleichen Tag, aber maximal 7 Tage nach dem Patientenkontakt) – Die Entscheidung, ob es sich um einen bestätigten Schub handelt wurde durch den behandelnden Neurologen getroffen, basierend auf dem EDSS/FS Score, der vom untersuchenden Neurologen festgestellt worden war <p>Sekundäre Zielkriterien:</p> <ul style="list-style-type: none"> • Ein sekundäres Zielkriterium war die jährliche Schubrate, definiert als die Anzahl bestätigter Schübe, die während der Behandlungsdauer pro Patientenjahr der Behandlung aufgetreten sind • Ein weiteres sekundäres Zielkriterium war die Änderung des FIS-Scores (change from baseline) in Woche 48. Und zwar als Gesamtwert (total score) und als Einzelwerte in den 3 Untergruppen (domain scores). <p>Der FIS (Fatigue Impact Scale) misst den Einfluss der Fatigue auf die 3 Untergruppen kognitive Funktion, physische Funktion und psychosoziale Funktionen.</p>

Item ^a	Charakteristikum	Studieninformation
		<p>Erhebungszeitpunkte: Randomisierungszeitpunkt, Woche 12, 24, 36, 48 und dann alle 24 Wochen bis zum Ende der Behandlungsdauer.</p> <ul style="list-style-type: none"> • Ein weiteres sekundäres Zielkriterium war die Änderung des TSQM-Scores (change from baseline) in Woche 48. <p>Der TSQM (Treatment Satisfaction Questionnaire for Medication) wurde in der Version 1.4 verwendet. Der TSQM deckt mit seinen Fragen verschiedene Bereiche ab:</p> <ul style="list-style-type: none"> – Effektivität (3 Punkte) – Nebenwirkungen (5 Punkte) – Bequemlichkeit (5 Punkte) – Generelle Zufriedenheit (1 Punkt) <p>Erhebungszeitpunkte: Woche 12, 24, 36, 48 und dann alle 24 Wochen bis zum Ende der Behandlungsdauer.</p> <ul style="list-style-type: none"> • Bewertung der Sicherheit und Verträglichkeit der zwei Dosen Teriflunomid im Vergleich mit Interferon beta-1a. <p>Entsprechende „Adverse event reportings“ erfolgten in Woche 2 und Woche 6 und dann alle 6 Wochen bis Woche 24, dann in Woche 36 und ab diesem Termin alle 12 Wochen bis zum Studienende.</p> <ul style="list-style-type: none"> • Optional war eine pharmakogenomische Substudie mit dem Ziel den möglichen Zusammenhang zwischen den hauptsächlich für die Metabolisierung von Teriflunomid zuständigen Enzymsystemen und Sicherheit in Bezug auf die Leber und andere Zusammenhänge mit z. B. Genvariationen oder klinische Endpunkte zu untersuchen.
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	Die Zielkriterien wurden nach Studienbeginn nicht geändert.
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	<p>Die Abschätzung der Fallzahl stützt sich auf einen Vergleich zwischen Teriflunomid und Interferon beta-1a in Bezug auf den primären Endpunkt „Time to Failure“.</p> <p>Für die Berechnung der Power wurde der nQuery Advisor 6.01 verwendet. Basierend auf vorhergehenden Erfahrungen mit Teriflunomid und Interferon beta-1a wurden Hazard Ratios berechnet und die Patientenzahl, die notwendig ist um mit einer Power von 81 % einen Unterschied mit einem zweiseitigen Signifikanzlevel von $\alpha=0,025$ nachzuweisen, als 100 bestimmt.</p> <p>Die Berechnung der notwendigen Fallzahlen basierte auf der Annahme, dass die Hazard rate 0,4186 für 14 mg Teriflunomid und 0,7440 für Beta-Interferon betrug und, dass die Rekrutierung ca. 1,5 Jahre dauerte. Das Signifikanzlevel wurde für die Berücksichtigung von</p>

Item ^a	Charakteristikum	Studieninformation
		Multiplizität spezifiziert. Die Berechnung der Hazard Ratios basierte auf vorhergehenden Erfahrung, aufgrund derer ein Anteil von 30 % für Teriflunomid und ein entsprechender Anteil von 46 % für Beta-Interferon an Patienten, die nach einem Jahr einen Schub erlitten haben, angenommen wurde. Die Entsprechenden Annahmen für vorzeitige Therapieabbrüche betragen 6 % und 12 % für 14 mg Teriflunomid und Beta-Interferon. Diese Ratios wurden bezüglich der Behandlungsdauer und Studienbedingungen extrapoliert. Andere Annahmen waren, dass time-to-event exponentiell mit einer konstanten Hazard Rate verteilt war und dass es sich bei Schüben und Behandlungsabbrüchen um konkurrierende Risiken handelt.
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	Es wurden keine Zwischenanalysen durchgeführt und die Studie auch nicht vorzeitig abgebrochen.
8	Randomisierung, Erzeugung der Behandlungsfolge	
8a	Methode zur Generierung der zufälligen Zuteilung	Die zufällige Zuteilung erfolgte über ein interaktives Spracherkennungssystem (interactive voice response system, IVRS).
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	Eine Stratifizierung erfolgte nach Region und EDSS-Baseline ($\leq 3,5$ und $> 3,5$).
9	Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment) Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war	Die Randomisierung erfolgte zentral über das IVRS. Das IVRS teilte die Nummer des Medikations-Kits zu, der Koordinator dokumentierte diese Nummer und die Patientennummer, und zwar auf dem „case report form“ und den „patient's source documents“. Darüber hinaus trug er die Nummer des Patienten auf dem „medication label“ des Kits ein bevor er es dem Patienten aushändigte. Die Geheimhaltung der Behandlung war bis zur Zuteilung gewährleistet.
10	Randomisierung, Durchführung Wer führte die Zuteilung durch, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	IVRS
11	Verblindung	
11a	Waren a) die Probanden / Patienten und / oder b) diejenigen, die die Intervention / Behandlung durchführten, und / oder c) diejenigen, die die Zielgrößen beurteilten, verblindet oder nicht verblindet, wie wurde die Verblindung vorgenommen?	a) in Bezug auf die Medikation (Teriflunomid oder Beta-Interferon) nicht verblindet, verblindet in Bezug auf die zwei unterschiedlichen Dosierungen von Teriflunomid (identisch aussehende Filmtabletten mit 7 bzw. 14 mg des Wirkstoffes) b) in Bezug auf die Medikation (Teriflunomid oder Beta-Interferon) nicht verblindet, verblindet in Bezug auf die zwei unterschiedlichen Dosierungen von Teriflunomid (identisch aussehende Filmtabletten mit 7 bzw. 14 mg des Wirkstoffes) c) verblindet

Item ^a	Charakteristikum	Studieninformation
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	<p>Die Anwendung von entweder Teriflunomid oder Beta-Interferon erfolgte in einem unverblindeten Design.</p> <p>Die zwei verschiedenen Dosierungen (7 mg und 14 mg) wurden verblindet appliziert, dafür wurden identisch aussehende weiß bis hellgelb gefärbte, bikonvexe Filmtabletten, die entweder 7 mg oder 14 mg Teriflunomid enthielten, in kindersichere Blister verpackt.</p>
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	<p><u>Analyse des primären Endpunktes („Time-to-Failure“):</u></p> <p>Time-to-failure wurde mit einem Log-Rang-Test mit „Time-to-Failure“ als der abhängigen Variable, der Behandlungsgruppe (Teriflunomid 7 mg, Teriflunomid 14 mg und Interferon-beta 1a) als Testvariable und prädefinierter geographischer Region und Baseline EDSS ($\leq 3,5$ und $>3,5$) als Stratum Variablen, getestet.</p> <p>Kaplan-Meier Schätzer und Kurven der kumulativen Inzidenz wurden verwendet um die Rate der therapieversagenden Patienten über die Zeitpunkte hinweg zu bestimmen. Die Durchführung der Analysen erfolgte mit dem SAS® PROC LIFETEST.</p> <p><u>Analysen der sekundären Endpunkte:</u></p> <p><u>a) jährliche Schubrate</u></p> <p>Das wichtigste statistische Modell und Anpassung nach Kovariaten:</p> <p>Die jährliche Schubrate wurde mit einem Poisson-Regressionsmodell analysiert, mit einer robusten Varianz des Fehlers, wodurch den möglicherweise übermäßig gestreuten Daten Rechnung getragen wurde. Das Modell beinhaltete die Gesamtzahl bestätigter Schübe, die zwischen Randomisierung und letzter Dosis aufgetreten sind, als Response-Variable und Behandlungsgruppe, EDSS und Region als Kovariate. Um die unterschiedlichen Behandlungsdauern der Patienten adäquat zu berücksichtigen, wurde die log-transformierte standardisierte Studiendauer in dem Modell als „Offset“-Variable integriert, was eine geeignete Berechnung der Schubrate ermöglichte.</p> <p>Die robuste Varianz des Fehlers wurde durch Spezifizierung der Patientenennung in der wiederholten Angabe geschätzt durch Verwendung des SAS PROC GENMID, das ein Äquivalent des GEE (generalized estimating equation)- Modells ist. Zweiseitige 95%-Konfidenzintervalle des Schubratenverhältnisses wurden für die Vergleiche jeder aktiven Behandlung gegen Interferon beta-1a bereitgestellt und die geschätzte Schubrate und ihr zweiseitiges 95%-Konfidenzintervall wurden für jede Behandlungsgruppe geliefert. Die groben Schätzungen der jährlichen Schubrate wurden auch für jede Behandlungsgruppe dargestellt.</p> <p>Die jährliche Schubrate wurde mit dem GEE-Modell statt mit dem Poisson-Modell analysiert, da der GEE-Schätzer sich gegen Verletzung der Korrelationsstruktur und der</p>

Item ^a	Charakteristikum	Studieninformation
		<p>Verteilungsannahmen als robust erwies.</p> <p><u>b) Fatigue Impact Scale Gesamtwert and Werte der Einzelbereiche</u> (kognitive, physische, and psychosoziale Funktion)</p> <p>Die Veränderung in Bezug auf den Basiswert in Woche 48 wurde mit einem MMRM (mixed-effect model with repeated measures) analysiert. Die MMRM-Analysen wurden mit Hilfe von PROC MIXED in SAS durchgeführt. Dies wurde erreicht durch Einbau der Veränderungen in Bezug auf den Basiswert zu allen post-Randomisierungs-Arztbesuchen in dem Behandlungszeitraum bis zur Woche 48.</p> <p>Das Modell beinhaltet Faktoren (fixed effects) für Behandlung, EDSS, Region, Arztbesuch, Behandlung - Arztbesuch-Interaktion, Basislinien-Wert und Basislinie-Arztbesuch-Interaktion. Der Faktor Arztbesuch mit nominalen Arztbesuchen hatte 4 Ebenen (Woche 12, Woche 24, Woche 36 und Woche 48)</p> <p>Eine unstrukturierte Korrelationsmatrix wurde verwendet, um den Fehler innerhalb der Daten eines Patienten zu modellieren. Parameter wurden mit der restricted maximum likelihood-Methode und dem Newton-Raphson Algorithmus geschätzt.</p> <p>Der gemeinsame Nenner der Freiheitsgrade wurde mit der Satterthwaite's Approximation geschätzt.</p> <p>Least squares mean (LS-Mean) Schätzungen der Veränderung in Bezug auf den Basiswert in Woche 48 wurden nach Behandlungsgruppe bereitgestellt, sowie die Unterschiede dieser Schätzungen im Vergleich zu Interferon beta-1a, zusammen mit ihren entsprechenden Standardabweichungen und zugehörigen 95%-Konfidenzintervalle.</p> <p>Student t-tests wurden angewendet, um die statistische Signifikanz des Vergleiches von jeder Teriflunomid-Dosis gegenüber Interferon beta-1a zu bestimmen.</p> <p>Zusätzlich dazu, wurden die Daten und die Veränderung in Bezug auf den Basiswert zusammengefasst, unter Verwendung beschreibender Statistik (Mittelwert, Median, Standardabweichung, Minimum, Maximum) sortiert nach Arztbesuch und basierend auf beobachteten Fällen.</p> <p>Graphische Darstellungen stellen die über die Zeit beobachteten Trends anschaulich dar.</p> <p><u>c) Treatment Satisfaction Questionnaire for Medication</u></p> <p>Vier Kategorien (Effektivität, Nebenwirkungen, Bequemlichkeit und Gesamtzufriedenheit) in Woche 48 wurden mit einem MMRM-Ansatz analysiert.</p> <p>Die MMRM Analysen wurden mit Hilfe des PROC MIXED in SAS durch Einbau der Werte von allen Post-Randomisierungs-Arztbesuchen in dem Behandlungszeitraum bis zu Woche 48. Das Modell beinhaltet Faktoren (fixed effects) für Behandlung, EDSS, Region, Arztbesuch und Behandlung-Arztbesuch-</p>

Item ^a	Charakteristikum	Studieninformation
		<p>Interaktion. Der Faktor Arztbesuch mit nominalen Arztbesuchen hatte 4 Ebenen (Woche 12, Woche 24, Woche 36 und Woche 48). Eine unstrukturierte Korrelationsmatrix wurde verwendet, um den Fehler innerhalb der Daten eines Patienten zu modellieren. Parameter wurden mit der restricted maximum likelihood-Methode und dem Newton-Raphson Algorithmus geschätzt.</p> <p>Der gemeinsame Nenner der Freiheitsgrade wurde mit der Satterthwaite's Approximation geschätzt.</p> <p>Least squares mean (LS-Mean) Schätzungen der Veränderung in Bezug auf den Basiswert in Woche 48 wurden nach Behandlungsgruppe bereitgestellt, sowie die Unterschiede dieser Schätzungen im Vergleich zu Interferon beta-1a, zusammen mit ihren entsprechenden Standardabweichungen und zugehörigen 95%-Konfidenzintervalle.</p> <p>Student t-tests wurden angewendet, um die statistische Signifikanz des Vergleiches von jeder Teriflunomid-Dosis gegenüber Interferon beta-1a zu bestimmen.</p> <p>Zusätzlich dazu, wurden die Daten und die Veränderung in Bezug auf den Basiswert zusammengefasst, unter Verwendung beschreibender Statistik (Mittelwert, Median, Standardabweichung, Minimum, Maximum) sortiert nach Arztbesuchen und basierend auf beobachteten Fällen.</p> <p>Graphische Darstellungen stellen die über die Zeit beobachteten Trends anschaulich dar.</p>
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	<p><u>Primärer Endpunkt („Time-to-Failure“):</u></p> <p>Als unterstützende Analyse, wurde der primäre Endpunkt mit einem PH (proportional hazards)- Modell untersucht, um einige Interaktionseffekte auszugleichen. Das Modell beinhaltete „Time-to-Failure“ als Response-Variable und Behandlungsgruppe, Baseline EDSS und Region als Kovariate. Zusätzlich wurden Interaktionen durch „Behandlung nach EDSS“ und „Behandlung nach Region“ entsprechend dem 5% Signifikanzlevel getrennt bewertet und zwar durch Hinzufügen des entsprechenden Interaktionsterms zu dem Modell. Wenn eine statistisch signifikante Interaktion bestimmt werden konnte, mussten wegen der Möglichkeit qualitativer Interaktionen weitere Untersuchungen durchgeführt werden. Die unterstützende Analyse erfolgte mit dem SAS PROG PHREG.</p> <p>Bezüglich folgender Faktoren wurden Subgruppenanalysen für den primären Endpunkt durchgeführt:</p> <ul style="list-style-type: none"> • Geschlecht • Altersgruppe (<38, ≥38) • Region • Baseline EDSS • Anzahl Schübe in den letzten 2 Jahren (0,1,2,3 und

Item ^a	Charakteristikum	Studieninformation
		<p>≥4)</p> <p>Das PH Modell wurde verwendet um festzustellen, ob es Belege für eine Inkonsistenz der Behandlungseffekte über die Ebenen jedes Subgruppenfaktors gibt. Dafür wurde ein Behandlung-Subgruppe-Interaktionstest angewendet. Jeder Subgruppenfaktor wurde getrennt bewertet. Das Modell beinhaltete Behandlung, Subgruppe der Wahl und Behandlung-Subgruppe-Interaktion als fixe Effekte. Wenn der p-Wert für eine Interaktion kleiner als 0,05 war, musste wegen der Möglichkeit einer qualitativen Interaktion eine weitere Untersuchung durchgeführt werden.</p> <p><u>Analysen der sekundären Endpunkte:</u></p> <p><u>a) jährliche Schubrate</u></p> <p>Zusätzlich wurden „Behandlung -EDSS“ und „Behandlung- Region“ am 5%-Signifikanzlevel getrennt ausgewertet. Dies wurde durch Hinzufügen des entsprechenden Interaktionsterms zu dem o. g. GEE-Modell erreicht. Wurde eine statistisch signifikante Interaktion festgestellt, machte das wegen der Möglichkeit qualitativer Interaktionen weitere Analysen notwendig.</p> <p><u>Verfahren mit Studienabbrechern</u></p> <p>Da das Modell nach Behandlungsdauer adjustiert war, entsprach sich die Ableitung der Analysenvariable für Abbrecher und Vollender.</p> <p>Wenn Patienten die Behandlung abgebrochen haben, wurde die standardisierte Studiendauer berechnet als (Datum der letzten Dosis-Randomisierungsdatum +1)/365,25. Die Gesamtzahl bestätigter Schübe, die während der Studie vor Behandlungsabbruch aufgetreten sind, wurde als Response gewertet und log-Transformierung der Standardstudiendauer als Offset-Variable.</p> <p><u>Subgruppenanalysen</u></p> <p>Für die jährliche Schubrate wurden die gleichen Subgruppenanalysen durchgeführt wie für den primären Endpunkt.</p>
-	Resultate	
13	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	
13a	Anzahl der Studienteilnehmer für jede durch Randomisierung gebildete Behandlungsgruppe, die a) randomisiert wurden, b) tatsächlich die geplante Behandlung/Intervention erhalten haben, c) in der Analyse des primären Zielkriteriums berücksichtigt wurden	<p><u>Teriflunomid 7 mg</u></p> <p>a) 109 Patienten wurden randomisiert b) 110 erhielten die Behandlung c) 109 wurden in der Analyse des primären Zielkriteriums berücksichtigt (ITT-Population)</p> <p><u>Teriflunomid 14 mg</u></p> <p>a) 111 Patienten wurden randomisiert b) 110 Patienten erhielten die Behandlung c) 111 Patienten wurden in der Analyse des primären</p>

Item ^a	Charakteristikum	Studieninformation
		<p>Zielkriteriums berücksichtigt (ITT-Population)</p> <p><u>Rebif 44</u></p> <p>a) 104 Patienten wurden randomisiert</p> <p>b) 101 Patienten erhielten die Behandlung</p> <p>c) 104 Patienten wurden in der Analyse des primären Zielkriteriums berücksichtigt (ITT-Population)</p>
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	<p><u>Teriflunomid 7 mg</u></p> <p>Ein weiterer Patient wurde nach Randomisierung der 7 mg Teriflunomid-Gruppe zugeteilt, erhielt aber für 2,5 Monate die 14 mg Dosis. Dieser Patient ist in der 7 mg Teriflunomid-Gruppe enthalten</p> <p>109 waren randomisiert und behandelt</p> <p>89 beendeten die Studie</p> <p>20 beendeten die Studie nicht</p> <p>11 beendeten die Studie auf eigenen Wunsch</p> <p>9 brachen die Studie aufgrund von Nebenwirkungen ab</p> <p>7 brachen die Studie wegen mangelnder Wirksamkeit ab</p> <p>1 „lost to follow-up“</p> <p>3 brachen die Studie aus anderen Gründen ab</p> <p><u>Teriflunomid 14 mg</u></p> <p>1 Patient wurde laut Randomisierung der 14 mg Teriflunomid-Gruppe zugeteilt, erhielt aber die 7 mg Dosis für 3 Monate. Für die Safety-Analyse wurde dieser Patient der 7 mg Gruppe zugeteilt.</p> <p>111 waren randomisiert und behandelt</p> <p>89 beendeten die Studie</p> <p>22 beendeten die Studie nicht</p> <p>14 beendeten die Studie auf eigenen Wunsch</p> <p>12 brachen die Studie aufgrund von Nebenwirkungen ab</p> <p>4 brachen die Studie wegen mangelnder Wirksamkeit ab</p> <p>1 „lost to follow-up“</p> <p>5 brachen die Studie aus anderen Gründen ab</p> <p><u>Rebif 44</u></p> <p>3 Patienten aus der Rebif-Behandlungsgruppe wurden zwar randomisiert, erhielten aber keine Studienmedikation. Dieser Patienten wurden aus der Safety-Population ausgeschlossen.</p> <p>101 waren randomisiert und behandelt</p> <p>71 beendeten die Studie</p> <p>30 beendeten die Studie nicht</p> <p>11 beendeten die Studie auf eigenen Wunsch</p>

Item ^a	Charakteristikum	Studieninformation
		22 brachen die Studie aufgrund von Nebenwirkungen ab 2 brachen die Studie wegen mangelnder Wirksamkeit ab 1 Protokollverletzung 5 brachen die Studie aus anderen Gründen ab
14	Aufnahme / Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	Studienaufnahme des ersten Patienten: 16 April 2009 Beendigung der Studie für den letzten Patienten: 14 September 2011 Für alle Patienten, die nicht in die Verlängerungsstudie eintreten wollten, erfolgte eine Wash-out-Prozedur und zwei Follow-up Arztbesuchen jeweils 4 und 16 Wochen nach Ende der Behandlung.
14b	Informationen, warum die Studie endete oder beendet wurde	Die Studie wurde plangemäß beendet nachdem der letzte aufgenommene Patient die Mindeststudiendauer von 48 Wochen erreicht hatte.
a: nach CONSORT 2010.		

Stellen Sie den Patientenfluss in einem Flow-Chart je Studie dar

Flow-Chart für die Studie TENERE

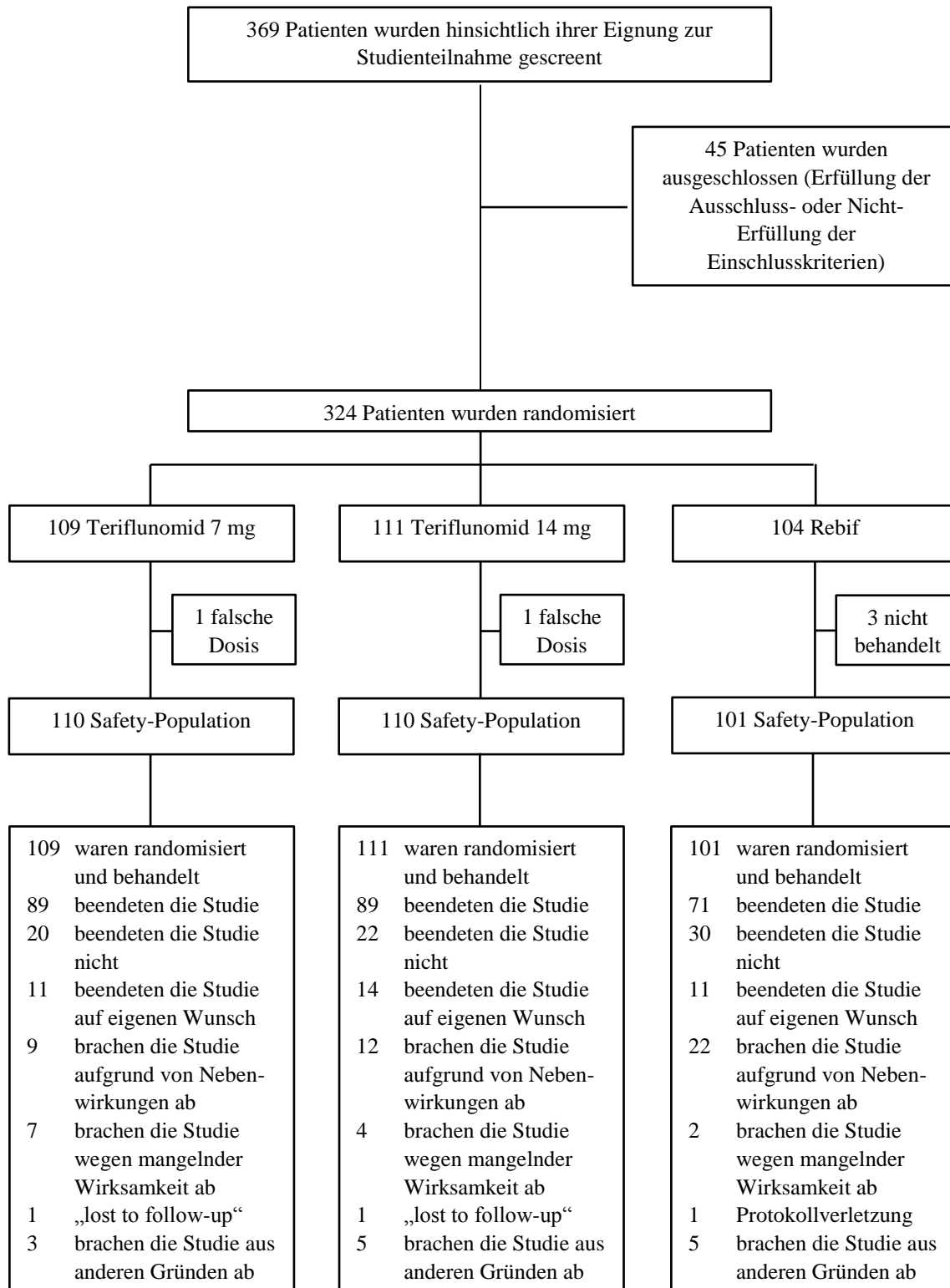


Tabelle 4-210 (Anhang): Studiendesign und -methodik für Studie PRISMS

Item ^a	Charakteristikum	Studieninformation
-	Studienziel	
2 b	Genaue Ziele, Fragestellung und Hypothesen	<p>Ziel der Studie war die Untersuchung des therapeutischen Effekts von zwei Dosen Interferon beta-1a s.c. bei Patienten mit schubförmig-remittierender Multipler Sklerose. Der therapeutische Effekt wurde anhand der patientenrelevanten Endpunkte Schubrate, Behinderung, Krankheitsaktivität und Krankheitslast bestimmt.</p> <p>Die Fragestellung basierte auf der Hypothese, dass durch die Behandlung mit Interferon beta-1a s. c. die Schubrate gesenkt werden kann.</p>
-	Methoden	
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	Bei PRISMS handelt es sich um eine randomisierte, doppelblinde, multizentrische placebokontrollierte Parallelgruppenstudie.
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	Keine Angabe
4	Probanden / Patienten	
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	<p>Einschlusskriterien:</p> <ul style="list-style-type: none"> • Erwachsene Patienten • RRMS • EDSS 0 – 5,0 • Mindestens 2 Schübe in den vorangegangenen 2 Jahren • Klinisch definierte oder Labor-gestützte definitive MS von mindestens 1 Jahr Dauer • Schriftliche Einverständniserklärung <p>Ausschlusskriterien</p> <ul style="list-style-type: none"> • Vorherige systemische Behandlung mit Interferonen, Bestrahlungen des Lymphsystems, Cyclophosphamid oder anderen immunmodulatorischen oder-suppressiven Behandlungen in den vorangegangenen 12 Monaten
4b	Studienorganisation und Ort der Studiendurchführung	22 Studienzentren in neun Ländern: Australien, Belgien, Kanada, ,Finnland, Deutschland, Niederlande, Schweden, Schweiz, Großbritannien
5	Interventionen	

Item ^a	Charakteristikum	Studieninformation
	Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	<p>Interferon beta-1a 22 µg, dreimal in der Woche als subkutane Injektion (0,5 ml)</p> <p>Interferon beta-1a 44 µg, dreimal in der Woche als subkutane Injektion (0,5 ml)</p> <p>Placebo, dreimal in der Woche als subkutane Injektion (0,5 ml)</p>
6	Zielkriterien	
6a	Klar definierte primäre und sekundäre Zielkriterien, Erhebungszeitpunkte, ggf. alle zur Optimierung der Ergebnisqualität verwendeten Erhebungsmethoden (z. B. Mehrfachbeobachtungen, Training der Prüfer) und ggf. Angaben zur Validierung von Erhebungsinstrumenten	<p>Primäre Zielkriterien:</p> <p>Das primäre Zielkriterium war die Anzahl der Schübe im Verlauf der Studie. Schübe waren definiert als das Auftreten eines neuen Symptoms oder die Verschlechterung eines alten Symptoms über die letzten 24 h, das eine MS-Aktivität zugeordnet werden konnte und von einer stabilen Phase oder einer Verbesserung gefolgt wurde (in den folgenden 30 Tagen). Die Schwere eines Schubes wurde mittels der „Scribbs neurological rating scale“ oder der "activities of daily living scale“ bewertet.</p> <p>Weitere Zielkriterien:</p> <ul style="list-style-type: none"> • Zeit bis zum ersten und zum zweiten Schub • Anteil schubfreier Patienten • Behinderungsprogression (Zunahme des EDSS um mindestens 1 Punkt, anhaltend über 3 Monate) • Benötigte Steroidtherapie und Hospitalisierung • Krankheitsaktivität (MRT) und Krankheitslast <p>Zusätzlich wurde noch die Fläche unter der Zeit/EDSS-Kurve bestimmt.</p> <p>Des Weiteren wurden 267 Patienten, die in englischsprachigen Ländern rekrutiert wurden, mit dem Beck's hopelessness scale auf die Entwicklung einer Depression hin untersucht.</p>
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	Keine Angaben
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	Die Studie hat eine Power von 80 % einen mittleren Unterschied von 0,64 in der mittleren Anzahl an Schüben zwischen der 44 µg Gruppe und der Placebo-Gruppe

Item ^a	Charakteristikum	Studieninformation
		festzustellen. Mit einer Effektgröße von 0,4 (Mittelwert 0,64; Standardabweichung 1,62) wurde für einen zweiseitigen Zwei-Stichproben- <i>t</i> -Test mit $p = 0,05$ eine minimale Gruppengröße von 100 Patienten benötigt.
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	Die Studie wurde nicht vorzeitig abgebrochen.
8	Randomisierung, Erzeugung der Behandlungsfolge	
8a	Methode zur Generierung der zufälligen Zuteilung	Die Randomisierungsliste wurde von Serono Biometrics Computer-generiert.
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	Eine Stratifizierung nach Studienzentrum wurde vorgenommen. Es handelte sich um eine Blockrandomisierung mit einer Blockgröße von 6.
9	Randomisierung, Geheimhaltung der Behandlungsfolge (allocation concealment) Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war	Entsprechend der Computer-generierten Liste, wurde die Studienmedikation verpackt und an die Studienzentren geliefert, so dass die Behandlungszuteilung verdeckt blieb.
10	Randomisierung, Durchführung Wer führte die Zuteilung durch, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	s. Items 8 und 9, keine weiteren Informationen angegeben
11	Verblindung	
11a	Waren a) die Probanden / Patienten und / oder b) diejenigen, die die Intervention / Behandlung durchführten, und / oder c) diejenigen, die die Zielgrößen beurteilten, verblindet oder nicht verblindet, wie wurde die Verblindung vorgenommen?	Das gesamte Personal war verblindet, die Patienten wurden von jeweils zwei Ärzten betreut: ein behandelnder Arzt, welcher für das gesamte medizinische Management des Patienten (inklusive Behandlung von unerwünschten Ereignissen) verantwortlich war und ein zweiter, auswertender Neurologe, mit Verantwortlichkeit für die Bestimmung und Follow-up von erfolgten Schüben. Die Patienten waren angewiesen, ihre Injektionsstellen entsprechend abzudecken. a) die Patienten waren verblindet b) der Behandler war verblindet c) der auswertende Neurologe war ebenfalls verblindet
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Nicht relevant.
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	Die Auswertungen wurden nach dem ITT-Prinzip durchgeführt. Alle Endpunkt-

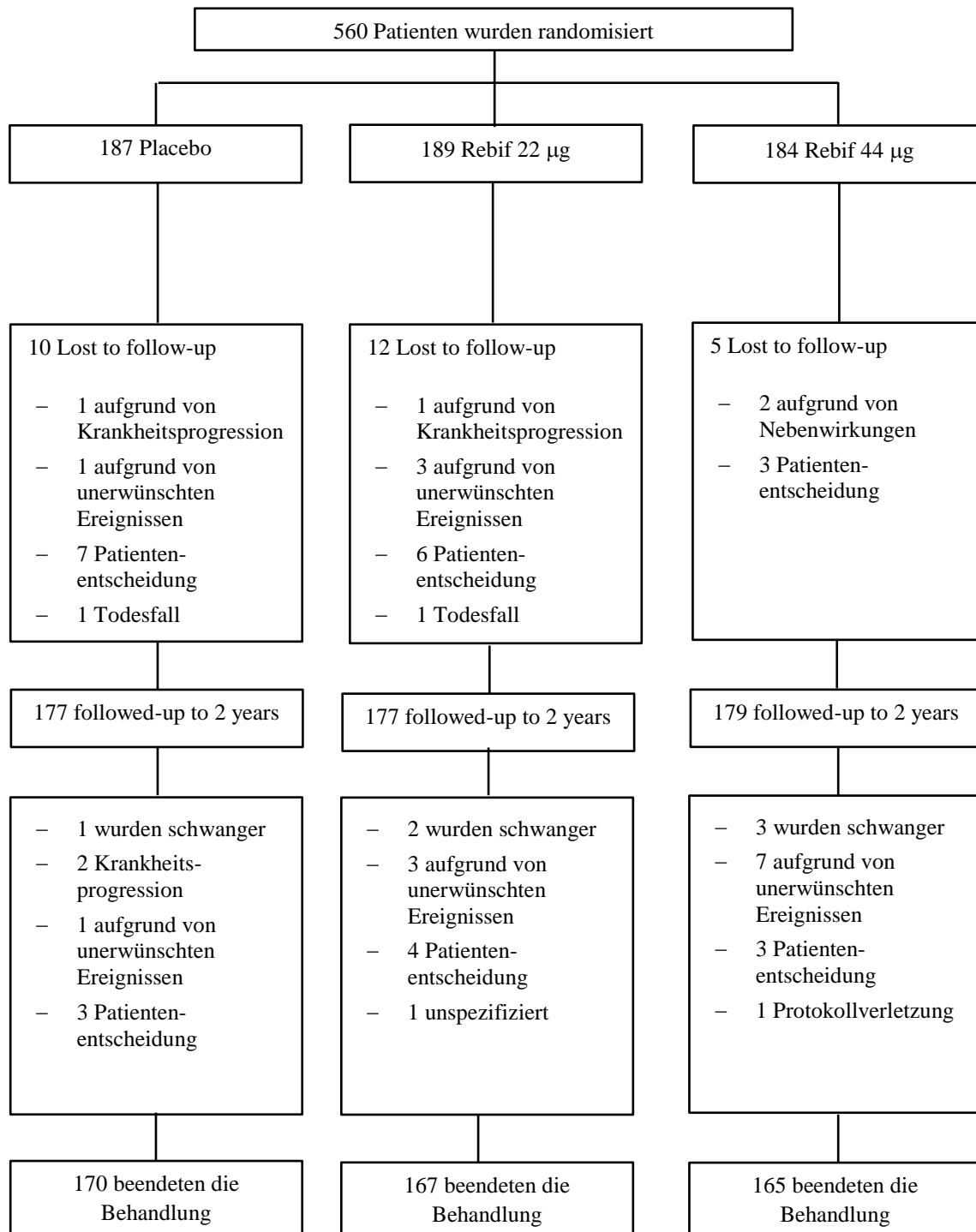
Item ^a	Charakteristikum	Studieninformation
		<p>Ergebnisse wurden eingeschlossen.</p> <p>Für jeden Endpunkt wurde ein globales Modell gefittet, das Zentrum und Behandlung mit einbezogen hat. Die Interaktion war jedoch nicht signifikant in irgendeinem Modell und wurde deswegen entfernt. Ein generalisiertes lineares Modell (GLM) mit einem log link und einer zum Mittelwert proportionalen Varianz wurde eingesetzt, um die Schubzahl zu analysieren. Für „Time-to-event“-Variablen wurde ein Cox proportional hazards Modell eingesetzt, logistische Regression für binäre Endpunkte, ANOVA für Rang-Daten anderer kontinuierlicher Endpunkte und Chi-Quadrat-Tests für Patienten mit speziellen Kategorien unerwünschter Ereignisse.</p>
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	Zusätzlich wurden Patienten mit einer schwereren Behinderung (EDSS > 3,5) in einem Modell analysiert, das Kontrollen für die Behandlung, das Zentrum, den Baseline EDSS und Behandlung-Baseline EDSS enthielt.
-	Resultate	
13	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	
13a	<p>Anzahl der Studienteilnehmer für jede durch Randomisierung gebildete Behandlungsgruppe, die</p> <p>a) randomisiert wurden,</p> <p>b) tatsächlich die geplante Behandlung/Intervention erhalten haben,</p> <p>c) in der Analyse des primären Zielkriteriums berücksichtigt wurden</p>	<p>Interferon beta-1a 22 µg, dreimal in der Woche als subkutane Injektion (0,5 ml):</p> <p>a) randomisiert wurden 189 Patienten</p> <p>b) 189 Patienten haben die Behandlung erhalten (keine genaue Angabe)</p> <p>c) 189 Patienten sind in der Analyse des primären Zielkriteriums berücksichtigt worden</p> <p>Interferon beta-1a 44 µg, dreimal in der Woche als subkutane Injektion (0,5 ml):</p> <p>a) randomisiert wurden 184 Patienten</p> <p>b) 184 Patienten haben die Behandlung erhalten (keine genaue Angabe)</p> <p>c) 184 Patienten sind in der Analyse des primären Zielkriteriums berücksichtigt worden</p> <p>Placebo, dreimal in der Woche als subkutane Injektion (0,5 ml):</p> <p>a) randomisiert wurden 187 Patienten</p> <p>b) 187 Patienten haben die Behandlung erhalten (keine genaue Angabe)</p> <p>c) 187 Patienten sind in der Analyse des</p>

Item ^a	Charakteristikum	Studieninformation
		primären Zielkriteriums berücksichtigt worden
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	<p>Interferon beta-1a 22 µg, dreimal in der Woche als subkutane Injektion (0,5 ml):</p> <ul style="list-style-type: none"> • 12 Patienten lost to follow-up <ul style="list-style-type: none"> – 1 aufgrund von Krankheitsprogression – 3 aufgrund von unerwünschten Ereignissen – 6 Patientenentscheidung – 1 Todesfall – 1 unspezifiziert • 10 weitere Patienten beendeten die Behandlung nicht <ul style="list-style-type: none"> – 2 wurden schwanger – 3 aufgrund von unerwünschten Ereignissen – 4 Patientenentscheidung – 1 unspezifiziert <p>Interferon beta-1a 44 µg, dreimal in der Woche als subkutane Injektion (0,5 ml):</p> <ul style="list-style-type: none"> • 5 Patienten lost to follow-up <ul style="list-style-type: none"> – 2 aufgrund von unerwünschten Ereignissen – 3 Patientenentscheidung • 14 weitere Patienten beendeten die Behandlung nicht <ul style="list-style-type: none"> – 3 wurden schwanger – 7 aufgrund von unerwünschten Ereignissen – 3 Patientenentscheidung – 1 Protokollverletzung <p>Placebo, dreimal in der Woche als subkutane Injektion (0,5 ml):</p> <ul style="list-style-type: none"> • 10 Patienten lost to follow-up <ul style="list-style-type: none"> – 1 aufgrund von Krankheitsprogression – 1 aufgrund von unerwünschten Ereignissen – 7 Patientenentscheidung – 1 Todesfall • 7 weitere Patienten beendeten die Behandlung nicht <ul style="list-style-type: none"> – 1 wurden schwanger – 2 Krankheitsprogression – 1 aufgrund von unerwünschten

Item^a	Charakteristikum	Studieninformation
		Ereignissen – 3 Patientenentscheidung
14	Aufnahme / Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	Die Patienten wurden zwischen Mai 1994 und Februar 1995 rekrutiert. Die Studie lief über 2 Jahre.
14b	Informationen, warum die Studie endete oder beendet wurde	Die Studie wurde planmäßig beendet.
a: nach CONSORT 2010.		

Stellen Sie den Patientenfluss in einem Flow-Chart je Studie dar.

Flow-Chart für die Studie PRISMS



Anhang 4-G: Bewertungsbögen zur Einschätzung von Verzerrungsaspekten

Beschreiben Sie nachfolgend die Verzerrungsaspekte jeder eingeschlossenen Studie (einschließlich der Beschreibung für jeden berücksichtigten Endpunkt). Erstellen Sie hierfür je Studie eine separate Version des nachfolgend dargestellten Bewertungsbogens.

Tabelle 4-211 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie <Studienbezeichnung>

Grundsätzliche Hinweise zum Ausfüllen des Bewertungsbogens

Dieser Bewertungsbogen dient der Dokumentation der Einstufung des Potenzials der Ergebnisse für Verzerrungen (Bias). Für jede Studie soll aus diesem Bogen nachvollziehbar hervorgehen, inwieweit die Ergebnisse für die einzelnen Endpunkte als möglicherweise verzerrt bewertet wurden, was die Gründe für die Bewertung waren und welche Informationen aus den Quellen dafür Berücksichtigung fanden.

Der Bogen gliedert sich in zwei Teile:

- **Kriterien zur Bewertung des Verzerrungspotenzials auf Studienebene.** In diesem Teil sind die endpunktübergreifenden Kriterien aufgelistet.
- **Kriterien zur Bewertung des endpunktspezifischen Verzerrungspotenzials.** In diesem Teil sind die Kriterien aufgelistet, die für jeden Endpunkt separat zu prüfen sind.

Für jedes Kriterium sind unter „Angaben zum Kriterium“ alle relevanten Angaben aus den Quellen zur Bewertung einzutragen (Stichworte reichen ggf., auf sehr umfangreiche Informationen in den Quellen kann verwiesen werden).

Grundsätzlich werden die Bögen studienbezogen ausgefüllt. Wenn mehrere Quellen zu einer Studie vorhanden sind, müssen die herangezogenen Quellen in der folgenden Tabelle genannt und jeweils mit Kürzeln (z. B. A, B, C ...) versehen werden. Quellenspezifische Angaben im weiteren Verlauf sind mit dem jeweiligen Kürzel zu kennzeichnen.

Studie: _____

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genauere Benennung der Quelle	Kürzel

Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse:

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A: Kriterien des Verzerrungspotenzials der Ergebnisse auf Studienebene:**Einstufung als randomisierte Studie**

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anhaltspunkte vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

1.

für randomisierte Studien:**Adäquate Erzeugung der Randomisierungssequenz**

ja: Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

unklar: Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

nein: Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:**Zeitliche Parallelität der Gruppen**

ja: Die Gruppen wurden zeitlich parallel verfolgt.

unklar: Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

nein: Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien:

Verdeckung der Gruppenzuteilung („allocation concealment“)

ja: Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierte Arzneimitteln/Arzneimittelbehältern
- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

unklar: Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau.

nein: Die Gruppenzuteilung erfolgte nicht verdeckt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:

Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

ja: Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anhaltspunkte dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere

Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

unklar: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau.

nein: Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und Behandlern

Patient

ja: Die Patienten waren verblindet.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Behandler bzw. Weiterbehandler

ja: Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, den primären Behandler (Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Behandler nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*
Zulässige Gründe sind:
 - *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
 - *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*
 - *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
- *Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.*
- *Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.*

Anzumerken ist, dass Anhaltspunkte für eine ergebnisgesteuerte Darstellung eines Endpunkts das Verzerrungspotenzial für die Ergebnisse der übrigen Endpunkte erhöhen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden

muss. Insbesondere bei Anhaltspunkten dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, ist das Verzerrungspotenzial für die anderen Endpunkte u. U. erhöht. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Erhöhung des Verzerrungspotenzials der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen.

Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und das Verzerrungspotenzial anderer Endpunkte nicht berührt.

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

5. Fehlen sonstiger (endpunktübergreifender) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien*
- intransparenter Patientenfluss*
- Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:*
 - Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).*
 - Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf. im Nachhinein von der Biometrie durchzuführen).*
 - Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.*
 - Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.*

ja **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

**Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene
(ausschließlich für randomisierte Studien durchzuführen):**

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

B: Kriterien des Verzerrungspotenzials der Ergebnisse pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: _____

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

- ja:** Eines der folgenden Merkmale trifft zu:
- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
 - Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
 - Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien,

statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Tabelle 4-212 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie TEMSO

Grundsätzliche Hinweise zum Ausfüllen des Bewertungsbogens

Dieser Bewertungsbogen dient der Dokumentation der Einstufung des Potenzials der Ergebnisse für Verzerrungen (Bias). Für jede Studie soll aus diesem Bogen nachvollziehbar hervorgehen, inwieweit die Ergebnisse für die einzelnen Endpunkte als möglicherweise verzerrt bewertet wurden, was die Gründe für die Bewertung waren und welche Informationen aus den Quellen dafür Berücksichtigung fanden.

Der Bogen gliedert sich in zwei Teile:

- **Kriterien zur Bewertung des Verzerrungspotenzials auf Studienebene.** In diesem Teil sind die endpunktübergreifenden Kriterien aufgelistet.
- **Kriterien zur Bewertung des endpunktspezifischen Verzerrungspotenzials.** In diesem Teil sind die Kriterien aufgelistet, die für jeden Endpunkt separat zu prüfen sind.

Für jedes Kriterium sind unter „Angaben zum Kriterium“ alle relevanten Angaben aus den Quellen zur Bewertung einzutragen (Stichworte reichen ggf., auf sehr umfangreiche Informationen in den Quellen kann verwiesen werden).

Grundsätzlich werden die Bögen studienbezogen ausgefüllt. Wenn mehrere Quellen zu einer Studie vorhanden sind, müssen die herangezogenen Quellen in der folgenden Tabelle genannt und jeweils mit Kürzeln (z. B. A, B, C ...) versehen werden. Quellenspezifische Angaben im weiteren Verlauf sind mit dem jeweiligen Kürzel zu kennzeichnen.

Studie: TEMSO

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genauere Benennung der Quelle	Kürzel
Studienbericht inklusive Studienprotokoll [32]	EFC6049 (TEMSO)
Publikation der Studienergebnisse [35]	O'Connor 2011
Publikation der Subgruppenanalysen [43]	Miller 2012

Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse:

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A: Kriterien des Verzerrungspotenzials der Ergebnisse auf Studienebene:**Einstufung als randomisierte Studie**

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anhaltspunkte vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

1.

für randomisierte Studien:**Adäquate Erzeugung der Randomisierungssequenz**

ja: Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

unklar: Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

nein: Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:**Zeitliche Parallelität der Gruppen**

ja: Die Gruppen wurden zeitlich parallel verfolgt.

unklar: Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

nein: Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien:

Verdeckung der Gruppenzuteilung („allocation concealment“)

ja: Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierte Arzneimitteln/Arzneimittelbehältern
- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

unklar: Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau.

nein: Die Gruppenzuteilung erfolgte nicht verdeckt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:

Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

ja: Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anhaltspunkte dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere

Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

unklar: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau.

nein: Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und Behandlern

Patient

ja: Die Patienten waren verblindet.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Behandler bzw. Weiterbehandler

ja: Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, den primären Behandler (Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Behandler nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:
Teriflunomid und Placebo wurden als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*

Zulässige Gründe sind:

 - *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
 - *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*
 - *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
- *Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.*
- *Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.*

Anzumerken ist, dass Anhaltspunkte für eine ergebnisgesteuerte Darstellung eines Endpunkts das Verzerrungspotenzial für die Ergebnisse der übrigen Endpunkte

erhöhen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anhaltspunkten dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, ist das Verzerrungspotenzial für die anderen Endpunkte u. U. erhöht. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Erhöhung des Verzerrungspotenzials der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen.

Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und das Verzerrungspotenzial anderer Endpunkte nicht berührt.

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

5. Fehlen sonstiger (endpunktübergreifender) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien*
- intransparenter Patientenfluss*
- Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:*
 - Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).*
 - Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf. im Nachhinein von der Biometrie durchzuführen).*
 - Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.*
 - Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.*

ja **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

**Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene
(ausschließlich für randomisierte Studien durchzuführen):**

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

B: Kriterien des Verzerrungspotenzials der Ergebnisse pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: Schübe

Das Verzerrungspotential wird für alle die Schübe betreffenden Endpunkte gemeinsam ausgewertet. Die Auswertung der jährlichen Schubrate, der Zeit bis zum Schub und des Anteils schubfreier Patienten basieren auf der gleichen Erhebung und werden durch die gleichen Faktoren beeinflusst. Eine adäquate Erhebung der Schübe ist für diese Endpunkte ausschlaggebend und potentiell der verzerrende Faktor.

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurden als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante

Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.

- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Zeit bis zur Behinderungsprogression

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurden als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der

Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Mittlere Änderung des MSFC-Z-Scores

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurden als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs,

Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen
- unplausible Angaben
- Anwendung inadäquater statistischer Verfahren

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Mittlere Änderung des EDSS-Scores**1. Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurden als Filmpillen verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr

genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen
- unplausible Angaben
- Anwendung inadäquater statistischer Verfahren

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Fatigue

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

- ja:** Der Endpunkt wurde verblindet erhoben.
- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurden als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren

möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).

- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Krankheitslast (Burden of Disease-MRT)**1. Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung
Teriflunomid und Placebo wurden als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppeneinteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

- ja**
- nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Lebensqualität (SF-36 und EQ-5D)

Für beide Fragebögen (SF-36 und EQ-5D) wird die Bewertung des Verzerrungspotentials gemeinsam durchgeführt. Es handelt sich bei beiden Fragebögen um etablierte Messinstrumente zur Lebensqualität, eine potentielle Verzerrung der Erhebung in der Studie TEMSO ist von den gleichen Faktoren abhängig (in diesem Fall vor allem von der Verblindung der Patienten, da es sich um patientenberichtete Endpunkte handelt).

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurden als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-

to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppeneinteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- unplausible Angaben*
- Anwendung inadäquater statistischer Verfahren*

- ja**
- nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

- niedrig:** Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.
- hoch:** Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Mortalität

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurden als Filmpillen verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch,

insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen
- un plausible Angaben
- Anwendung inadäquater statistischer Verfahren

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Unerwünschte Ereignisse

Alle Endpunkte zu unerwünschten Ereignissen (Gesamtraten unerwünschter Ereignisse (AEs, SAEs, Behandlungsabbrüche) und alle unerwünschten Ereignisse von besonderem Interesse) werden in Bezug auf eine mögliche Verzerrung gemeinsam bewertet. Den in diesem Bereich durchgeführten Auswertungen liegen allesamt die in der Studie erhobenen Daten zu unerwünschten Ereignissen zugrunde, dementsprechend werden sie von den gleichen potentiell verzerrenden Faktoren beeinflusst.

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurden als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante

Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.

- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Für die Analyse der Safety-Endpunkte wurde anstatt der ITT-Population die Safety-Population herangezogen. Die Safety-Population unterscheidet sich von der ITT-Population dahingehend, dass Patienten entsprechend der erhaltenen Medikation untersucht wurden. Für die Auswertung der Verträglichkeitsdaten ist dies der weniger verzerrende Ansatz. Die Safety-Population weicht in keinem Behandlungsarm über 5 % von der ITT-Population ab.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Tabelle 4-213 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie TOWER

Grundsätzliche Hinweise zum Ausfüllen des Bewertungsbogens

Dieser Bewertungsbogen dient der Dokumentation der Einstufung des Potenzials der Ergebnisse für Verzerrungen (Bias). Für jede Studie soll aus diesem Bogen nachvollziehbar hervorgehen, inwieweit die Ergebnisse für die einzelnen Endpunkte als möglicherweise verzerrt bewertet wurden, was die Gründe für die Bewertung waren und welche Informationen aus den Quellen dafür Berücksichtigung fanden.

Der Bogen gliedert sich in zwei Teile:

- **Kriterien zur Bewertung des Verzerrungspotenzials auf Studienebene.** In diesem Teil sind die endpunktübergreifenden Kriterien aufgelistet.
- **Kriterien zur Bewertung des endpunktspezifischen Verzerrungspotenzials.** In diesem Teil sind die Kriterien aufgelistet, die für jeden Endpunkt separat zu prüfen sind.

Für jedes Kriterium sind unter „Angaben zum Kriterium“ alle relevanten Angaben aus den Quellen zur Bewertung einzutragen (Stichworte reichen ggf., auf sehr umfangreiche Informationen in den Quellen kann verwiesen werden).

Grundsätzlich werden die Bögen studienbezogen ausgefüllt. Wenn mehrere Quellen zu einer Studie vorhanden sind, müssen die herangezogenen Quellen in der folgenden Tabelle genannt und jeweils mit Kürzeln (z. B. A, B, C ...) versehen werden. Quellenspezifische Angaben im weiteren Verlauf sind mit dem jeweiligen Kürzel zu kennzeichnen.

Studie: TOWER

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genaue Benennung der Quelle	Kürzel
Studienbericht inklusive Studienprotokoll [31]	EFC10531 (TOWER)

Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse:

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A: Kriterien des Verzerrungspotenzials der Ergebnisse auf Studienebene:**Einstufung als randomisierte Studie**

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anhaltspunkte vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

1.

für randomisierte Studien:**Adäquate Erzeugung der Randomisierungssequenz**

ja: Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

unklar: Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

nein: Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:**Zeitliche Parallelität der Gruppen**

ja: Die Gruppen wurden zeitlich parallel verfolgt.

unklar: Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

nein: Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien:

Verdeckung der Gruppenzuteilung („allocation concealment“)

ja: Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierte Arzneimitteln/Arzneimittelbehältern
- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

unklar: Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau.

nein: Die Gruppenzuteilung erfolgte nicht verdeckt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:

Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

ja: Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anhaltspunkte dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere

Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

unklar: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau.

nein: Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und Behandlern

Patient

ja: Die Patienten waren verblindet.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Behandler bzw. Weiterbehandler

ja: Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, den primären Behandler (Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Behandler nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Teriflunomid und Placebo wurde als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*

Zulässige Gründe sind:

 - *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
 - *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*
 - *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
- *Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.*
- *Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.*

Anzumerken ist, dass Anhaltspunkte für eine ergebnisgesteuerte Darstellung eines Endpunkts das Verzerrungspotenzial für die Ergebnisse der übrigen Endpunkte

erhöhen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anhaltspunkten dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, ist das Verzerrungspotenzial für die anderen Endpunkte u. U. erhöht. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Erhöhung des Verzerrungspotenzials der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen.

Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und das Verzerrungspotenzial anderer Endpunkte nicht berührt.

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

5. Fehlen sonstiger (endpunktübergreifender) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien*
- intransparenter Patientenfluss*
- Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:*
 - Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).*
 - Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf. im Nachhinein von der Biometrie durchzuführen).*
 - Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.*
 - Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.*

ja **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

**Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene
(ausschließlich für randomisierte Studien durchzuführen):**

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

B: Kriterien des Verzerrungspotenzials der Ergebnisse pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: Schübe

Das Verzerrungspotential wird für alle die Schübe betreffenden Endpunkte gemeinsam ausgewertet. Die Auswertung der jährlichen Schubrate, der Zeit bis zum Schub und des Anteils schubfreier Patienten basieren auf der gleichen Erhebung und werden durch die gleichen Faktoren beeinflusst. Eine adäquate Erhebung der Schübe ist für diese Endpunkte ausschlaggebend und potentiell der verzerrende Faktor.

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu

anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung
Teriflunomid und Placebo wurde als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante

Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.

- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Zeit bis zur Behinderungsprogression

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurde als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die

Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- unplausible Angaben*
- Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch

mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Mittlere Änderung des EDSS-Wertes

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurde als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung

angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Fatigue

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurde als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- Anwendung inadäquater statistischer Verfahren

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: SF-36 (Lebensqualität)

5. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurde als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

6. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

7. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

8. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der

Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Mortalität

5. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurde als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

6. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu

definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppeneinteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

7. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

8. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Unerwünschte Ereignisse

Alle Endpunkte zu unerwünschten Ereignissen (Gesamtraten unerwünschter Ereignisse (AEs, SAEs, Behandlungsabbrüche) und alle unerwünschten Ereignisse von besonderem Interesse) werden in Bezug auf eine mögliche Verzerrung gemeinsam bewertet. Den in diesem Bereich durchgeführten Auswertungen liegen allesamt die in der Studie erhobenen Daten zu unerwünschten Ereignissen zugrunde, dementsprechend werden sie von den gleichen potentiell verzerrenden Faktoren beeinflusst.

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Teriflunomid und Placebo wurde als Filmtabletten verabreicht, die optisch nicht unterscheidbar waren, weder das behandelnde Personal noch die Patienten wurden über die Gruppenzuteilung informiert.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen

Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Für die Analyse der Safety-Endpunkte wurde anstatt der ITT-Population die Safety-Population herangezogen. Die Safety-Population unterscheidet sich von der ITT-Population dahingehend, dass Patienten entsprechend der erhaltenen Medikation untersucht wurden. Für die Auswertung der Verträglichkeitsdaten ist dies der weniger verzerrende Ansatz. Die Safety-Population weicht in keinem Behandlungsarm über 5 % von der ITT-Population ab.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Tabelle 4-214 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie TENERE

Grundsätzliche Hinweise zum Ausfüllen des Bewertungsbogens

Dieser Bewertungsbogen dient der Dokumentation der Einstufung des Potenzials der Ergebnisse für Verzerrungen (Bias). Für jede Studie soll aus diesem Bogen nachvollziehbar hervorgehen, inwieweit die Ergebnisse für die einzelnen Endpunkte als möglicherweise verzerrt bewertet wurden, was die Gründe für die Bewertung waren und welche Informationen aus den Quellen dafür Berücksichtigung fanden.

Der Bogen gliedert sich in zwei Teile:

- **Kriterien zur Bewertung des Verzerrungspotenzials auf Studienebene.** In diesem Teil sind die endpunktübergreifenden Kriterien aufgelistet.
- **Kriterien zur Bewertung des endpunktspezifischen Verzerrungspotenzials.** In diesem Teil sind die Kriterien aufgelistet, die für jeden Endpunkt separat zu prüfen sind.

Für jedes Kriterium sind unter „Angaben zum Kriterium“ alle relevanten Angaben aus den Quellen zur Bewertung einzutragen (Stichworte reichen ggf., auf sehr umfangreiche Informationen in den Quellen kann verwiesen werden).

Grundsätzlich werden die Bögen studienbezogen ausgefüllt. Wenn mehrere Quellen zu einer Studie vorhanden sind, müssen die herangezogenen Quellen in der folgenden Tabelle genannt und jeweils mit Kürzeln (z. B. A, B, C ...) versehen werden. Quellenspezifische Angaben im weiteren Verlauf sind mit dem jeweiligen Kürzel zu kennzeichnen.

Studie: TENERE

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genaue Benennung der Quelle	Kürzel
Studienbericht inklusive Studienprotokoll [35]	EFC10891 (TENERE)

Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse:

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A: Kriterien des Verzerrungspotenzials der Ergebnisse auf Studienebene:**Einstufung als randomisierte Studie**

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anhaltspunkte vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

1.

für randomisierte Studien:**Adäquate Erzeugung der Randomisierungssequenz**

ja: Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

unklar: Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

nein: Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:**Zeitliche Parallelität der Gruppen**

ja: Die Gruppen wurden zeitlich parallel verfolgt.

unklar: Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

nein: Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien:

Verdeckung der Gruppenzuteilung („allocation concealment“)

ja: Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierte Arzneimitteln/Arzneimittelbehältern
- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

unklar: Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau.

nein: Die Gruppenzuteilung erfolgte nicht verdeckt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:

Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

ja: Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anhaltspunkte dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere

Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

unklar: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau.

nein: Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und Behandlern

Patient

ja: Die Patienten waren verblindet.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Aufgrund der unterschiedlichen Applikationsform (Filmtablette vs. Spritze) wäre eine Verblindung der Patienten lediglich durch Double-Dummy-Design möglich. Da sich aber auch durch ein solches Design die unterschiedlichen Reaktionen an der Injektionspritze nicht hätten verblinden lassen, wäre ein solches Design im Falle dieser Studie ethisch nicht vertretbar gewesen.

Hinsichtlich der zwei unterschiedlichen Dosierungen von Teriflunomid waren die Patienten verblindet.

Behandler bzw. Weiterbehandler

ja: Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, den primären Behandler (Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Behandler nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Der behandelnde Arzt war bezüglich der Medikation (Teriflunomid oder Beta-Interferon) aus den oben genannten Gründen nicht verblindet. Der auswertende Neurologe war jedoch verblindet, was als angemessene Verblindung der weiteren an der Behandlung beteiligten Personen angesehen werden kann.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*

Zulässige Gründe sind:

- *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
- *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*

- *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
- *Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.*
- *Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.*

Anzumerken ist, dass Anhaltspunkte für eine ergebnisgesteuerte Darstellung eines Endpunkts das Verzerrungspotenzial für die Ergebnisse der übrigen Endpunkte erhöhen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anhaltspunkten dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, ist das Verzerrungspotenzial für die anderen Endpunkte u. U. erhöht. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Erhöhung des Verzerrungspotenzials der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen.

Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und das Verzerrungspotenzial anderer Endpunkte nicht berührt.

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Kritisch ist anzumerken, dass der Endpunkt „Behinderungsprogression“ in der Studie nicht vorgesehen ist, obwohl es sich um einen üblichen Endpunkt handelt. Allerdings ist die Studiendauer für diesen Endpunkt auch zu kurz und der Endpunkt wurde posthoc ausgewertet, so dass dies insgesamt nicht zu einer Abstufung der Studie hin zu einem hohen Verzerrungspotential führt.

5. Fehlen sonstiger (endpunktübergreifender) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien*
- *intransparenter Patientenfluss*

- Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:
 - Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).
 - Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf. im Nachhinein von der Biometrie durchzuführen).
 - Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.
 - Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

**Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene
(ausschließlich für randomisierte Studien durchzuführen):**

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

B: Kriterien des Verzerrungspotenzials der Ergebnisse pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: Schübe

Das Verzerrungspotential wird für alle die Schübe betreffenden Endpunkte gemeinsam ausgewertet. Die Auswertung der jährlichen Schubrate, der Zeit bis zum Schub und des Anteils schubfreier Patienten basieren auf der gleichen Erhebung und werden durch die gleichen Faktoren beeinflusst. Eine adäquate Erhebung der Schübe ist für diese Endpunkte ausschlaggebend und potentiell der verzerrende Faktor.

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Der behandelnde Arzt war bezüglich der Medikation (Teriflunomid oder Beta-Interferon) aus den oben genannten Gründen (Bewertung des Verzerrungspotentials auf Studienebene) nicht verblindet. Der auswertende Neurologe war jedoch verblindet, was als angemessene Verblindung der weiteren an der Behandlung beteiligten Personen angesehen werden kann.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten

ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen
- unplausible Angaben
- Anwendung inadäquater statistischer Verfahren

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Behinderungsprogression

Alle Posthoc-Analysen von Daten aus der TENERE-Studie basieren auf den in der Studie erfolgten Erhebungen des EDSS-Wertes, weswegen bei der Einstufung des Verzerrungspotentials keine Differenzierung nach verschiedenen Endpunkten sinnvoll ist.

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung
Die Behinderungsprogression ist in der TENERE-Studie posthoc ausgewertet worden. Die Patienten und der behandelnde Neurologe waren in dieser Studie nicht verblindet, durch das Hinzuziehen eines verblindeten auswertenden Neurologen konnten die Schübe dennoch verblindet erhoben werden. Eine analoge adäquate Verblindung der regelmäßigen Erhebung des EDSS-Scores ist im Studienbericht und –protokoll nicht vorgeschrieben, so dass das genaue Procedere unklar bleibt. Es ist möglich, dass der Endpunkterheber nicht verblindet war.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- Anwendung inadäquater statistischer Verfahren

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Der Endpunkterheber war bezüglich der Therapie möglicherweise nicht verblindet.
Der Endpunkt war nicht im Studienprotokoll vorgesehen, es handelt sich um eine Posthoc-Analyse.

Endpunkt: Therapieversagen (Time to Failure)

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Der behandelnde Arzt war bezüglich der Medikation (Teriflunomid oder Beta-Interferon) aus den oben genannten Gründen (Bewertung des Verzerrungspotentials auf Studienebene) nicht verblindet. Der auswertende Neurologe war jedoch verblindet, was als angemessene Verblindung der weiteren an der Behandlung beteiligten Personen angesehen werden kann.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien,

statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Fatigue**1. Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Es handelt sich bei dem angewendeten Messinstrument um einen Fragebogen, der von dem Patienten selbst auszufüllen war. Da die Patienten bezüglich der Behandlung nicht verblindet waren, kann auch der Endpunkt nicht als „verblindet erhoben“ eingestuft werden.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

- ja**
- nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

- niedrig:** Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Der Endpunkt wurde nicht verblindet erhoben. Das kann, muss aber nicht, zu einer relevanten Verzerrung des Ergebnisses führen.

Endpunkt: Patientenzufriedenheit (TSQM = Treatments Satisfaction Questionnaire for Medication)

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Es handelt sich bei dem angewendeten Messinstrument um einen Fragebogen, der von dem Patienten selbst auszufüllen war. Da die Patienten bezüglich der Behandlung nicht verblindet waren, kann auch der Endpunkt nicht als „verblindet erhoben“ eingestuft werden.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten

entsprechend ihrer Gruppeneinteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen
- unplausible Angaben
- Anwendung inadäquater statistischer Verfahren

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Der Endpunkt wurde nicht verblindet erhoben. Das kann, muss aber nicht, zu einer relevanten Verzerrung des Ergebnisses führen.

Endpunkt: Mortalität

9. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

- ja:** Der Endpunkt wurde verblindet erhoben.
- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Weder die Patienten noch der behandelnde Neurologe war verblindet.
Bezogen auf die auftretenden Todesfälle führt dies jedoch nicht zu einer
Verzerrung der Ergebnisse.

10. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

- ja:** Eines der folgenden Merkmale trifft zu:
- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante

Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.

- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

11. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

12. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Unerwünschte Ereignisse

Alle Endpunkte zu unerwünschten Ereignissen (Gesamtraten unerwünschter Ereignisse (AEs, SAEs, Behandlungsabbrüche) und alle unerwünschten Ereignisse von besonderem Interesse) werden in Bezug auf eine mögliche Verzerrung gemeinsam bewertet. Den in diesem Bereich durchgeführten Auswertungen liegen allesamt die in der Studie erhobenen Daten zu unerwünschten Ereignissen zugrunde, dementsprechend werden sie von den gleichen potentiell verzerrenden Faktoren beeinflusst. Eine mögliche Verzerrung von Ergebnissen zu einzelnen unerwünschten Ereignissen von besonderem Interesse, die in der Natur des betrachteten Ereignissen zu suchen sind, wird im Rahmen der Diskussion eingebracht.

5. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden,

wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

- ja:** Der Endpunkt wurde verblindet erhoben.
- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Weder die Patienten noch der behandelnde Neurologe war bezüglich der Behandlung verblindet. In Bezug auf die unerwünschten Ereignisse kann dies zu einer relevanten Verzerrung des Ergebnisses führen, je nachdem welcher Natur das unerwünschte Ereignis ist.

6. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

- ja:** Eines der folgenden Merkmale trifft zu:
- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante

Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.

- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Für die Analyse der Safety-Endpunkte wurde anstatt der ITT-Population die Safety-Population herangezogen. Die Safety-Population unterscheidet sich von der ITT-Population dahingehend, dass Patienten entsprechend der erhaltenen Medikation untersucht wurden. Für die Auswertung der Verträglichkeitsdaten ist dies der weniger verzerrende Ansatz. Die Safety-Population weicht in keinem Behandlungsarm über 5 % von der ITT-Population ab.

7. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

8. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Die Ergebnisse für den Endpunkt unerwünschte Ereignisse sind für einzelne Ereignisse möglicherweise relevant verzerrt, da die Patienten bezüglich ihrer Studienmedikation nicht verblindet waren.

Tabelle 4-215 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie PRISMS

Grundsätzliche Hinweise zum Ausfüllen des Bewertungsbogens

Dieser Bewertungsbogen dient der Dokumentation der Einstufung des Potenzials der Ergebnisse für Verzerrungen (Bias). Für jede Studie soll aus diesem Bogen nachvollziehbar hervorgehen, inwieweit die Ergebnisse für die einzelnen Endpunkte als möglicherweise verzerrt bewertet wurden, was die Gründe für die Bewertung waren und welche Informationen aus den Quellen dafür Berücksichtigung fanden.

Der Bogen gliedert sich in zwei Teile:

- **Kriterien zur Bewertung des Verzerrungspotenzials auf Studienebene.** In diesem Teil sind die endpunktübergreifenden Kriterien aufgelistet.
- **Kriterien zur Bewertung des endpunktspezifischen Verzerrungspotenzials.** In diesem Teil sind die Kriterien aufgelistet, die für jeden Endpunkt separat zu prüfen sind.

Für jedes Kriterium sind unter „Angaben zum Kriterium“ alle relevanten Angaben aus den Quellen zur Bewertung einzutragen (Stichworte reichen ggf., auf sehr umfangreiche Informationen in den Quellen kann verwiesen werden).

Grundsätzlich werden die Bögen studienbezogen ausgefüllt. Wenn mehrere Quellen zu einer Studie vorhanden sind, müssen die herangezogenen Quellen in der folgenden Tabelle genannt und jeweils mit Kürzeln (z. B. A, B, C ...) versehen werden. Quellenspezifische Angaben im weiteren Verlauf sind mit dem jeweiligen Kürzel zu kennzeichnen.

Studie: PRISMS

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genaue Benennung der Quelle	Kürzel
Publikation: Randomised double-blind placebo-controlled study of interferon β -1a in relapsing/remitting multiple sclerosis PRISMS (Prevention of Relapses and Disability by Interferon β -1a Subcutaneously in Multiple Sclerosis) Study Group Lancet. 1998 Nov 7;352:1498-504[78]	Ebers et al. [78]

Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse:

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A: Kriterien des Verzerrungspotenzials der Ergebnisse auf Studienebene:**Einstufung als randomisierte Studie**

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anhaltspunkte vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

1.

für randomisierte Studien:**Adäquate Erzeugung der Randomisierungssequenz**

ja: Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

unklar: Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

nein: Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:**Zeitliche Parallelität der Gruppen**

ja: Die Gruppen wurden zeitlich parallel verfolgt.

unklar: Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

nein: Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien:

Verdeckung der Gruppenzuteilung („allocation concealment“)

ja: Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierte Arzneimitteln/Arzneimittelbehältern
- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

unklar: Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau.

nein: Die Gruppenzuteilung erfolgte nicht verdeckt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:

Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

ja: Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anhaltspunkte dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere

Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

unklar: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau.

nein: Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und Behandlern

Patient

ja: Die Patienten waren verblindet.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Behandler bzw. Weiterbehandler

ja: Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, den primären Behandler (Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die Behandler nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Sowohl die Studienmedikation Interferon beta-1a als auch Placebo wurde dreimal wöchentlich über subkutane Injektionen verabreicht, die Patienten waren angewiesen die Injektionsstellen bei der Untersuchung entsprechend abzudecken, so dass keine Entblindung des Arztes aufgrund von typischen Reaktionen an der Injektionsstelle stattfinden konnte.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*

Zulässige Gründe sind:

- *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
- *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*
- *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
- *Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.*
- *Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.*

Anzumerken ist, dass Anhaltspunkte für eine ergebnisgesteuerte Darstellung eines Endpunkts das Verzerrungspotenzial für die Ergebnisse der übrigen Endpunkte erhöhen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anhaltspunkten dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, ist das Verzerrungspotenzial für die anderen Endpunkte u. U. erhöht. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Erhöhung des Verzerrungspotenzials der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen.

Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und das Verzerrungspotenzial anderer Endpunkte nicht berührt.

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

5. Fehlen sonstiger (endpunktübergreifender) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien*
- *intransparenter Patientenfluss*
- *Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:*
 - *Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).*
 - *Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf. im Nachhinein von der Biometrie durchzuführen).*
 - *Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.*

- *Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

**Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene
(ausschließlich für randomisierte Studien durchzuführen):**

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

B: Kriterien des Verzerrungspotenzials der Ergebnisse pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: Jährliche Schubrate

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

- ja:** Der Endpunkt wurde verblindet erhoben.
- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Sowohl die Studienmedikation Interferon beta-1a als auch Placebo wurde dreimal wöchentlich über subkutane Injektionen verabreicht, die Patienten waren angewiesen die Injektionsstellen bei der Untersuchung entsprechend abzudecken, so dass keine Entblindung des Arztes aufgrund von typischen Reaktionen an der Injektionsstelle stattfinden konnte.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppeneinteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

- ja:** Eines der folgenden Merkmale trifft zu:
- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.

- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Zeit bis zur Behinderungsprogression**1. Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Sowohl die Studienmedikation Interferon beta-1a als auch Placebo wurde dreimal wöchentlich über subkutane Injektionen verabreicht, die Patienten waren angewiesen die Injektionsstellen bei der Untersuchung entsprechend abzudecken, so dass keine Entblindung des Arztes aufgrund von typischen Reaktionen an der Injektionsstelle stattfinden konnte.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche

Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen
- unplausible Angaben
- Anwendung inadäquater statistischer Verfahren

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Mortalität

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Sowohl die Studienmedikation Interferon beta-1a als auch Placebo wurde dreimal wöchentlich über subkutane Injektionen verabreicht, die Patienten waren angewiesen die Injektionsstellen bei der Untersuchung entsprechend abzudecken, so dass keine Entblindung des Arztes aufgrund von typischen Reaktionen an der Injektionsstelle stattfinden konnte.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).

- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: unerwünschte Ereignisse

Alle Endpunkte zu unerwünschten Ereignissen (Gesamtraten unerwünschter Ereignisse (AEs, SAEs, Behandlungsabbrüche) und alle unerwünschten Ereignisse von besonderem Interesse) werden in Bezug auf eine mögliche Verzerrung gemeinsam bewertet. Den in diesem Bereich durchgeführten Auswertungen liegen allesamt die in der Studie erhobenen Daten zu unerwünschten Ereignissen zugrunde, dementsprechend werden sie von den gleichen potentiell verzerrenden Faktoren beeinflusst. Eine mögliche Verzerrung von Ergebnissen zu einzelnen unerwünschten Ereignissen von besonderem Interesse, die in der Natur des betrachteten Ereignissen zu suchen sind, wird im Rahmen der Diskussion eingebracht.

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

ja: Der Endpunkt wurde verblindet erhoben.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Sowohl die Studienmedikation Interferon beta-1a als auch Placebo wurde dreimal wöchentlich über subkutane Injektionen verabreicht, die Patienten waren angewiesen die Injektionsstellen bei der Untersuchung entsprechend abzudecken, so dass keine Entblindung des Arztes aufgrund von typischen Reaktionen an der Injektionsstelle stattfinden konnte.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“-)Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

ja: Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenanzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien,

statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

unklar: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

nein: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

ja: Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

unklar: Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

nein: Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

ja

nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich

für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

niedrig: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

hoch: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:
