Dokumentvorlage, Version vom 18.11.2025

Dossier zur Nutzenbewertung gemäß § 35a SGB V

Zuranolon (Zurzuvae™)

Biogen GmbH

Modul 4 A

Behandlung von Erwachsenen mit postpartaler Depression (PPD) nach Entbindung

Medizinischer Nutzen und
medizinischer Zusatznutzen,
Patientengruppen mit therapeutisch
bedeutsamem Zusatznutzen

Stand: 15.07.2026

Inhaltsverzeichnis

Seite

Tabellenverzeichnis **4**
Abbildungsverzeichnis **10**
Abkürzungsverzeichnis **11**
4.1 Zusammenfassung der Inhalte von Modul 4 16
4.2 Methodik 34
4.2.1 Fragestellung 34
4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung 36
4.2.3 Informationsbeschaffung 38
4.2.3.1 Studien des pharmazeutischen Unternehmers 38
4.2.3.2 Bibliografische Recherche 38
4.2.3.3 Suche in Studienregistern/Studienergebnisdatenbanken 40
4.2.3.4 Suche auf der Internetseite des Gemeinsamen Bundesausschuss 41
4.2.3.5 Selektion relevanter Studien 42
4.2.4 Bewertung der Aussagekraft der Nachweise 42
4.2.5 Informationssynthese und -analyse 44
4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien 44
4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien 45
4.2.5.2.1 Analysepopulationen 46
4.2.5.2.2 Demografische und krankheitsspezifische Patientencharakteristika 46
4.2.5.2.3 Patientenrelevante Endpunkte 48
4.2.5.2.4 Patientenrelevanz, Validität und Operationalisierung der Endpunkte 52
4.2.5.2.4.1 Mortalität 52
4.2.5.2.4.2 Morbidität 52
4.2.5.2.4.3 Lebensqualität 56
4.2.5.2.4.4 Sicherheit 57
4.2.5.3 Meta-Analysen 58
4.2.5.4 Sensitivitätsanalysen 61
4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren 62
4.2.5.6 Indirekte Vergleiche 64
4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen 66
4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel 67
4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel 67
4.3.1.1.1 Studien des pharmazeutischen Unternehmers 67
4.3.1.1.2 Studien aus der bibliografischen Recherche 68
4.3.1.1.3 Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken 70
4.3.1.1.4 Studien aus der Suche auf der Internetseite des Gemeinsamen Bundesausschusses 71
4.3.1.1.5 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel 72

4.3.1.2 Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel ... 73
4.3.1.2.1 Studiendesign und Studienpopulationen ... 73
4.3.1.2.2 Verzerrungspotenzial auf Studienebene ... 93
4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien ... 95
4.3.1.3.1.1 Mortalität – RCT ... 96
4.3.1.3.1.2 Morbidität – RCT ... 104
4.3.1.3.1.3 Lebensqualität – RCT ... 149
4.3.1.3.1.4 Sicherheit – RCT ... 184
4.3.1.3.2 Subgruppenanalysen – RCT ... 203
4.3.2 Weitere Unterlagen ... 233
4.3.2.1 Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien ... 233
4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche ... 233
4.3.2.1.2 Charakteristika der Studien für indirekte Vergleiche ... 234
4.3.2.1.3 Ergebnisse aus indirekten Vergleichen ... 234
4.3.2.1.3.1 <Endpunkt xxx> – indirekte Vergleiche aus RCT ... 234
4.3.2.1.3.2 Subgruppenanalysen – indirekte Vergleiche aus RCT ... 237
4.3.2.2 Nicht randomisierte vergleichende Studien ... 237
4.3.2.2.1 Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien ... 237
4.3.2.2.2 Charakteristika der nicht randomisierten vergleichenden Studien ... 238
4.3.2.2.3 Ergebnisse aus nicht randomisierten vergleichenden Studien ... 239
4.3.2.2.3.1 <Endpunkt xxx> – nicht randomisierte vergleichende Studien ... 240
4.3.2.2.3.2 Subgruppenanalysen – nicht randomisierte vergleichende Studien ... 241
4.3.2.3 Weitere Untersuchungen ... 241
4.3.2.3.1 Ergebnis der Informationsbeschaffung – weitere Untersuchungen ... 241
4.3.2.3.2 Charakteristika der weiteren Untersuchungen ... 242
4.3.2.3.3 Ergebnisse aus weiteren Untersuchungen ... 243
4.3.2.3.3.1 <Endpunkt xxx> – weitere Untersuchungen ... 243
4.3.2.3.3.2 Subgruppenanalysen – weitere Untersuchungen ... 244
4.4 Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens ... 244
4.4.1 Beurteilung der Aussagekraft der Nachweise ... 244
4.4.2 Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß ... 245
4.4.3 Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht ... 260
4.5 Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte ... 261
4.5.1 Begründung für die Vorlage indirekter Vergleiche ... 261
4.5.2 Begründung für die Vorlage nicht randomisierter vergleichender Studien und weiterer Untersuchungen ... 261
4.5.3 Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen ... 261
4.5.4 Verwendung von Surrogatendpunkten ... 262
4.6 Referenzliste ... 263
Anhang 4-A : Suchstrategien – bibliografische Recherche ... 269

Anhang 4-B : Suchstrategien – Suche in Studienregistern/Studienergebnisdatenbanken 272
Anhang 4-C : Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Recherche) 274
Anhang 4-D : Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern/Studienergebnisdatenbanken) 275
Anhang 4-E : Methodik der eingeschlossenen Studien – RCT 279
Studie: ROBIN, Teil B 280
Studie: SKYLARK 299
Anhang 4-F : Bewertungsbögen zur Einschätzung von Verzerrungsaspekten 314
Studie: ROBIN, Teil B 315
Studie: SKYLARK 319
Verzerrung auf Endpunktebene für die Studien ROBIN (Teil B) und SKYLARK 323
Anhang 4-G : Ergebnisse der Meta-Analyse 350
Anhang 4-H : Ergebnisse der Studie ROBIN 351
Anhang 4-I : Ergebnisse der Studie SKYLARK 352

Tabellenverzeichnis

Seite

Tabelle 4-1: Endpunkte mit signifikanten Ergebnissen zum ersten Messzeitpunkt 19
Tabelle 4-2: Endpunkte mit signifikanten Ergebnissen zum Ende der Kurzzeitbehandlung und zum Ende der Beobachtung sowie Ereigniszeitanalysen über den gesamten Zeitraum ... 21
Tabelle 4-3: Inzidenz von UE nach Zeit des Auftretens 30
Tabelle 4-4: Übersicht der Ein-/Ausschlusskriterien der berücksichtigten Studien 37
Tabelle 4-5: Demografischen und krankheitsspezifische Charakteristika die für die Studienpopulationen der Studien ROBIN und SKYLARK abgefragt wurden 47
Tabelle 4-6: Übersicht über patientenrelevante Endpunkte und Analysen 48
Tabelle 4-7: Prädefinierte Subgruppen und zugehörige Kategorien der Studien ROBIN und SKYLARK 63
Tabelle 4-8: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel 68
Tabelle 4-9: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel 68
Tabelle 4-10: Relevante Studien (auch laufende Studien) aus der Suche in Studienregistern/ Studienergebnisdatenbanken – RCT mit dem zu bewertenden Arzneimittel 70
Tabelle 4-11: Relevante Studien aus der Suche auf der Internetseite des Gemeinsamen Bundesausschusses – RCT mit dem zu bewertenden Arzneimittel 71
Tabelle 4-12: Studienpool – RCT mit dem zu bewertenden Arzneimittel 72
Tabelle 4-13: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel 74
Tabelle 4-14: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel 76
Tabelle 4-15: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel 77
Tabelle 4-16: Verteilung der Patienten nach Schwere der Erkrankung mittels MADRS-Gesamtwert zu Baseline in der ITT-Population der Studien ROBIN und SKYLARK 90
Tabelle 4-17: Tatsächliche Behandlungsdosis und -dauer in der Studie ROBIN 92
Tabelle 4-18: Tatsächliche Behandlungsdosis und -dauer in der Studie SKYLARK 92
Tabelle 4-19: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel 94
Tabelle 4-20: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel 96
Tabelle 4-21: Operationalisierung von Todesfällen 101
Tabelle 4-22: Bewertung des Verzerrungspotenzials für Todesfälle in RCT mit dem zu bewertenden Arzneimittel 101

Tabelle 4-23: Ergebnisse für Todesfälle aus RCT mit dem zu bewertenden Arzneimittel.... 102
Tabelle 4-24: Operationalisierung von HAMD-17 104
Tabelle 4-25: Bewertung des Verzerrungspotenzials für HAMD-17 in RCT mit dem zu bewertenden Arzneimittel 105
Tabelle 4-26: Rücklaufrate für HAMD-17 aus der Meta-Analyse mit dem zu bewertenden Arzneimittel..... 106
Tabelle 4-27: Ergebnisse für HAMD-17-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 106
Tabelle 4-28: Ergebnisse für Zeit bis zum ersten HAMD-17-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 109
Tabelle 4-29: Ergebnisse für HAMD-17-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 111
Tabelle 4-30: Ergebnisse für Zeit bis zur ersten HAMD-17-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 113
Tabelle 4-31: Operationalisierung von MADRS 116
Tabelle 4-32: Bewertung des Verzerrungspotenzials für MADRS in RCT mit dem zu bewertenden Arzneimittel 117
Tabelle 4-33: Rücklaufrate für MADRS aus der Meta-Analyse mit dem zu bewertenden Arzneimittel..... 118
Tabelle 4-34: Ergebnisse für MADRS-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 118
Tabelle 4-35: Ergebnisse für Zeit bis zum ersten MADRS-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 120
Tabelle 4-36: Ergebnisse für MADRS-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 122
Tabelle 4-37: Ergebnisse für Zeit bis zur ersten MADRS-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 123
Tabelle 4-38: Operationalisierung von EPDS 126
Tabelle 4-39: Bewertung des Verzerrungspotenzials für EPDS in RCT mit dem zu bewertenden Arzneimittel 127
Tabelle 4-40: Rücklaufrate für EPDS aus der Meta-Analyse mit dem zu bewertenden Arzneimittel..... 128
Tabelle 4-41: Ergebnisse für EPDS-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 128
Tabelle 4-42: Ergebnisse für Zeit bis zum ersten EPDS-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 131
Tabelle 4-43: Operationalisierung von PHQ-9 133
Tabelle 4-44: Bewertung des Verzerrungspotenzials für PHQ-9 in RCT mit dem zu bewertenden Arzneimittel 135

Tabelle 4-45: Rücklaufrate für PHQ-9 aus der Meta-Analyse mit dem zu bewertenden Arzneimittel ... 135
Tabelle 4-46: Ergebnisse für PHQ-9-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel ... 136
Tabelle 4-47: Ergebnisse für Zeit bis zum ersten PHQ-9-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel ... 138
Tabelle 4-48: Operationalisierung von HAM-A ... 141
Tabelle 4-49: Bewertung des Verzerrungspotenzials für HAM-A in RCT mit dem zu bewertenden Arzneimittel ... 143
Tabelle 4-50: Rücklaufrate für HAM-A aus der Meta-Analyse mit dem zu bewertenden Arzneimittel ... 143
Tabelle 4-51: Ergebnisse für HAM-A-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel ... 144
Tabelle 4-52: Ergebnisse für Zeit bis zum ersten HAM-A-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel ... 146
Tabelle 4-53: Operationalisierung von SF-36 ... 149
Tabelle 4-54: Bewertung des Verzerrungspotenzials für SF-36 in RCT mit dem zu bewertenden Arzneimittel ... 150
Tabelle 4-55: Rücklaufraten des SF-36 aus RCT mit dem zu bewertenden Arzneimittel ... 151
Tabelle 4-56: Ergebnisse für das PCS-Ansprechen auf dem SF-36 aus RCT mit dem zu bewertenden Arzneimittel ... 151
Tabelle 4-57: Ergebnisse für Zeit bis zum ersten PCS-Ansprechen des SF-36 aus RCT mit dem zu bewertenden Arzneimittel ... 153
Tabelle 4-58: Ergebnisse für MCS-Ansprechen auf dem SF-36 aus RCT mit dem zu bewertenden Arzneimittel ... 154
Tabelle 4-59: Ergebnisse für Zeit bis zum ersten MCS-Ansprechen auf dem SF-36 aus RCT mit dem zu bewertenden Arzneimittel ... 155
Tabelle 4-60: Ergebnisse für das Ansprechen auf der Subskala „Körperliche Schmerzen“ des SF-36 aus der RCT mit dem zu bewertenden Arzneimittel ... 157
Tabelle 4-61: Ergebnisse für Zeit bis zum ersten Ansprechen auf der Subskala „Körperliche Schmerzen“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel ... 159
Tabelle 4-62: Ergebnisse für das Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel ... 160
Tabelle 4-63: Ergebnisse für Zeit bis zum ersten Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel ... 162
Tabelle 4-64: Ergebnisse für das Ansprechen auf der Subskala „Psychisches Wohlbefinden“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel ... 163
Tabelle 4-65: Ergebnisse für Zeit bis zum ersten Ansprechen auf der Subskala „Psychisches Wohlbefinden“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel ... 165
Tabelle 4-66: Operationalisierung von BIMF ... 167

Tabelle 4-67: Bewertung des Verzerrungspotenzials für BIMF in RCT mit dem zu bewertenden Arzneimittel 169
Tabelle 4-68: Rücklaufrate für BIMF aus RCT mit dem zu bewertenden Arzneimittel..... 170
Tabelle 4-69: Ergebnisse für BIMF-Gesamtwert-Ansprechen aus RCT mit dem zu bewertenden Arzneimittel 170
Tabelle 4-70: Ergebnisse für Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen Gesamtwert aus RCT mit dem zu bewertenden Arzneimittel 173
Tabelle 4-71: Ergebnisse für Ansprechen auf der Subskala „Kompetenz der Mutter“ des BIMF aus RCT mit dem zu bewertenden Arzneimittel 175
Tabelle 4-72: Ergebnisse für Zeit bis zum ersten Ansprechen auf der Subskala „Kompetenz der Mutter“ des BIMF aus RCT mit dem zu bewertenden Arzneimittel 177
Tabelle 4-73: Ergebnisse für Ansprechen auf der Subskala „Bedürfnisse der Mutter“ aus RCT mit dem zu bewertenden Arzneimittel 179
Tabelle 4-74: Ergebnisse für Zeit bis zum ersten Ansprechen auf der Subskala „Bedürfnisse der Mutter“ des BIMF aus RCT mit dem zu bewertenden Arzneimittel 182
Tabelle 4-75: Operationalisierung von C-SSRS 184
Tabelle 4-76: Bewertung des Verzerrungspotenzials für C-SSRS in RCT mit dem zu bewertenden Arzneimittel 185
Tabelle 4-77: Ergebnisse für C-SSRS zu Baseline aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 186
Tabelle 4-78: Ergebnisse für C-SSRS an Tag 3 aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 187
Tabelle 4-79: Ergebnisse für C-SSRS an Tag 15 aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 187
Tabelle 4-80: Ergebnisse für C-SSRS bei der letzten Bewertung nach Baseline aus der Meta-Analyse mit dem zu bewertenden Arzneimittel..... 188
Tabelle 4-81: Operationalisierung von UE 190
Tabelle 4-82: Bewertung des Verzerrungspotenzials für Sicherheit in RCT mit dem zu bewertenden Arzneimittel 192
Tabelle 4-83: Ergebnisse für Anzahl der Patientinnen mit mindestens einem UE aus der Meta-Analyse mit dem zu bewertenden Arzneimittel..... 192
Tabelle 4-84: Ergebnisse für Anzahl der Patientinnen mit mindestens einem UE nach SOC aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 193
Tabelle 4-85: Ergebnisse für Anzahl der Patientinnen mit mindestens einem UE nach PT aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 197
Tabelle 4-86: Ergebnisse für Anzahl der Patientinnen mit mindestens einem schweren UE aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 200
Tabelle 4-87: Ergebnisse für Anzahl der Patientinnen mit mindestens einem SUE aus der Meta-Analyse mit dem zu bewertenden Arzneimittel..... 201
Tabelle 4-88: Ergebnisse für Therapieabbruch aufgrund von UE aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 202

Tabelle 4-89 Matrix der durchgeführten Subgruppenanalysen 205
Tabelle 4-90: Ergebnis des Interaktionsterms der Subgruppenanalysen je Endpunkt für die Meta-Analyse 206
Tabelle 4-91: Ergebnisse für Subgruppe BMI im Endpunkt HAMD-17-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 211
Tabelle 4-92: Ergebnisse für Subgruppe BMI im Endpunkt HAMD-17-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 215
Tabelle 4-93: Ergebnisse für Subgruppe BMI im Endpunkt Zeit bis zur ersten MADRS-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 219
Tabelle 4-94: Ergebnisse für Subgruppe BMI im Endpunkt Zeit bis zum ersten HAM-A-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel 222
Tabelle 4-95: Ergebnisse für Subgruppe „Schwere der Erkrankung“ im Endpunkt Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel 224
Tabelle 4-96: Ergebnisse für Subgruppe „Schwere der Erkrankung“ im Endpunkt BIMF-Gesamtwert-Ansprechen aus RCT mit dem zu bewertenden Arzneimittel 226
Tabelle 4-97: Ergebnisse für Subgruppe Abstammung im Endpunkt Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen aus RCT mit dem zu bewertenden Arzneimittel 229
Tabelle 4-98: Ergebnisse für Subgruppe Abstammung im Endpunkt Zeit bis zum ersten Ansprechen nach -Subskala „Kompetenz der Mutter“ des BIMF aus RCT mit dem zu bewertenden Arzneimittel 230
Tabelle 4-99: Ergebnisse für Subgruppe Beginn der PPD im Endpunkt Zeit bis zum ersten Ansprechen auf der Subskala „Bedürfnisse der Mutter“ des BIMF aus RCT mit dem zu bewertenden Arzneimittel 231
Tabelle 4-100: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche 234
Tabelle 4-101: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden 235
Tabelle 4-102: Operationalisierung von <Endpunkt xxx> 235
Tabelle 4-103: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche 236
Tabelle 4-104: Ergebnisse für <Endpunkt xxx> aus RCT für indirekte Vergleiche 236
Tabelle 4-105: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien 239
Tabelle 4-106: Matrix der Endpunkte in den eingeschlossenen nicht randomisierten vergleichenden Studien 240
Tabelle 4-107: Operationalisierung von <Endpunkt xxx> 240
Tabelle 4-108: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien 240
Tabelle 4-109: Matrix der Endpunkte in den eingeschlossenen weiteren Untersuchungen 243
Tabelle 4-110: Operationalisierung von <Endpunkt xxx> – weitere Untersuchungen 243

Tabelle 4-111: Signifikante Endpunkte zum ersten Messzeitpunkt 247
Tabelle 4-112: Zusammenfassung der Ergebnisse 249
Tabelle 4-113: Inzidenz von UE nach Zeit des Auftretens 257
Tabelle 4-114: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens 261
Tabelle 4-115 (Anhang): Studiendesign und -methodik für Studie ROBIN 280
Tabelle 4-116 (Anhang): Studiendesign und -methodik für Studie SKYLARK 299
Tabelle 4-117 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie ROBIN (Teil B) 315
Tabelle 4-118 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie SKYLARK 319

Abbildungsverzeichnis

Seite

Abbildung 4-1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel 69
Abbildung 4-2: Histogramm der MADRS-Gesamtwerte zu Baseline nach Schweregrad..... 91
Abbildung 4-3: Kaplan-Meier-Plot für Zeit bis zum ersten HAMD-17-Ansprechen..... 110
Abbildung 4-4: Kaplan-Meier-Plot für Zeit bis zur ersten HAMD-17-Remission..... 115
Abbildung 4-5: Kaplan-Meier-Plot für Zeit bis zum ersten MADRS-Ansprechen 121
Abbildung 4-6: Kaplan-Meier-Plot für Zeit bis zur ersten MADRS-Remission 125
Abbildung 4-7: Kaplan-Meier-Plot für Zeit bis zum ersten EPDS-Ansprechen..... 132
Abbildung 4-8: Kaplan-Meier-Plot für Zeit bis zum ersten PHQ-9-Ansprechen 140
Abbildung 4-9: Kaplan-Meier-Plot für Zeit bis zum ersten HAM-A-Ansprechen..... 148
Abbildung 4-10: Kaplan-Meier-Plot für Zeit bis zum ersten PCS-Ansprechen des SF-36 ... 154
Abbildung 4-11: Kaplan-Meier-Plot für Zeit bis zum ersten MCS-Ansprechen auf dem SF-36 157
Abbildung 4-12: Kaplan-Meier-Plot für Zeit bis zum ersten Ansprechen auf der Subskala „Körperliche Schmerzen“ des SF-36 160
Abbildung 4-13: Kaplan-Meier-Plot für Zeit bis zum ersten Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“ des SF-36 163
Abbildung 4-14: Kaplan-Meier-Plot für Zeit bis zum ersten Ansprechen auf der Subskala „Psychisches Wohlbefinden“ des SF-36 166
Abbildung 4-15: Kaplan-Meier-Plot für Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen..... 174
Abbildung 4-16: Kaplan-Meier-Plot für Zeit bis zum ersten Ansprechen auf der Subskala „Kompetenz der Mutter“ des BIMF..... 179
Abbildung 4-17: Kaplan-Meier-Plot für Zeit bis zum ersten Ansprechen auf der Subskala „Bedürfnisse der Mutter“ des BIMF 183
Abbildung 4-18: Patientenfluss der Studie ROBIN 298
Abbildung 4-19: Patientenfluss der Studie SKYLARK..... 313

Abkürzungsverzeichnis

Abkürzung	Bedeutung
ACOG	American College of Obstetricians and Gynecologists
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
APA	American Psychiatric Association
ARR	Absolute Risikoreduktion
BAI	Beck Anxiety Inventory
Bewertungsumfang	Bewertungsumfang sind gemäß Artikel 2 Nummer 9 der Verordnung (EU) 2021/2282 die von den Mitgliedsstaaten der Europäischen Union gemeinsam in Auftrag gegebenen Parameter für eine gemeinsame klinische Bewertung in Bezug auf Patientenpopulation, Intervention, Komparatoren und gesundheitsbezogene Endpunkte, die im Verfahren nach Artikel 8 Absatz 6 der Verordnung (EU) 2021/2282 festgelegt werden.
BIMF	Barkin Index of Maternal Functioning
BMI	Body-Mass-Index
BSC	Best Supportive Care
CANMAT	Canadian Network for Mood and Anxiety Treatments
CBT	Cognitive-Behavioral Therapy (Kognitive Verhaltenstherapie)
CGI-I	Clinical Global Impression-Improvement
CGI-S	Clinical Global Impression-Severity
CONSORT	Consolidated Standards of Reporting Trials
COST	European Cooperation in Science and Technology
CoxPH	Cox Proportional Hazards Modell
C-SSRS	Columbia Suicide Severity Rating Scale
CTCAE	Common Terminology Criteria for Adverse Events
CYP	Cytochrom P450
DGPPN	Deutsche Gesellschaft für Psychiatrie und Psychotherapie, Psychosomatik und Nervenheilkunde
DIMDI	Deutsches Institut für Medizinische Dokumentation
DSM-5	Diagnostic and Statistical Manual of Mental Disorders
EBM	Einheitlicher Bewertungsmaßstab
EG	Europäische Gemeinschaft
EKG	Elektrokardiogramm

EMA	European Medicines Agency
EPAR	European Public Assessment Report
EPDS	Edinburgh Postnatal Depression Scale
EU	Europäische Union
EU-Dossier	Europäisches Dossier sind die im nach Artikel 10 Absatz 2 der Verordnung (EU) 2021/2282 zur Durchführung einer gemeinsamen klinischen Bewertung vorgelegten Dossier enthaltenen und die nach Artikel 10 Absatz 5 Satz 2 der Verordnung (EU) 2021/2282, auf Aufforderung nach Artikel 11 Absatz 2 Satz 1 der Verordnung (EU) 2021/2282 oder in Folge einer Information nach Artikel 11 Absatz 2 Satz 3 der Verordnung (EU) 2021/2282 nachgereichten Informationen, Daten, Analysen und sonstigen Nachweise.
FAS	Full Analysis Set
GABA	Gamma-Aminobutyric Acid
GAD-7	Generalized Anxiety Scale
G-BA	Gemeinsamer Bundesausschuss
GEE	Generalized Estimating Equation
Gemeinsame klinische Bewertung	Gemeinsame klinische Bewertung eines Arzneimittels im Sinne des Artikels 2 Nummer 6 der Verordnung (EU) 2021/2282 des Europäischen Parlaments und des Rates vom 15. Dezember 2021 über die Bewertung von Gesundheitstechnologien und zur Änderung der Richtlinie 2011/24/EU (ABl. L 458 vom 22.12.2021, S. 1; L, 2024/90313, 28.5.2024) nach den Vorgaben der Verordnung (EU) 2021/2282
GKV	Gesetzliche Krankenversicherung
HAM-A	Hamilton Rating Scale for Anxiety
HAMD-17	Hamilton Rating Scale for Depression
HBsAg	Hepatitis-B-Oberflächenantigen
HCRU	Healthcare Resource Utilization
HCV	Hepatitis-C-Virus
HIV	Humanes Immundefizienz Virus
HPA-Achse	Hypothalamus-Hypophysen-Nebennierenachse
HR	Hazard Ratio
ICD	International Classification of Diseases
ICF	Einwilligungsformular (Informed Consent Form)
IPD	Individuelle Patientendaten
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

IRB	Ethikkommission (Institutional Review Board)
IRT	Interactive Response Technology
ITT	Intention to treat
IU	International Unit
KI	Konfidenzintervall
LLT	Lower Level Term
LPFV	Last patient first visit
LPI	Last patient in
MADRS	Montgomery-Asberg Depression Rating Scale
MCID	Minimal Clinically Important Difference
MCS	psychischer Summenscore (Mental Component Summary)
MDD	Major Depressive Disorder
MedDRA	Medical Dictionary for Regulatory Activities
MHRA	Britische Arzneimittel- und Gesundheitsbehörde (Medicines and Healthcare Products Regulatory Agency)
MMRM	Mixed effect Model Repeat Measurement
MTC	Mixed Treatment Comparison
NVL	Nationale Versorgungsleitlinie
OR	Odds Ratio
PCS	körperlicher Summenscore (Physical Component Summary)
PHQ-9	Patient Health Questionnaire-9
PiT	Patientenindividuelle Therapie
PPD	Postpartum/Peripartum Depression
PT	Preferred Terms nach MedDRA
PWC-20	20-Item Physician Withdrawal Checklist
RAAD	Rapid-Acting Antidepressants
RCT	Randomized Controlled Trial
Riseup-PPD	Research Network in Peripartum Depression Disorder
RR	Relatives Risiko
SCID-5-CT	Structured Clinical Interview for DSM-5 Clinical Trials Version
SD	Standardabweichung (Standard Deviation)
SF-36	Short Form 36
SGB	Sozialgesetzbuch

SMQs	Standardised MedDRA Queries
SOC	System Organ Class nach MedDRA
SSNRI	Selektiver Serotonin-Noradrenalin-Wiederaufnahmehemmer
SSRI	Selektive Serotonin-Wiederaufnahmehemmer
SSS	Stanford Sleepiness Scale
STE	Surrogate Threshold Effects
STROBE	Strengthening the Reporting of Observational Studies in Epidemiology
SUE	Schwerwiegendes UE
TEAE	Treatment Emergent Adverse Event
TREND	Transparent Reporting of Evaluations with Non-Randomized Design
TTE	Time to event
UE	Unerwünschtes Ereignis
VerfO	Verfahrensordnung des Gemeinsamen Bundesausschusses
Verordnung (EU) 2021/2282	Verordnung (EU) 2021/2282 des Europäischen Parlaments und des Rates vom 15. Dezember 2021 über die Bewertung von Gesundheitstechnologien und zur Änderung der Richtlinie 2011/24/EU
WHO	World Health Organization
zVT	Zweckmäßige Vergleichstherapie

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.

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

Im Falle einer vorangegangenen gemeinsamen klinischen Bewertung nach der Verordnung (EU) 2021/2282 müssen pharmazeutische Unternehmen keine Informationen, Daten, Analysen oder sonstige Nachweise vorlegen, die bereits auf Unionsebene vorgelegt wurden.

Wurde für ein Arzneimittel ein EU-Dossier vorgelegt und wurde die gemeinsame klinische Bewertung des Arzneimittels nicht nach Artikel 10 Absatz 6 Satz 1 der Verordnung (EU) 2021/2282 eingestellt, hat der pharmazeutische Unternehmer gemäß 5. Kapitel § 9 Absatz 2a VerfO anzugeben, ob und welche Nachweise aus dem EU-Dossier Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, indem er durch Verweise in den betroffenen Abschnitten des vorliegenden Dossiers auf diese Nachweise Bezug nimmt.

Die Verweise sind dabei bis zur untersten vorhandenen Gliederungsebene und auf Abschnittsebene zu spezifizieren. Bei Verweisen auf Tabellen oder Abbildungen ist zusätzlich die jeweilige Tabellen- beziehungsweise Abbildungsnummerierung anzugeben.

Sind in Fällen einer vorangegangenen gemeinsamen klinischen Bewertung nach der Verordnung (EU) 2021/2282 Angaben zum Nachweis des Zusatznutzens gegenüber der zweckmäßigen Vergleichstherapie bisher teilweise oder vollständig nicht im EU-Dossier vorgelegt worden, so sind diese Angaben in den betroffenen Abschnitten des Moduls 4 jeweils zu ergänzen beziehungsweise die jeweilige Datei in Modul 5 vorzulegen.

Die Abschnitte 4.2.1 und 4.4, die sich auf die Fragestellung für die Nutzenbewertung nach § 35a SGB V in Bezug auf den deutschen Versorgungskontext beziehungsweise auf die abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens beziehen, sind unabhängig von einer vorangegangenen gemeinsamen klinischen Bewertung nach der Verordnung (EU) 2021/2282 ohne Verweise auszufüllen.

Sofern für ein Arzneimittel bis zum für die Einreichung des nationalen Dossiers maßgeblichen Zeitpunkt kein europäisches Dossier vorgelegt wurde oder die gemeinsame klinische Bewertung des Arzneimittels nach Artikel 10 Absatz 6 Satz 1 Verordnung (EU) 2021/2282 eingestellt wurde, sind Verweise auf bereits im EU-Dossier vorgelegte Informationen, Daten, Analysen oder sonstige Nachweise nicht möglich. In diesem Fall hat der pharmazeutische Unternehmer alle erforderlichen Angaben in Modul 4 ohne Verweise auszufüllen und die zugehörigen Dateien in Modul 5 vorzulegen.

4.1 Zusammenfassung der Inhalte von Modul 4

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

Fragestellung

Das Ziel des Dossiers ist die Quantifizierung des Zusatznutzens von Zuranolon bei erwachsenen Patientinnen mit postpartaler Depression (PPD) nach Entbindung. Das Anwendungsgebiet umfasst zwei Patientenpopulationen mit entsprechender zweckmäßiger Vergleichstherapie (zVT):

a) Für Patientinnen mit **leichter** depressiver Symptomatik nach der Geburt bestimmt der Gemeinsame Bundesausschuss (G-BA) Best Supportive Care (BSC) als zVT.

b) Für Patientinnen mit **mittelschwerer bis schwerer** depressiver Symptomatik bestimmt der G-BA individualisierte Therapie unter Auswahl von BSC und einem selektiven Serotonin-Wiederaufnahmehemmer (SSRI). Die Studien ROBIN und SKYLARK können meta-analytisch zusammengefasst werden und belegen einen Zusatznutzen für Zuranolon im Vergleich zur vom G-BA festgelegten zVT für die Patientenpopulation b) [1]. Die Ergebnisse werden in den vier patientenrelevanten Dimensionen Mortalität, Morbidität, Lebensqualität und Sicherheit dargestellt.

Datenquellen

Für die Patientenpopulation a) mit leichter depressiver Symptomatik, die in der Regel nicht medikamentös behandelt wird, können keine relevanten Datenquellen identifiziert werden. Für die Population b) mit mittelschwerer bis schwerer depressiver Symptomatik konnten im Rahmen der systematischen Literaturrecherche in Studienregistern zwei Studien identifiziert werden. Die Studien ROBIN und SKYLARK sind randomisierte, kontrollierte klinische Studien der Phase 3, die in einer Meta-Analyse zusammengefasst wurden und dienen dem Nachweis eines Zusatznutzens von Zuranolon im Vergleich zur vom G-BA festgelegten zVT.

Ein-/Ausschlusskriterien für Studien

Relevante Studien wurden anhand definierter Ein- und Ausschlusskriterien (siehe Tabelle 4-4) für die Bewertung selektiert. Diese Kriterien grenzen die Studien hinsichtlich der Patientenpopulation, der Intervention, der Vergleichstherapie, den Endpunkten, dem Studientyp und der Studiendauer ein.

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

Die Aussagesicherheit der erbrachten Nachweise wurde auf der Basis der Vorgaben der Dossiervorlage sowie der Vorgaben in Anhang 4-F bewertet. Die für die Bewertung relevanten Studien wurden gemäß den Anforderungen des Consolidated Standards Of Reporting Trials (CONSORT)-Statements für randomisierte Studien beschrieben (Anhang 4-E). Die Studienpopulation wurde anhand demografischer Daten sowie krankheitsspezifischer Charakteristika beschrieben.

Da zwei geeignete Studien vorliegen, wurde eine Meta-Analyse durchgeführt. Hierfür wurden die relevanten Studien hinsichtlich ihrer Vergleichbarkeit in Bezug auf Fragestellung, Studiendesign, Studienpopulation, Kontrollarm und Endpunkte untersucht. Die Studien ROBIN und SKYLARK weisen ähnliche Protokolle und Einschlusskriterien auf und vergleichen die Sicherheit und Wirksamkeit von Zuranolon in Kombination mit einer patientenindividuellen Therapie (PiT) gegenüber einem PiT-Vergleichsarm. Die PiT entspricht der definierten zVT, da die Patientinnen SSRI sowie weitere supportive Medikation bzw. Therapien – entsprechend BSC – erhielten. Die Meta-Analyse basiert auf individuellen Patientendaten (IPD), was als Goldstandard gilt, und verwendet feste Effekte aufgrund der homogenen Studienlage. Die Analyseverfahren umfassen Mixed Model for Repeated Measures (MMRM) für stetige Variablen, ein Generalisiertes Schätzgleichungsmodell (GEE) für binäre Variablen und ein Cox-Proportional-Hazard-Modell (CoxPH) für Ereigniszeit-Analysen. Damit liegen Ergebnisse mit höchster Ergebnissicherheit vor, die gemäß § 5 Abs. 6 AM-NutzenV der Evidenzstufe Ia entsprechen.

Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen

Zur Bewertung und Quantifizierung des medizinischen Nutzens und Zusatznutzens werden im vorliegenden Modul 4A Ergebnisse patientenrelevanter Endpunkte der Nutzendimensionen Mortalität, Morbidität, gesundheitsbezogene Lebensqualität und Sicherheit der Zulassungsstudien ROBIN und SKYLARK herangezogen.

Beide Studien untersuchten die Wirksamkeit und Sicherheit von Zuranolon zur Behandlung mittelschwerer bis schwerer PPD als randomisierte, doppelblinde, kontrollierte Parallelgruppenstudien mit einer 14-tägigen Behandlungsdauer und einer Nachbeobachtung bis Tag 45. Es wurden Patientinnen mit einem Hamilton Rating Scale for Depression (HAMD)-17-Wert von ≥ 26 eingeschlossen, was auf eine schwere Depression hinweist. Zusätzliche

Auswertungen nach Montgomery-Asberg Depression Rating Scale (MADRS) beschreiben ein Spektrum von mittelschweren bis schweren Symptomen. Die Studienpopulationen sind demografisch und krankheitsspezifisch vergleichbar.

Obwohl die Datenlage zur PPD in Deutschland begrenzt ist, stimmen das Durchschnittsalter der Patientinnen und die Diagnosemethoden mit deutschen Standards überein. Daher wird von einer Übertragbarkeit auf den deutschen Versorgungskontext ausgegangen.

Die Ableitung des Zusatznutzens von Zuranolon für die Teilpopulation b) (mittelschwer – schwer) beruht maßgeblich auf den Ergebnissen zur Mortalität, Morbidität und Sicherheit aus einer Meta-Analyse der beiden Studien ROBIN und SKYLARK. Zusätzlich zu Wirksamkeit und Sicherheit liegen aus der Studie ROBIN Daten für die Lebensqualität vor. Die Patientinnen der Teilpopulation a) (mild) waren in den Studien nicht eingeschlossen, da diese oft durch therapeutische Interventionen oder unterstützende Maßnahmen wie körperliche Aktivität und nicht medikamentös behandelt werden.

Die vorliegenden Daten, bestehend aus einer Meta-Analyse der höchsten Evidenzstufe Ia sowie ergänzend statistisch signifikante Ergebnisse auch auf Ebene der Einzelstudien (siehe Anhang 4-H und Anhang 4-I), verdeutlichen eindrucksvoll den Zusatznutzen der Zuranolon+PiT-Therapie bei der mittelschweren – schweren PPD.

Insbesondere der schnelle Wirkungseintritt bereits zu Beginn der Behandlung ermöglicht eine rasche und für die Patientinnen deutlich wahrnehmbare Reduktion der depressiven Symptomatik und trägt maßgeblich zur Verbesserung der Lebensqualität betroffener Mütter bei. Damit trägt Zuranolon der besonderen Akuität der PPD Rechnung, die in einer Phase hoher psychischer und funktioneller Vulnerabilität der Mutter auftritt und eine zeitnahe therapeutische Intervention erfordert.

Im Folgenden fassen Tabelle 4-1 und Tabelle 4-2 die Ergebnisse zusammen, die anschließend bezüglich der medizinischen Bedeutung und Ausmaß des Zusatznutzens in Kontext gesetzt werden.

Schnell einsetzende Wirkung stellt einen Zusatznutzen dar

Die PPD stellt eine Erkrankung dar, die in eine Phase maximaler psychischer, funktioneller und biologischer Vulnerabilität der Mutter fällt. Vor diesem Hintergrund ist ein rascher Wirkungseintritt ein zentrales Kriterium für die Wahl einer geeigneten Therapie [2; 3]. Studien zeigen, dass die Symptome einer PPD bei einem Großteil der Patientinnen innerhalb von 3 bis 6 Monaten abklingen [4; 5]. Vor diesem Hintergrund fügt sich eine 14-tägige Therapie mit akutem, klar zeitlich begrenztem Behandlungsverlauf wie Zuranolon gut in das klinische Bild der Erkrankung ein. PPD ist einerseits eine akut auftretende und zeitkritische Erkrankung, erfordert andererseits nicht in allen Fällen eine langfristige Pharmakotherapie [4; 5]. Zuranolon ist aufgrund seiner Eigenschaft als Rapid-Acting Antidepressant (RAAD) und seiner spezifischen Entwicklung für die Pathophysiologie der PPD als einmalige 14-tägige orale Kurzzeittherapie konzipiert.

Die Ergebnisse in Tabelle 4-1 zeigen, dass die Kombination von Zuranolon mit PiT im Vergleich zu PiT allein eine schnelle und statistisch signifikante Verbesserung der depressiven Symptomatik bewirkt, und somit die Akutphase drastisch verkürzt. Schon an Tag 3 war der Unterschied zwischen den Behandlungsarmen für mehrere Endpunkte statistisch signifikant.

Tabelle 4-1: Endpunkte mit signifikanten Ergebnissen zum ersten Messzeitpunkt

Endpunkt	**Art der Analyse**	**Meta-Analyse Zuranolon+PiT (N = 176) vs. PiT (N = 177) Effektmaß [95 %-KI], p-Wert**	**Ableitung des Ausmaßes des Zusatznutzens**
Morbidität			
Depressive Symptomatik (HAMD-17, MADRS) - Symptomverbesserung und Remission unter Zuranolon			
HAMD-17-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen Tag 3	RR: 1,723 [1,186; 2,503], 0,0043	Beträchtlich
HAMD-17-Remission	Anteil der Studienteilnehmerinnen mit Ansprechen Tag 3	RR: 2,414 [1,515; 5,064], 0,0197	Beträchtlich
MADRS-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen Tag 8	RR: 1,423 [1,125; 1,799], 0,0032	Beträchtlich
Patientenberichtete Endpunkte (EPDS) - Verbesserung der selbstberichteten PPD-Symptome unter Zuranolon			
EPDS-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen Tag 3	RR: 1,454 [1,031; 2,049], 0,0326	Gering
Angstsymptomatik (HAM-A) - Reduktion begleitender Angstsymptome unter Zuranolon			
HAM-A-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen Tag 3	RR: 1,304 [1,009; 1,686], 0,0423	Gering
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.			

Insbesondere die schnelle Ansprechrate bei der HAMD-17 und MADRS sowie das frühzeitige Erreichen einer Remission bei der HAMD-17 in **beträchtlichem Ausmaß** unterstreichen die Bedeutung der Behandlung für Patientinnen mit mittelschwerer bis schwerer PPD. Die schnelle Wirkung ist entscheidend, um die zeitkritischen Auswirkungen der Erkrankung auf die Mutter-Kind-Beziehung und die familiäre Dynamik zu minimieren. Neben den eindrücklichen Ergebnissen bereits am ersten Messzeitpunkt, verdeutlichen auch die Ergebnisse zum Ende der

Kurzzeitbehandlung, am Tag 15, und zum Ende der Beobachtung an Tag 45 die schnelle Verbesserung der Symptome und den deutlichen Vorteil der Zuranolon+PiT-Behandlung.

Tabelle 4-2: Endpunkte mit signifikanten Ergebnissen zum Ende der Kurzzeitbehandlung und zum Ende der Beobachtung sowie Ereigniszeitanalysen über den gesamten Zeitraum

Endpunkt	**Art der Analyse**	**Meta-Analyse** **Zuranolon+PiT (N = 176) vs. PiT (N = 177)** **Effektmaß [95 %-KI], p-Wert**		**Ableitung des Ausmaßes des Zusatznutzens**	
		Tag 15 Ende der Kurzzeitbehandlung	Tag 45 Ende der Beobachtung	Tag 15	Tag 45
Mortalität					
Anzahl der Todesfälle		In beiden Studien wurden keine Todesfälle beobachtet.			
Morbidität					
Depressive Symptomatik (HAMD-17, MADRS) - Schnellere und häufigere Symptomverbesserung und Remission unter Zuranolon					
HAMD-17-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,507 [1,219; 1,863], 0,0002	RR: 1,269 [1,051; 1,532], 0,0130	Beträchtlich	Gering
	Zeit bis zum Ansprechen	HR: 1,854 [1,4372; 2,3920], < 0,0001		Erheblich	
HAMD-17-Remission	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,773 [1,222; 2,574], 0,0026	RR: 1,634 [1,213; 2,200], 0,0012	Beträchtlich	Beträchtlich
	Zeit bis zur Remission	HR: 1,753 [1,2818; 2,3962], 0,0004		Beträchtlich	
MADRS-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,528 [1,233; 1,894], 0,0001	RR: 1,315 [1,086; 1,592], 0,0050	Beträchtlich	Gering
	Zeit bis zum Ansprechen	HR: 1,497 [1,1630; 1,9258], 0,0017		Beträchtlich	
MADRS-Remission	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,690 [1,258; 2,271], 0,0005	RR: 1,409 [1,089; 1,823], 0,0091	Beträchtlich	Gering
	Zeit bis zur Remission	HR: 1,683 [1,2676; 2,2350], 0,0003		Beträchtlich	

Endpunkt	**Art der Analyse**	**Meta-Analyse** **Zuranolon+PiT (N = 176) vs. PiT (N = 177)** **Effektmaß [95 %-KI], p-Wert**		**Ableitung des Ausmaßes des Zusatznutzens**	
Patientenberichtete Endpunkte (EPDS, PHQ-9) - Schnellere und häufigere Verbesserung der selbstberichteten PPD-Symptome unter Zuranolon					
EPDS-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,169 [1,018; 1,342], 0,0272	RR: 1,162 [1,017; 1,328], 0,0269	Gering	Gering
	Zeit bis zum Ansprechen	HR: 1,244 [0,9933; 1,5575], 0,0572		Nicht belegt	
PHQ-9-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,177 [1,025; 1,352], 0,0209	RR: 1,148 [1,001; 1,315], 0,0480	Gering	Gering
	Zeit bis zum Ansprechen	HR: 1,306 [1,0426; 1,6348], 0,0201		Gering	
Angstsymptomatik (HAM-A) - Schnellere und häufigere Reduktion begleitender Angstsymptome unter Zuranolon					
HAM-A-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,323 [1,129; 1,552], 0,0006	RR: 1,178 [1,014; 1,369], 0,0333	Beträchtlich	Gering
	Zeit bis zum Ansprechen	HR: 1,356 [1,0738; 1,7128], 0,0105		Gering	
Lebensqualität					
SF-36 und BIMF (nur in Studie ROBIN erhoben) - Häufigere Verbesserung körperlicher Schmerzen und sozialer Funktion unter Zuranolon					
SF-36 PCS-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,500 [0,718; 3,133], 0,2802	RR: 1,543 [0,659; 3,615], 0,0322	Nicht belegt	Nicht belegt
	Zeit bis zum Ansprechen	HR: 1,619 [0,7953; 3,2953], 0,1841		Nicht belegt	
SF-36 MCS-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,058 [0,835; 1,341], 0,6393	RR: 1,223 [0,995; 1,502], 0,0555	Nicht belegt	Nicht belegt
	Zeit bis zum Ansprechen	HR: 1,051 [0,7375; 1,4967], 0,7845		Nicht belegt	

Endpunkt	**Art der Analyse**	**Meta-Analyse** **Zuranolon+PiT (N = 176) vs. PiT (N = 177)** **Effektmaß [95 %-KI], p-Wert**		**Ableitung des Ausmaßes des Zusatznutzens**	
SF-36 “Körperliche Schmerzen”-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,204 [0,877; 1,652], 0,2513	RR: 1,461 [1,051; 2,032], 0,0243	Nicht belegt	Gering
	Zeit bis zum Ansprechen	HR: 1,229 [0,8214; 1,8390], 0,3159		Nicht belegt	
SF-36 “Soziale Funktionsfähigkeit”-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,149 [0,899; 1,470], 0,2671	RR: 1,218 [1,011; 1,467], 0,0378	Nicht belegt	Gering
	Zeit bis zum Ansprechen	HR: 1,102 [0,7751; 1,5675], 0,5879		Nicht belegt	
SF-36 “Psychisches Wohlbefinden”-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,149 [0,918; 1,439], 0,2253	RR: 1,240 [1,028; 1,497], 0,0249	Nicht belegt	Gering
	Zeit bis zum Ansprechen	HR: 1,022 [0,7245; 1,4427], 0,8999		Nicht belegt	
BIMF-Gesamtwert-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,616 [1,040; 2,513], 0,0330	RR: 1,223 [0,914; 1,636], 0,1753	Gering	Nicht belegt
	Zeit bis zum Ansprechen	HR: 1,040 [0,7028; 1,5387], 0,8449		Nicht belegt	
BIMF-“Kompetenz der Mutter”-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,142 [0,725; 1,800], 0,5668	RR: 1,373 [0,921; 2,046], 0,1194	Nicht belegt	Nicht belegt
	Zeit bis zum Ansprechen	HR: 0,986 [0,6289; 1,5444], 0,9492		Nicht belegt	
BIMF-“Bedürfnisse der Mutter”-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,054 [0,747; 1,487], 0,7654	RR: 1,165 [0,905; 1,501], 0,2364	Nicht belegt	Nicht belegt
	Zeit bis zum Ansprechen	HR: 0,912 [0,6241; 1,3316], 0,6322		Nicht belegt	

Endpunkt	**Art der Analyse**	**Meta-Analyse** **Zuranolon+PiT (N = 176) vs. PiT (N = 177)** **Effektmaß [95 %-KI], p-Wert**		**Ableitung des Ausmaßes des Zusatznutzens**	
Relevante Ergebnisse von Kliniker*innen bewertet bezüglich Schweregrad (CGI-S) der PPD und Verbesserung des Zustands (CGI-I)					
CGI-S-Ansprechen [1]	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,348 [1,113; 1,633], 0,0023	RR: 1,271 [1,073; 1,505], 0,0054	Beträchtlich	Gering
	Zeit bis zum Ansprechen	HR: 1,688 [1,3209; 2,1562], < 0,0001		Beträchtlich	
CGI-I-Ansprechen [1]	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,393 [1,151; 1,686], 0,0007	RR: 1,181 [0,992; 1,406], 0,0615	Beträchtlich	Nicht belegt
	Zeit bis zum Ansprechen	HR: 1,660 [1,3042; 2,1135], < 0,0001		Beträchtlich	
Sicherheit [2]					
C-SSRS (Frage 1 – 10)	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 0,95 [0,388; 2,339], 0,9157	RR: 0,43 [0,135; 1,367], 0,1525	Nicht belegt	
UE		RR: 1,21 [1,009; 1,449], 0,0396		Ergänzend dargestellt	
SOC Erkrankungen des Nervensystems		RR: 1,66 [1,230; 2,248], 0,0010		Geringerer	
PT Somnolenz		RR: 2,84 [1,569; 5,142], 0,0006		Geringerer Transient zu Beginn der Behandlung (vorwiegend Tag 1-2)	
PT Sedierung		RR: 14,57 [1,946; 109,126], 0,0091		Geringerer Transient zu Beginn der Behandlung (vorwiegend Tag 1-2)	
Schwere UE		RR: 1,46 [0,419; 5,074], 0,5540		Nicht belegt	
SUE		RR: 2,91 [0,306; 27,748], 0,3521		Nicht belegt	
Therapieabbruch aufgrund von UE		RR: 2,43 [0,478; 12,351], 0,2848		Nicht belegt	

Endpunkt	**Art der Analyse**	**Meta-Analyse** **Zuranolon+PiT (N = 176) vs. PiT (N = 177)** **Effektmaß [95 %-KI], p-Wert**	**Ableitung des Ausmaßes des Zusatznutzens**
[1] Das CGI-S-Ansprechen war definiert als Reduktion von ≥ 1,05 Punkte (15 % der Skalenspannweite 0 – 7) von Baseline. Das CGI-I-Ansprechen war definiert als Antwort „stark verbessert“ und „sehr stark verbessert“ der 7 möglichen Antworten. Alle Analysen zu CGI-Endpunkten befinden sich in Anhang 4-G. [2] Die Sicherheitsanalysen wurden für die Sicherheits-Population durchgeführt: Zuranolon+PiT (N = 176) und PiT (N = 171). Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.			

Mortalität

Die Mortalität wird in den Studien durch die Anzahl der während des Studienzeitraums aufgetretenen Todesfälle erfasst. Eine metaanalytische Zusammenfassung ergab sich nicht, da in beiden Studien keine Todesfälle auftraten.

Morbidität

Die Morbidität wird durch verschiedene Instrumente gemessen. Der primäre Endpunkt beider Studien war die HAMD-17 am Tag 15 (Ende der Kurzzeitbehandlung). Die HAMD-17 bewertet die Schwere depressiver Symptome anhand von 17 Items, die sowohl psychopathologische als auch somatische Aspekte wie Schlafstörungen oder Suizidgedanken umfassen. Höhere Werte auf der Skala (0 – 52) deuten auf schwerere Symptome hin. Die Ergebnisse für die HAMD-17 zeigen, dass die Kombination von Zuranolon mit PiT im Vergleich zu PiT allein statistisch signifikante Verbesserungen bei der Ansprechrate erzielte. Das Ansprechen war *a priori* definiert als Reduktion von ≥ 50 % gegenüber dem Gesamtwert zu Baseline. Zum Ende der 14-tägigen Kurzzeitbehandlung mit Zuranolon betrug das Relative Risiko (RR) 1,507 mit einem p-Wert von 0,0002 und einem 95 %-Konfidenzintervall (KI) von [1,219; 1,863], was auf einen **beträchtlichen** Zusatznutzen hinweist. Zusätzlich zeigen die Analysen zur Zeit bis zum HAMD-17-Ansprechen einen statistisch signifikanten Vorteil zugunsten von Zuranolon und PiT. Patientinnen unter Zuranolon sprachen deutlich schneller an (Median 8 vs. 20 Tage), was einer signifikant erhöhten Rate des Ansprechens über die Zeit um 85,4 % entspricht (HR = 1,854 [1,4372; 2,3920], p < 0,0001). Dieser Effekt entspricht einem schnellen Ansprechen in **erheblichem** Ausmaß und blieb bis zum Ende der Beobachtung an Tag 45 bestehen.

Die MADRS findet häufig Anwendung in der klinischen Praxis und konzentriert sich auf zentrale depressive Merkmale wie Traurigkeit und Suizidgedanken. Sie misst Veränderungen der Symptomatik mit 10 Items auf einer Skala von 0 – 60. Das MADRS-Ansprechen war auch *a priori* definiert als Reduktion von ≥ 50 % gegenüber dem Gesamtwert zu Baseline. Im Zuranolon+PiT-Arm sprachen statistisch signifikant mehr Patientinnen an als im Vergleichsarm. Zum Ende der Behandlung lag die Wahrscheinlichkeit, eine relevante Verbesserung, also ein Ansprechen auf der MADRS zu erreichen bei der Behandlung mit Zuranolon in Kombination mit PiT um 52,8 % höher als im Vergleichsarm (RR: 1,528 [1,233; 1,894], p = 0,0001). Dies gilt als **beträchtlicher** Zusatznutzen. Auch hier wurde ein Ansprechen statistisch signifikant – um 49,7 % (8,5 vs. 17 Tage im Median) – schneller im Zuranolon+PiT-Arm als im Vergleichsarm erreicht (HR: 1,497 [1,1630; 1,9258], p = 0,0017), was einem ebenfalls **beträchtlichen** Zusatznutzen entspricht. Diese schnell eintretende, deutlich überlegene Wirksamkeit von Zuranolon+PiT gegenüber der Vergleichstherapie wird durch den anhaltenden, statistisch signifikanten Effekt untermauert, der bis zum Abschluss der Beobachtungsphase an Tag 45 nachweisbar ist.

Auf den Instrumenten HAMD-17 und MADRS wurde nicht nur das Ansprechen, sondern auch die Remission analysiert. Eine Remission war auf der HAMD-17 definiert als Wert von ≤ 7 Punkten und bei der MADRS ≤ 10 Punkten. Übereinstimmend zeigten beide Instrumente zum Ende der Behandlung einen statistisch signifikanten Vorteil zugunsten der Zuranolon+PiT-

Behandlung. Zum Ende der Behandlung lag die Wahrscheinlichkeit, eine Remission nach HAMD-17 bzw. MADRS zu erreichen bei der Behandlung mit Zuranolon in Kombination mit PiT um 77,3 % bzw. 69 % höher als im Vergleichsarm (RR: 1,773 [1,222; 2,574], p = 0,0026 bzw. RR: 1,690 [1,258; 2,271], p = 0,0005). Die Dauer bis zur Remission unterschied sich dabei auch statistisch signifikant und wurde im Zuranolon+PiT-Arm schneller erreicht (HAMD-17 um 75,3 %, 29 vs. 50 Tage; MADRS um 68,3 %, 16 vs. 56 Tage), als im Vergleichsarm (HR: 1,753 [1,2818; 2,3962], p = 0,0004 bzw. HR: 1,683 [1,2676; 2,2350], p = 0,0003). Mit einer statistisch signifikanten Überlegenheit gegenüber dem Vergleichsarm bezüglich der Remission und dem schnellen Einsetzen dieser – beides mit dem Ausmaß **beträchtlich** – lindert Zuranolon die Symptome der PPD gemessen mittels HAMD-17 und MADRS und ermöglicht Patientinnen schnell eine Remission zu erreichen. Die statistisch signifikante Überlegenheit der Zuranolon+PiT-Behandlung blieb bis zum Ende der Beobachtung mit einem **beträchtlichen** bzw. **geringen** Ausmaß für die Remission im HAMD-17 bzw. MADRS erhalten.

Die Edinburgh Postnatal Depression Scale (EPDS) ist ein selbstberichteter Fragebogen, der speziell für die Erkennung und Bewertung der PPD entwickelt wurde. Mit 10 Items und einer Skala von 0 – 30 bezieht sie sich auf emotionale Symptome. Das Ansprechen, heißt eine klinisch relevante Verbesserung, war für die EPDS – in Ermangelung prädefinierter Grenzen – definiert als Zunahme um ≥ 15 % der Skalenspannweite. Für das EPDS-Ansprechen zeigte sich zu Behandlungsende (RR: 1,169 [1,018; 1,342], p = 0,0272) sowie zum Ende der Beobachtung (RR: 1,162 [1,017; 1,328], p = 0,0269) ein statistisch signifikanter Vorteil der Zuranolon+PiT-Behandlung gegenüber dem Vergleichsarm. Dies entspricht einem Zusatznutzen in **geringem** Ausmaß. Die Zeit bis zum EPDS-Ansprechen zeigte für beide Behandlungsgruppen keinen statistisch signifikanten Unterschied.

Der Patient Health Questionnaire (PHQ)-9, ein weiteres Selbstbeurteilungsinstrument, erfasst die Schwere depressiver Symptome anhand von 9 Items (Skala 0 – 27) und legt besonderen Wert auf Symptome wie Anhedonie und Suizidgedanken. Eine relevante Verbesserung der Symptome war definiert als Zunahme um ≥ 15 % der Skalenspannweite und wurde als Ansprechen ausgewertet. Für das PHQ-9-Ansprechen zeigte sich zu Behandlungsende ein statistisch signifikanter Vorteil der Zuranolon+PiT-Behandlung gegenüber dem Vergleichsarm (RR: 1,177 [1,025; 1,352], p = 0,0209). Die Nachhaltigkeit der Verbesserung und Überlegenheit gegenüber dem Vergleichsarm zeigte sich auch zum Ende der Beobachtung (RR: 1,148 [1,001; 1,315], p = 0,0480). Das Ansprechen wurde statistisch signifikant – um 30,6 % (5 vs. 7 Tage zum 25. Perzentil) – schneller im Zuranolon+PiT-Arm als im Vergleichsarm erreicht (HR: 1,306 [1,0426; 1,6348], 0,0201). Dies entspricht einem Zusatznutzen in **geringem** Ausmaß.

Zu den typischen Symptomen einer PPD gehören Angststörungen, insbesondere Panikattacken und übersteigerte Sorgen um das Wohl des Kindes bzw. für das Wohl des Kindes nicht ausreichend sorgen zu können. Die Hamilton Rating Scale for Anxiety (HAM-A) als weiterer Morbiditätsendpunkt misst die Schwere von Angstzuständen mit 14 Items (Skala von 0 – 56 Punkten), die sowohl psychische als auch somatische Angst bewerten. Das klinisch relevante Ansprechen wird über eine 15 %-ige Verbesserung auf der Skalenspannweite der HAM-A angezeigt. Im Zuranolon+PiT-Arm sprachen statistisch signifikant mehr Patientinnen

an als im Vergleichsarm. Zum Ende der Behandlung lag die Wahrscheinlichkeit, eine relevante Verbesserung bezüglich der Angstsymptomatik zu erreichen bei der Behandlung mit Zuranolon in Kombination mit PiT um 32,3 % höher als im Vergleichsarm (RR: 1,323 [1,129; 1,552], p = 0,0006). Dies entspricht einem **beträchtlichen** Zusatznutzen. Auch hier wurde ein Ansprechen statistisch signifikant – um 35,6 % (3 vs. 4 Tage zum 25. Perzentil) – schneller im Zuranolon+PiT-Arm als im Vergleichsarm erreicht (HR: 1,356 [1,0738; 1,7128], p = 0,0105). Diese schnell eintretende, deutlich überlegene Wirksamkeit von Zuranolon gegenüber der Vergleichstherapie wird bis zum Ende der Beobachtungsphase beibehalten (RR: 1,178 [1,014; 1,369], p = 0,0333) und entspricht einem **geringen** Zusatznutzen.

Relevante Ergebnisse von Kliniker*innen bewertet

Die Clinical Global Impression-Skalen (CGI) sind standardisierte Instrumente, die häufig in klinischen Studien eingesetzt werden, um den Schweregrad einer Erkrankung sowie die Verbesserung des Zustands von Patient*innen zu bewerten. Zum Gesamtbild in Bezug auf Schweregrad der PPD bzw. Veränderung des Zustands der Patientin kann die Einschätzung von Kliniker*innen einen wertvollen Beitrag leisten. Die zuvor dargestellten Endpunkte werden hierdurch nicht nur ergänzt, sondern in ihrer Robustheit unterstützt.

Die CGI-S-Skala (CGI-Severity) dient in diesem Studiensetting dazu, den Gesundheitszustand der PPD-Patientinnen zum Zeitpunkt der Bewertung einzuschätzen. Dabei wird eine 7-Punkte-Likert-Skala verwendet, die von „normal, nicht krank“ (1) bis „extrem krank“ (7) reicht. Die Bewertung erfolgt relativ zur bisherigen klinischen Erfahrung der Bewertungsperson mit Patientinnen, die dieselbe Diagnose haben, und bildet die Grundlage für spätere Beurteilungen der Verbesserung. Der Gesundheitszustand gemessen mittels CGI-S verbessert sich im Zuranolon+PiT-Arm statistisch signifikant bei mehr Patientinnen als im Vergleichsarm. Zum Ende der Behandlung haben Patientinnen unter Zuranolon eine 34,8 % höhere Wahrscheinlichkeit eine klinisch relevante Verbesserung zu erreichen als unter der Vergleichstherapie (RR: 1,348 [1,113; 1,633], p = 0,0023). Diese Verbesserung wird auch statistisch signifikant schneller erreicht (HR: 1,688 [1,3209; 2,1562], p < 0,0001) Dies entspricht einem **beträchtlichen** Effekt.

Die CGI-I-Skala (CGI-Improvement) misst hingegen die Verbesserung des Zustands eines Patienten nach der Behandlung im Vergleich zu Baseline. Auch hier wird eine 7-Punkte-Likert-Skala verwendet, die von „sehr stark verbessert“ (1) bis „sehr stark verschlechtert“ (7) reicht. Im CGI-I zeigte sich ebenfalls für statistisch signifikant mehr Patientinnen eine Verbesserung im Zuranolon+PiT-Arm als im Vergleichsarm. Zum Ende der Behandlung haben Patientinnen unter Zuranolon eine ca. 40 % höhere Wahrscheinlichkeit die Verbesserung zu erreichen als unter der Vergleichstherapie (RR: 1,393 [1,151; 1,686], p = 0,0007). Diese Verbesserung wird auch statistisch signifikant schneller erreicht (HR: 1,660 [1,3042; 2,1135], < 0,0001). Hieraus kann in Übereinstimmung mit den Ergebnissen des CGI-S ein **beträchtlicher** Effekt abgeleitet werden.

Der statistisch signifikante Unterschied zwischen den Behandlungsgruppen ist auch nachhaltig und wurde zum Ende der Beobachtung in beiden CGI-Endpunkten bestätigt.

Lebensqualität

Die Lebensqualität wird durch den Short Form (SF)-36 und den Barkin Index of Maternal Functioning (BIMF) erfasst. Beide Instrumente wurden nur in der Studie ROBIN erhoben, daher können hier nur die Ergebnisse auf Ebene einer RCT dargestellt werden.

Der SF-36 ist ein standardisierter Fragebogen mit 36 Items, der die allgemeine gesundheitsbezogene Lebensqualität mit 8 Subskalen misst, die in körperliche und psychische Summenscores unterteilt werden können. Der BIMF hingegen bewertet die mütterliche Funktionsfähigkeit in den ersten 12 Monaten nach der Geburt. Mit 20 Items und einer Skala von 0 – 120 deckt er Bereiche wie Selbstfürsorge, Mutter-Kind-Interaktion und soziale Unterstützung ab. Beide Instrumente sind validiert und gelten als patientenrelevant, da sie zentrale Aspekte der Lebensqualität und der Bewältigung des Alltags erfassen. Es wurden Responderanalysen bezüglich des Ansprechens durchgeführt, wobei für alle Skalen eine relevante Verbesserung definiert war als Zunahme um ≥ 15 % des gesamten Skalenbereich.

Statistisch signifikant positive Effekte auf die Lebensqualität zeigten sich bei drei der acht Subskalen des SF-36. Es konnte am Ende der Beobachtung für „Körperliche Schmerzen“ ein statistisch signifikanter Vorteil zugunsten von Zuranolon+PiT gegenüber der Vergleichstherapie beobachtet werden (RR: 1,461 [1,051; 2,032], p = 0,0243). Ähnlich positive Effekte waren auch auf den Subskalen „Soziale Funktionsfähigkeit“ (RR: 1,218 [1,011; 1,467], p = 0,0378) und „Psychisches Wohlbefinden“ (RR: 1,240 [1,028; 1,497], p = 0,0249) zu sehen. Für den Gesamtwert des BIMF wurden bereits zum Ende der Zuranolon-Behandlung statistisch signifikant mehr Patientinnen mit Ansprechen im Zuranolon+Pit-Arm beobachtet als im Vergleichsarm (RR: 1,616 [1,040; 2,513], p = 0,0330). Den statistisch signifikanten Ergebnissen zugunsten von Zuranolon stehen keine gegenteiligen Effekte auf irgendeiner Skala des SF-36 oder BIMF entgegen. Zusammenfassend kann für die Lebensqualität ein Zusatznutzen in **geringem** Ausmaß entnommen werden.

Sicherheit

Erfassung von Suizidalität mittels C-SSRS

In den Studien ROBIN und SKYLARK wurden suizidale Gedanken bei wenigen Patientinnen beobachtet, Aufgrund der insgesamt sehr geringen Ereigniszahlen konnte kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen festgestellt werden.

Unerwünschte Ereignisse

Im Zuranolon+PiT-Arm trat bei 63,6 % der Patientinnen mindestens ein UE auf, während im Vergleichsarm bei 52,6 % der Patientinnen mindestens ein UE beobachtet wurde. Innerhalb der SOC „Erkrankungen des Nervensystems“ wurden statistisch signifikant häufiger UE im Zuranolon+PiT-Arm festgestellt als im Vergleichsarm (RR: 1,66 [1,230; 2,248], p = 0,0010). Nebenwirkungen aus dieser SOC sind für Zuranolon bekannt und werden auch in der Fachinformation genannt [1]. Die PTs „Somnolenz“ und „Sedierung“ traten im Zuranolon+PiT-Arm signifikant häufiger auf (RR: 2,84 [1,569; 5,142], p = 0,0006 bzw. RR: 14,57 [1,946; 109,126], p = 0,0091). Die Ereignisse waren mild bis moderat und traten, wie in Tabelle 4-3 dargestellt, vorwiegend in den ersten 1-2 Tagen auf [6; 7].

Tabelle 4-3: Inzidenz von UE nach Zeit des Auftretens

UE **Zeitintervall**	**Zuranolon+PiT** **(N = 176)**	**PiT** **(N = 171)**
Somnolenz [1], n (%)		
1 – 2 Tage	32 (18,2)	12 (7,0)
3 – 7 Tage	10 (5,7)	1 (0,6)
8 – 15 Tage	4 (2,3)	1 (0,6)
> 15 Tage	0	0
Sedierung [2]		
1 – 2 Tage	10 (5,7)	0
3 – 7 Tage	5 (2,8)	1 (0,6)
8 – 15 Tage	2 (1,1)	0
> 15 Tage	0	0

[1]: Somnolenz umfasst die PTs Somnolenz und Hypersomnie.

[2]: Sedierung umfasst den PT Sedierung

UE werden basierend auf den Zeitintervallen ihres Auftretens klassifiziert, entsprechend der individuellen Expositionsdauer jeder Teilnehmerin zum Zeitpunkt des Auftretens des UE. Die Zeit bis zum Auftreten wird berechnet als Datum des UE-Beginns – Datum der ersten Dosis + 1. Für Teilnehmerinnen, die denselben Ereignistyp mehrfach und in verschiedenen Zeitintervallen erlebten, wird das Ereignis in jedem Intervall, in dem es auftrat, einmal gezählt. UE mit unvollständigen Beginn-Daten werden von diesen Analysen ausgeschlossen.

Quelle: [7]

Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.

Sowohl Somnolenz als auch Sedierung gehören zu den bekannten, in der Fachinformation genannten, sehr häufigen Nebenwirkungen, die jedoch zumeist transient zu Beginn der Behandlung auftreten und im Zusammenhang mit dem Wirkmechanismus von Zuranolon stehen [1].

Hinsichtlich des Auftretens schwerer UE zeigte sich kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen. Ebenso gab es keinen signifikanten Unterschied im Auftreten von SUE zwischen den Gruppen. Diese Ereignisse waren insgesamt selten.

Die Therapieabbrüche aufgrund von UE waren nicht signifikant unterschiedlich und insgesamt selten mit 2,8 % im Zuranolon+PiT-Arm und 1,2 % im Vergleichsarm.

Insgesamt zeigte Zuranolon in der Indikation PPD ein gut verträgliches Sicherheitsprofil. Die Mehrheit der UE war von leichter bis moderater Schwere. Häufig berichtete UE betrafen das zentrale Nervensystem, darunter Somnolenz und Sedierung, die transient zum Beginn der Behandlung beobachtet wurden.

Schlussfolgerungen zum Zusatznutzen und zum therapeutisch bedeutsamen Zusatznutzen

Die PPD ist eine ernsthafte Erkrankung, deren depressive Symptomatik nicht nur die betroffene Mutter und die Entwicklung des Säuglings erheblich beeinträchtigt, sondern auch die gesamte familiäre Dynamik, inklusive Partner*innen und Geschwisterkinder, mit kurz- und langfristigen Konsequenzen belastet.

Bei mittelschweren bis schweren Formen der PPD sind die Symptome wie anhaltende Traurigkeit, häufiges Weinen, deutliche Ängstlichkeit, Schlafstörungen, Konzentrationsschwierigkeiten und möglicherweise Appetitveränderungen oft so ausgeprägt, dass sie die alltäglichen Funktionen der Mutter erheblich einschränken. Das Leben der Betroffenen mit mittelschwerer bis schwerer PPD kann akut durch Selbstmordgedanken oder gar Selbstmordversuche bedroht sein. Symptome der Mutter schränken die Fähigkeit, sich um sich selbst und den Säugling zu kümmern deutlich ein und erfordern eine engmaschige Beobachtung und eine angemessene Behandlung. Die Auswirkungen auf die Mutter-Kind-Interaktion können zudem langfristige Folgen für die Entwicklung des Säuglings haben, einschließlich emotionaler und kognitiver Beeinträchtigungen. Endpunkte, die die Symptome der Mutter bewerten, sind somit nicht nur für die Mutter selbst von Bedeutung, sondern auch für die langfristige Gesundheit des Säuglings. Darüber hinaus können die depressiven Symptome der Mutter negative Konsequenzen im gesamten familiären Umfeld nach sich ziehen, einschließlich eines erhöhten Depressionsrisikos für den Partner oder die Partnerin sowie emotionaler und sozialer Entwicklungsstörungen bei Geschwisterkindern [8-10]. Daher ist es unerlässlich, alle Endpunkte in diesem Kontext als schwerwiegend zu betrachten, um die umfassenden Auswirkungen der Erkrankung auf die gesamte Familie zu erkennen und zu adressieren.

Alle Ergebnisse bestätigen: Zuranolon zeigt eine schnelle und anhaltende Wirksamkeit bei der Behandlung von mittelschwerer bis schwerer PPD. Diese schnelle Wirkung ist besonders wertvoll, da eine PPD nicht nur die Lebensqualität der Betroffenen erheblich beeinträchtigt, sondern auch schwerwiegende Folgen für die Familie haben kann, einschließlich eines erhöhten Risikos für Suizid [11; 12], einer gestörten Mutter-Kind-Bindung und kindlicher Mangelernährung im ersten Lebensjahr [3]. Im Vergleich zu herkömmlichen Antidepressiva, die oft Wochen bis Monate benötigen, um eine Wirkung zu zeigen, bietet Zuranolon durch seinen raschen Wirkungseintritt eine dringend benötigte Entlastung in einer kritischen Phase [13-15]. Dies beschleunigt nicht nur die Genesung, sondern mindert auch die langfristigen sozialen und gesundheitlichen Folgen einer PPD. Bereits nach der zweiten täglichen Dosis geht es den Patientinnen deutlich besser, was die HAMD-17-, EPDS- und HAM-A-Ergebnisse an Tag 3 mit einem statistisch signifikanten Vorteil gegenüber dem Vergleichsarm zeigen (siehe Tabelle 4-1).

Die Ergebnisse zum Ende der Behandlung sowie auch zum Ende der Beobachtung aus der Meta-Analyse belegen signifikante Verbesserungen.

Schnelle Wirkung mit beträchtlichem Zusatznutzen für Patientinnen und Familien

Unter Zuranolon+PiT zeigen die Patientinnen eine besonders schnelle und statistisch signifikante Verbesserung der depressiven Symptomatik, insbesondere durch hohe Ansprech- und Remissionsraten bei HAMD-17 und MADRS. Diese schnelle Wirkung ist entscheidend für die Patientin selbst sowie auch für die Mutter-Kind-Beziehung, die unter den Folgen der Erkrankung massiv beeinträchtigt sein können. Insgesamt stellt die frühzeitige Verbesserung der Symptome einen zentralen Vorteil in **beträchtlichem** Ausmaß dar.

Morbidität: beträchtlicher Zusatznutzen

Signifikant mehr Patientinnen unter Zuranolon gegenüber der Vergleichstherapie erreichten am Ende der Behandlung eine 50%-ige Verbesserung der Depressionssymptome gemessen mittels HAMD-17 und MADRS. Dies wird als „beträchtlicher Zusatznutzen“ bewertet. Die TTE-Analysen zeigten einen schnellen Wirkeintritt von beträchtlichem Ausmaß für MADRS sowie erheblichem Ausmaß für HAMD-17, was besonders in kritischen Phasen der Erkrankung, wie bei Suizidalität, von Bedeutung ist. Die signifikante Überlegenheit gegenüber der Vergleichstherapie und damit der Zusatznutzen bleibt bis zum Ende der Beobachtung erhalten. Die Responderanalysen sowie auch TTE-Analysen der HAMD-17 sowie auch der MADRS bezüglich einer Remission unter Zuranolon zeigen ebenfalls signifikante Vorteile gegenüber der Vergleichstherapie in beträchtlichem Ausmaß. Langfristige Daten und Real-World-Analysen unterstützen die Stabilität der Remission über den klinischen Studienzeitraum hinaus. Die Ergebnisse zeigen, dass Patientinnen, die eine Remission erreichen, in der Regel über längere Zeiträume stabil bleiben, was durch longitudinale Studien und reale Versicherungsdaten gestützt wird [6]. Dies unterstreicht die Bedeutung einer Kurzzeitbehandlung mit Zuranolon als schnelle und effektive Therapieoption für Patientinnen mit PPD, die eine nachhaltige Symptomlinderung bietet.

Auch für die Instrumente EPDS, PHQ-9 und HAM-A zeigte sich zum Ende der Behandlung ein deutlicher Zusatznutzen. Der Effekt war über die TTE-Analysen für PHQ-9 und HAM-A signifikant schneller eingetreten und blieb bei allen drei Endpunkten bis zum Ende der Behandlung als signifikante Überlegenheit sichtbar.

Folglich ergibt sich für Zuranolon in der Endpunktkategorie Morbidität ein **beträchtlicher Zusatznutzen.**

Einschätzung von Kliniker*innen: Schnelle Abnahme der Symptome und der Krankheitsschwere

Die CGI-Daten liefern eine klinische Einschätzung der Reduktion des Schweregrads der PPD. Die klinische Einschätzung, wie sie durch Skalen wie CGI-I und CGI-S erfolgt, ermöglicht es, eine objektive Verbesserung der Patientin im Kontext der klinischen Praxis zu bewerten. Diese Einschätzungen sind oft entscheidend für die Therapieanpassung und das Management der Erkrankung. Die Ergebnisse zeigen statistisch signifikante Überlegenheit von Zuranolon und ergänzen die patientenberichteten Endpunkte und die Ergebnisse der Meta-Analyse, insbesondere in Bezug auf die schnell einsetzende Symptomlinderung und die Reduktion des Schweregrads der Erkrankung.

Lebensqualität: geringer Zusatznutzen

Zusätzlich zur Morbidität zeigte sich der Zusatznutzen auch für die Lebensqualität. Patientinnen berichten statistisch signifikant häufiger Verbesserungen in den SF-36-Subskalen „körperlicher Schmerz", „soziale Funktion" und „mentale Gesundheit". Dies wird als **geringer Zusatznutzen** eingestuft.

Sicherheit: transiente milde bis moderate Nebenwirkungen mit rascher Abnahme

Das Sicherheitsprofil von Zuranolon ist insgesamt gut. Schwere oder schwerwiegende UE sind insgesamt sehr selten. Die im Vergleich zum Kontrollarm häufiger auftretenden UE, wie Somnolenz und Sedierung waren mild bis moderat und traten vorwiegend lediglich in den ersten 1 – 2 Tagen auf und nahmen dann ab [6]. Diese Ereignisse sind gut behandelbar, bzw. werden gut toleriert, was durch die niedrige Rate von Therapieabbrüchen aufgrund von UE bestätigt wird.

Zusammenfassung

Insgesamt belegen die vorgelegten Daten in Form der Meta-Analyse als höchster Evidenzstufe Ia und der Einzelstudien, dass durch die Anwendung von Zuranolon nicht nur die Symptome der PPD signifikant gelindert werden, sondern auch ein wichtiger Beitrag zur Verbesserung der Lebensqualität der betroffenen Mütter und zur Förderung einer gesunden Bindung zwischen Mutter und Säugling geleistet wird. Die vorgelegten Daten unterstreichen die schnelle und über die gesamte Studie anhaltende Wirksamkeit einer Zuranolon+PiT-Behandlung bei der mittelschweren bis schweren PPD und resultieren in einem **Beleg für einen beträchtlichen Zusatznutzen.**

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.

Sofern Angaben zur Methodik im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist in dem jeweiligen Abschnitt des vorliegenden Dossiers darauf zu verweisen.

Sofern Angaben zur Methodik bisher teilweise oder vollständig nicht im EU-Dossier eingereicht wurden, so sind diese Angaben in den betroffenen Abschnitten des Moduls 4 jeweils zu ergänzen.

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

Unter Endpunkte sind dabei alle für die frühe Nutzenbewertung relevanten Endpunkte anzugeben (das heißt nicht nur solche, die gegebenenfalls in den relevanten Studien untersucht wurden).

Die Benennung der Vergleichstherapie in Modul 4 muss zur Auswahl der zweckmäßigen Vergleichstherapie im zugehörigen Modul 3 konsistent sein.

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.

Im Falle einer vorangegangenen gemeinsamen klinischen Bewertung nach der Verordnung (EU) 2021/2282 formulieren Sie unabhängig von dem im Verfahren der gemeinsamen klinischen Bewertung definierten Bewertungsumfang die vollständige Fragestellung für die Nutzenbewertung nach § 35a SGB V in Bezug auf den deutschen Versorgungskontext.

Ziel des vorliegenden Dossiers ist die Quantifizierung des Ausmaßes des Zusatznutzens von Zuranolon bei erwachsenen Patientinnen mit einer PPD nach Entbindung anhand einer Meta-Analyse der Studien ROBIN und SKYLARK.

Patientenpopulation

Gemäß der Zulassung umfasst die Patientenpopulation Erwachsene mit einer PPD nach der Entbindung.

Intervention

Die zu bewertenden Intervention ist Zuranolon nach Maßgabe der Fachinformation [1].

Vergleichstherapie

Der Gemeinsame Bundesausschuss (G-BA) bestimmte im Beratungsgespräch vom 13.06.2025 unter der Vorgangsnummer 2025-B-099 zwei Patientenpopulationen mit dazu aufgeführter zweckmäßigen Vergleichstherapie (zVT) [16]:

a) Erwachsene Patientinnen nach der Geburt mit einer PPD mit **leichter** depressiver Symptomatik:

- Best Supportive Care

b) Erwachsene Patientinnen nach der Geburt mit einer PPD mit **mittelschwerer bis schwerer** depressiver Symptomatik:

- Individualisierte Therapie unter Auswahl von:
    - Best Supportive Care
    - einem selektiven Serotonin-Wiederaufnahmehemmer (SSRI)

Patientenrelevante Endpunkte

Die Ergebnisse werden für die vier patientenrelevanten Dimensionen Mortalität, Morbidität, Lebensqualität und Sicherheit dargestellt.

Mortalität

Die Mortalität wurde im Rahmen der Sicherheit anhand der UE, die zum Tod führen, erhoben.

Morbidität

- HAMD-17
- MADRS
- EPDS
- PHQ-9
- HAM-A

Lebensqualität

- SF-36

- BIMF

Sicherheit

- Columbia-Suicide Severity Rating Scale (C-SSRS)
- UE (Gesamtraten und nach SOC und PT)
- Schwere UE (Gesamtraten und nach SOC und PT)
- SUE (Gesamtraten und nach SOC und PT)
- Therapieabbrüche aufgrund von UE

Studientypen

Für die Nutzenbewertung von Zuranolon liegen keine Daten für die Patientenpopulation a) vor. Diese Patientinnen werden in der Regel wegen der leichten Symptomatik nicht mit Arzneimitteln behandelt. Für die Patientenpopulation b) Patientinnen nach der Geburt mit einer PPD mit mittelschwerer bis schwerer depressiver Symptomatik liegen 2 randomisierte, kontrollierte klinische Studien der Phase 3 vor. Die Studien wurden zu einer Meta-Analyse zusammengefasst um eine Bewertung von Zuranolon gegenüber der vom G-BA festgelegten zVT „individualisierte Therapie unter Auswahl von BSC und einem SSRI“ zu ermöglichen. Damit liegen Ergebnisse mit höchster Ergebnissicherheit vor, die gemäß § 5 Abs. 6 AM-NutzenV der Evidenzstufe Ia entsprechen.

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. Dabei ist zu beachten, dass eine Studie nicht allein deshalb ausgeschlossen werden soll, weil keine in einer Fachzeitschrift veröffentlichte Vollpublikation vorliegt. Eine Bewertung der Studie kann beispielsweise auch auf Basis eines ausführlichen Ergebnisberichts aus einem Studienregister/ einer Studienergebnisdatenbank erfolgen, während ein Kongressabstrakt allein in der Regel nicht für eine Studienbewertung ausreicht.

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. Erstellen Sie dabei für unterschiedliche Themen der Recherche (zum Beispiel unterschiedliche Fragestellungen) jeweils eine separate Übersicht.

Sofern Informationen zu Ein- und Ausschlusskriterien für Studien zum medizinischen Nutzen und Zusatznutzen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung

nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Um die für die Beantwortung der Fragestellung relevanten Studien zu selektieren, wurden Ein- und Ausschlusskriterien für die Population, die Intervention, die Vergleichstherapie, die Endpunkte, den Studientyp, den Publikationstyp sowie die Studiendauer definiert. Die Ein- und Ausschlusskriterien für die Berücksichtigung von Studien zur Nutzenbewertung von Zuranolon werden in Tabelle 4-4 zusammengefasst.

Tabelle 4-4: Übersicht der Ein-/Ausschlusskriterien der berücksichtigten Studien

Kriterium		Einschlusskriterium	Ausschlusskriterium	Begründung
1	**Patientenpopulation**	Erwachsenen nach der Geburt mit einer postpartalen Depression	Patientenpopulation abweichend	Zugelassene Population im betrachteten Anwendungsgebiet (Zielpopulation)
2	**Intervention**	Zuranolon	Intervention abweichend	Der Einsatz von Zuranolon muss zulassungskonform oder in adäquater Dosierung erfolgen
3	**Vergleichstherapie**	• SSRI • BSC	Vergleichstherapie abweichend	Entspricht der vom G-BA benannten zVT [16]
4	**Endpunkte**	Mindestens ein patientenrelevanter Endpunkt zu • Mortalität • Morbidität • Gesundheitsbezogener Lebensqualität • Sicherheit	Keine patientenrelevanten Endpunkt	Anforderungen von AMNutzenV und IQWiG Methodenpapier
5	**Studiendesign**	RCT	Studien, die nicht randomisiert und nicht kontrolliert sind Extensionsstudien ohne Aufrechterhaltung der RCT-Bedingungen	Studientyp mit der höchsten Ergebnis-sicherheit, siehe § 5 Abs. 3 VerfO
6	**Studiendauer**	Mindestens 14 Tage Behandlungszeit	Behandlungsdauer unter 14 Tage	Eine einmalige 14-tägige Gabe ist laut Fachinformation vorgesehen [1]
7	**Publikationstyp**	Vollpublikation oder Bericht verfügbar, der den Kriterien des CONSORT- bzw. TREND-Statements genügt	Berichterstattung liefert keine ausreichenden Informationen zur Beurteilung von Methodik/Ergebnissen	Anforderung gemäß VerfO des G-BA

Kriterium		Einschlusskriterium	Ausschlusskriterium	Begründung
		und so eine Einschätzung der Studienergebnisse ermöglicht	(z. B. narrativer Review, Letter, Editorial, Errata, Note, Konferenz-Abstract oder Paper, Short Survey, unsystematische Übersichtsarbeit)	
8	**Studienstatus**	Abgeschlossene oder laufende Studie mit vorliegenden Ergebnissen	Rekrutierende Studie oder abgeschlossene oder laufende Studie ohne verfügbare Ergebnisse	
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.				

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).

Sofern Angaben zur Methodik der Informationsbeschaffung im EU-Dossier enthalten sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Das Datum der Recherche soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen. Liegt die Recherche im EU-Dossier mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt, ist die Recherche zu aktualisieren. Die aktualisierte Recherche ist im Dossier darzustellen.

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 Recherche

Die Durchführung einer bibliografischen Recherche ist erforderlich, um sicherzustellen, dass ein vollständiger Studienpool in die Bewertung einfließt.

Eine bibliografische Recherche muss für RCT mit dem zu bewertenden Arzneimittel (Abschnitt 4.3.1) immer durchgeführt werden. Für indirekte Vergleiche auf Basis von RCT (Abschnitt 4.3.2.1), nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2) sowie weitere Untersuchungen (Abschnitt 4.3.2.3) muss eine bibliografische Recherche immer dann durchgeführt werden, wenn auf Basis solcher Studien der medizinische Zusatznutzen bewertet wird.

Das Datum der Recherche soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Die bibliografische Recherche soll mindestens in den Datenbanken MEDLINE (inklusive „in-process & other non-indexed citations) sowie in der Cochrane-Datenbank „Cochrane Central Register of Controlled Trials (CENTRAL)“ durchgeführt werden. Optional kann zusätzlich eine Suche in weiteren Datenbanken (zum Beispiel EMBASE, PsycINFO et cetera) 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 gegebenenfalls Studientypen, aufgebaut werden. Wird eine Einschränkung der Strategien auf bestimmte Studientypen vorgenommen (zum Beispiel randomisierte kontrollierte Studien), sollen aktuelle validierte Filter hierfür verwendet werden. Alle Suchstrategien sind in Anhang 4-A zu dokumentieren.

Beschreiben Sie nachfolgend für alle durchgeführten Recherchen, in welchen Datenbanken eine bibliografische Recherche durchgeführt wurde. Begründen Sie Abweichungen von den oben beschriebenen Vorgaben. Geben Sie auch an, wenn bei der Recherche generelle Einschränkungen vorgenommen wurden (zum Beispiel Sprach- oder Jahreseinschränkungen), und begründen Sie diese.

Sofern Angaben zu diesem Abschnitt im EU-Dossier enthalten sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Eine bibliografische Literaturrecherche wurde am 16.04.2026 in der Datenbank MEDLINE und der Cochrane Library durchgeführt. In jeder Datenbank wurde eine angepasste Suchstrategie verwendet. Es wurden Blöcke nach Indikation und Studientyp in der Suchstrategie mit dem Arzneimittel verknüpft und für MEDLINE auch ein RCT-Filter verwendet. Die Treffer wurden anschließend nach den zur Fragestellung passenden Selektionskriterien (Tabelle 4-4) bewertet. Bei der Suche in der Cochrane Library wurden nur Treffer aus der Kategorie „Trials“ berücksichtigt. Die jeweiligen Suchstrategien sind in Anhang 4-A dokumentiert.

4.2.3.3 Suche in Studienregistern/Studienergebnisdatenbanken

Eine Suche in öffentlich zugänglichen Studienregistern/Studienergebnisdatenbanken ist grundsätzlich durchzuführen, um sicherzustellen, dass laufende Studien sowie abgeschlossene Studien auch von Dritten vollständig identifiziert werden und in Studienregistern/Studienergebnisdatenbanken vorliegende Informationen zu Studienmethodik und –ergebnissen in die Bewertung einfließen.

Eine Suche in Studienregistern/ Studienergebnisdatenbanken muss für RCT mit dem zu bewertenden Arzneimittel (Abschnitt 4.3.1) immer durchgeführt werden. Für indirekte Vergleiche auf Basis von RCT (Abschnitt 4.3.2.1), nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2) sowie weitere Untersuchungen (Abschnitt 4.3.2.3) muss eine Suche in Studienregistern sowie Studienergebnisdatenbanken immer dann durchgeführt werden, wenn auf Basis solcher Studien der medizinische Zusatznutzen bewertet wird.

Das Datum der Recherche soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Die Suche soll mindestens in den Studienregistern/Studienergebnisdatenbanken National Library of Medicine (US) Clinicaltrials.gov (www.clinicaltrials.gov), EU Clinical Trials Register (EU-CTR, www.clinicaltrialsregister.eu), Clinical Trials Information System (CTIS) (https://euclinicaltrials.eu/), Clinical Data Suchportal der European Medicines Agency (https://clinicaldata.ema.europa.eu) durchgeführt werden.

Die Suche soll in jedem Studienregister/Studienergebnisdatenbank einzeln und mit einer für das jeweilige Studienregister/Studienergebnisdatenbank adaptierten Suchstrategie durchgeführt werden. Die Suche soll abgeschlossene, abgebrochene und laufende Studien erfassen. Alle Suchstrategien sind in Anhang 4-B zu dokumentieren.

Für Clinical Data (Suchportal der European Medicines Agency) genügt hingegen die Suche nach Einträgen mit Ergebnisberichten zu Studien, die bereits anderweitig (zum Beispiel über die bibliografische Recherche und Studienregistersuche) identifiziert wurden. Eine Dokumentation der zugehörigen Suchstrategie ist nicht erforderlich.

Beschreiben Sie nachfolgend für alle durchgeführten Recherchen, in welchen Studienregistern/Studienergebnisdatenbanken die Suche durchgeführt wurde. Begründen Sie dabei Abweichungen von den oben beschriebenen Vorgaben. Geben Sie auch an, wenn bei der Recherche generelle Einschränkungen vorgenommen wurden (zum Beispiel Jahreseinschränkungen), und begründen Sie diese.

Sofern Angaben zu diesem Abschnitt im EU-Dossier enthalten sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Die Suche für RCT zu oben genannter Fragestellung erfolgte entsprechend den Vorgaben in den Studienregistern/Studienergebnisdatenbanken clinicaltrials.gov, EU Clinical Trials

Register, Clinical Trials Information System (CTIS) sowie dem Clinical Data Suchportal der European Medicines Agency (EMA). Eine komplette Suche wurde am 16.04.2026 in den Studienregistern/Studienergebnisdatenbanken anhand separater Suchstrategien durchgeführt. Im Rahmen dieser Suchen wurden abgeschlossene, abgebrochene und laufende Studien erfasst. Es wurden keine generellen Einschränkungen der Recherche vorgenommen. Die jeweiligen Suchstrategien für die Studienregister sind in Anhang 4-B dokumentiert. Die Suchstrategie für das Clinical Data Portal der EMA wurde nicht dokumentiert.

4.2.3.4 Suche auf der Internetseite des Gemeinsamen Bundesausschuss

Die Internetseite des Gemeinsamen Bundesausschusses ist grundsätzlich zu durchsuchen, um sicherzustellen, dass alle vorliegenden Daten zu Studienmethodik und –ergebnissen von relevanten Studien in die Bewertung einfließen.

Auf der Internetseite des Gemeinsamen Bundesausschusses werden Dokumente zur frühen Nutzenbewertung nach §35a SGB V veröffentlicht. Diese enthalten teilweise anderweitig nicht veröffentlichte Daten zu Studienmethodik und –ergebnissen[1]. Solche Daten sind dabei insbesondere in den Modulen 4 der Dossiers pharmazeutischer Unternehmer, in IQWiG-Nutzenbewertungen sowie dem Beschluss des Gemeinsamen Bundesausschusses einschließlich der Tragenden Gründe und der Zusammenfassenden Dokumentation zu erwarten.

Die Suche auf der Internetseite des Gemeinsamen Bundesausschusses muss für RCT mit dem zu bewertenden Arzneimittel (Abschnitt 4.3.1) immer durchgeführt werden. Für indirekte Vergleiche auf Basis von RCT (Abschnitt 4.3.2.1), nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2) sowie weitere Untersuchungen (Abschnitt 4.3.2.3) muss eine Suche auf der Internetseite des Gemeinsamen Bundesausschusses immer dann durchgeführt werden, wenn auf Basis solcher Studien der medizinische Zusatznutzen bewertet wird. Die Suche ist dann sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie durchzuführen. Es genügt die Suche nach Einträgen zu Studien, die bereits anderweitig (zum Beispiel über die bibliografische Recherche und Studienregistersuche) identifiziert wurden. Eine Dokumentation der zugehörigen Suchstrategie ist nicht erforderlich.

Das Datum der Recherche soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Beschreiben Sie nachfolgend das Vorgehen für die Suche. Benennen Sie die Wirkstoffe und die auf der Internetseite des Gemeinsamen Bundesausschusses *genannten zugehörigen Vorgangsnummern, zu denen Sie eine Suche durchgeführt haben.*

[1] Köhler M, Haag S, Biester K, Brockhaus AC, McGauran N, Grouven U, Kölsch H, Seay U, Hörn H, Moritz G, Staeck K, Wieseler B. Information on new drugs at market entry: retrospective analysis of health technology assessment reports, journal publications, and registry reports. BMJ 2015;350:h796

Begründen Sie Abweichungen von den oben beschriebenen Vorgaben.

Sofern Angaben zu diesem Abschnitt im EU-Dossier enthalten sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Auf der Internetseite des G-BA wurde am 16.04.2026 eine Suche nach Einträgen zu Studien, die bereits anderweitig (z. B. über die bibliografische Literaturrecherche und Studienregistersuche) identifiziert wurden, durchgeführt.

4.2.3.5 Selektion relevanter Studien

Beschreiben Sie das Vorgehen bei der Selektion relevanter Studien aus dem Ergebnis der in den Abschnitten 4.2.3.2, 4.2.3.3 und 4.2.3.4 beschriebenen Rechercheschritte. Begründen Sie das Vorgehen, falls die Selektion nicht von zwei Personen unabhängig voneinander durchgeführt wurde.

Sofern Angaben zu diesem Abschnitt im EU-Dossier enthalten sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Die durch die Suche in bibliografischen Datenbanken identifizierten Quellen wurden in einem ersten Schritt durch die verwendete Literaturverwaltungs-Software (Endnote) um etwaige Dubletten bereinigt. In einem weiteren Schritt wurden verbliebene Dubletten manuell aussortiert. Die Selektion der Treffer erfolgte anhand ihres Titels und – sofern vorhanden – ihres Abstracts anhand der in Abschnitt 4.2.2 aufgeführten Kriterien. Die Selektion erfolgte durch zwei Personen unabhängig voneinander. Etwaige Diskrepanzen in der Bewertung wurden durch Diskussion zwischen den zwei Personen aufgelöst.

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-F):

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 beziehungsweise Berücksichtigung prognostisch relevanter Faktoren *(bei nicht randomisierten vergleichenden Studien)*
– Verblindung des Patienten sowie der behandelnden Personen
– 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 der Verzerrungsaspekte und des Verzerrungspotenzials eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.

Sofern im EU-Dossier die Methodik zur Bewertung der Verzerrungsaspekte hinterlegt ist und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Verzerrungspotenzial auf Studienebene

Die Bewertung des Verzerrungspotenzials auf Studienebene berücksichtigte zentrale Aspekte des Studiendesigns, wie beispielsweise die Art der Studie (RCT oder unkontrollierte, nicht-

randomisierte offene Studie). Im Falle einer RCT wurden die Erstellung der Randomisierungssequenz, die Verdeckung der Gruppenzuteilung sowie die Verblindung von Patientinnen und behandelnden Personen geprüft. Zusätzlich wurde analysiert, ob eine verzerrte Berichterstattung der Ergebnisse vorlag. Hierfür erfolgte ein Abgleich der Angaben aus dem Studienprotokoll, dem statistischen Analyseplan und dem Studienbericht. Weitere mögliche Verzerrungen wurden durch die Bewertung der Transparenz und Plausibilität des Patientenflusses sowie der Vergleichbarkeit der Gruppen (z. B. hinsichtlich Alter, Geschlecht, Konstitution, Begleiterkrankungen und -behandlungen) identifiziert. Auch die Klarheit der Beschreibung und Durchführung der Methodik sowie etwaige nachträgliche Änderungen wurden berücksichtigt. Basierend auf dieser Analyse wurde das Verzerrungspotenzial auf Studienebene entweder als „niedrig“ oder „hoch“ eingestuft. Eine Einstufung als „niedrig“ erfolgte, wenn bei einer RCT keine relevanten Verzerrungen der Ergebnisse zu erwarten waren und keine Hinweise auf verzerrende Faktoren vorlagen, deren Behebung die Grundaussage der Ergebnisse wesentlich verändert hätte.

Verzerrungspotenzial auf Endpunktebene

Es wurden verschiedene Faktoren geprüft, die potenziell zu Verzerrungen auf Endpunktebene führen könnten. Dazu zählten unter anderem die Verblindung der Endpunkterhebenden, die Durchführung der Datenanalyse und die Darstellung der Ergebnisse. Ein „niedriges“ Verzerrungspotenzial wurde dann angenommen, wenn keine Hinweise auf Verzerrungen identifiziert wurden, die die Aussagekraft der Ergebnisse bei einer Korrektur wesentlich verändert hätten. Die Methodik zur Bewertung des Verzerrungspotenzials orientierte sich dabei an den Bewertungsbögen, die in Anhang 4-F detailliert beschrieben sind.

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)[2]. Die Darstellung nicht randomisierter Interventionsstudien und epidemiologischer Beobachtungsstudien soll mindestens den Anforderungen des TREND-[3] beziehungsweise

[2] Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. BMJ 2010; 340: c332.
[3] 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.

STROBE-Statements[4] 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.

Sofern im EU-Dossier eine Beschreibung hinterlegt ist, nach welchen Standards und mit welchen Informationen (Items) das Design und die Methodik der eingeschlossenen Studien dargestellt sind, und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Die Bewertung der für die Nutzenbewertung relevanten Studien basierte auf dem CONSORT-Statement, wie in Anhang 4-E dargestellt. Zur Beschreibung des Designs und der Methodik der in die Nutzenbewertung einbezogenen Studien wurden gemäß CONSORT-Statement und Abschnitt 4.3.1 jeweils das Studienprotokoll, der statistische Analyseplan sowie der Studienbericht als Informationsquellen herangezogen [14; 15; 17-20].

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.

Benennen Sie die Patientencharakteristika und patientenrelevanten Endpunkte, die in den relevanten Studien erhoben wurden. 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 (zum Beispiel zur Validierung der eingesetzten Fragebögen). Geben Sie für den jeweiligen Endpunkt an, ob unterschiedliche Operationalisierungen innerhalb der Studien und zwischen den Studien verwendet wurden. Benennen Sie die für die Bewertung herangezogene(n) Operationalisierung(en) und begründen Sie die Auswahl. Beachten Sie bei der Berücksichtigung von Surrogatendpunkten Abschnitt 4.5.4.

Sofern zur Berechnung von Ergebnissen von Standardverfahren und –software abgewichen wird (insbesondere beim Einsatz spezieller Software oder individueller Programmierung), sind

[4] 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.

die Berechnungsschritte und gegebenenfalls verwendete Software explizit abzubilden. Insbesondere der Programmcode ist in lesbarer Form anzugeben.

Sofern Informationen zu Patientencharakteristika, zu in den relevanten Studien erhobenen patientenrelevanten Endpunkten und/oder zu Angaben zu Berechnungsschritten und zur verwendeten Software (insbesondere zum Programmcode) im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Die Bewertung des Ausmaßes des Zusatznutzens von Zuranolon basiert auf den randomisierten, kontrollierten klinischen Phase-3-Studien ROBIN und SKYLARK. Die Studienergebnisse sind in Form einer Meta-Analyse in Anhang 4-G dargestellt, während die Studienergebnisse der einzelnen Studien in Anhang 4-H und Anhang 4-I zu finden sind.

4.2.5.2.1 Analysepopulationen

Für die Studien SKYLARK und ROBIN wurden folgende Populationen für die Analyse vordefiniert:

ITT-Population:

Alle randomisierten Patientinnen, eingeteilt nach ihrer ursprünglich randomisierten Behandlungsgruppe, unabhängig davon, ob sie tatsächlich das Studienmedikament erhalten haben. In den ursprünglichen Studiendokumenten der ROBIN-Studie kommt mit dem „Full Analysis Set“ eine zugeschnittene Form der ITT vor, während in der SKYLARK-Studie die ITT-Population als „Randomized Set“ bezeichnet wird.

Safety-Population:

Alle Patientinnen die mindestens eine Dosis der Studienmedikation erhalten haben. Die überwiegende Mehrheit, jedoch nicht alle, der randomisierten Patientinnen erhielten tatsächlich eine Dosis der Studienmedikation, deshalb ist die Safety-Population etwas kleiner als die ITT-Population. Zudem erhielten in der ROBIN-Studie ein Teil der Patientinnen eine andere Studienmedikation als ihnen zur Randomisierung zugewiesen wurde. In der Safety-Population sind die Patientinnen dem Behandlungsarm mit der Behandlung zugeordnet, die sie tatsächlich erhielten.

4.2.5.2.2 Demografische und krankheitsspezifische Patientencharakteristika

Für die in die Nutzenbewertung eingeschlossenen Studien ROBIN und SKYLARK werden folgende Charakteristika der Patientinnen zu Baseline dargestellt:

Tabelle 4-5: Demografischen und krankheitsspezifische Charakteristika die für die Studienpopulationen der Studien ROBIN und SKYLARK abgefragt wurden

Studien: ROBIN und SKYLARK	**Statistische Angaben oder Ausprägungen**	**Erhoben und dargestellt für:**	
		ROBIN	**SKYLARK**
Demografische Charakteristika			
Alter zum Zeitpunkt der Einwilligung (Jahre)	n; Mittelwert (SD); Median; Q1, Q3; Min, Max	X	X
Alter zum Zeitpunkt der Einwilligung (Jahre) - Kategorie; n (%)	18 – 24 Jahre; 25 – 45 Jahre	X	X
Ethnizität; n (%)	Nicht hispanisch oder lateinamerikanisch; Hispanisch oder lateinamerikanisch; Nicht angegeben	X	X
Abstammung; n (%)	Weiß; Schwarz oder Afroamerikanisch; Mehr als eine Ethnie; Andere; amerikanischer Ureinwohner oder Ureinwohner Alaskas; Asiatisch; Hawaiianischer Ureinwohner oder anderer pazifischer Inselbewohner; Nicht angegeben	X	X
Abstammung - Kategorie; n (%)	Schwarz oder Afroamerikanisch; Weiß; Andere	X	X
Bildungsniveau; n (%)	≤ 12. Klasse, kein Abschluss; Abschluss der 12. Klasse oder GED; Einige College-Kurse, aber kein Abschluss; Beruflicher Associate-Abschluss; Akademischer Associate-Abschluss; Bachelor-Abschluss; Master-Abschluss; Beruflicher Abschluss; Doktorgrad	--	X
Familien-/Ehestand; n (%)	Verheiratet; Nie verheiratet; Geschieden; Lebenspartner; Ledig	--	X
Beschäftigungsstatus; n (%)	Vollzeit (≥ 35 h pro Woche); Teilzeit (< 35 h pro Woche); Im Ruhestand; Arbeitslos; Andere	--	X
Land; n (%)	USA; Rest der Welt	--	X
Körpergröße (cm)	n; Mittelwert (SD); Median; Q1, Q3; Min, Max	X	X
Gewicht (kg)	n; Mittelwert (SD); Median; Q1, Q3; Min, Max	X	X
BMI (kg/m^2)	n; Mittelwert (SD); Median; Q1, Q3; Min, Max	X	X
BMI (kg/m^2) - Kategorie; n (%)	≤ 18,4; 18,5 – 24,9; 25 – 29,9; ≥ 30	X	X
BMI (kg/m^2) - Kategorie 2; n (%)	≤ 24,9; 25 – 29,9; ≥ 30	X	X
Krankheitsspezifische Charakteristika			

Studien: ROBIN und SKYLARK	Statistische Angaben oder Ausprägungen	Erhoben und dargestellt für:	
		ROBIN	SKYLARK
Antidepressiva zu Baseline; n (%)	Ja; Nein	X	X
Baseline-HAMD-17-Gesamtpunktzahl - Kategorie; n (%)	< 28 (Median des HAMD-17-Gesamtwertes zu Baseline); ≥ 28 (Median des HAMD-17-Gesamtwertes zu Baseline)	X	X
Beginn der PPD; n (%)	3. Trimester; Nach der Geburt	X	X
PPD in der Anamnese; n (%)	Erste Episode; Wiederkehrende Episode	--	X
PPD in der Familienanamnese; n (%)	Ja; Nein/Unbekannt	X	X
COVID-19 in der Anamnese; n (%) [2]	COVID-19; Vermutetes COVID-19; Patientenisolation; SARS-CoV-2-Test positiv; Krankenhausaufenthalt; Nicht betroffen	--	X
Depression mit erhöhter Angst; basierend auf der HAMD-17-Angst-Subskala; n (%)	≥ 7 (Normalisierter Score ≥ 39); < 7 (Normalisierter Score < 39)	--	X
Depression mit erhöhter Angst; basierend auf der HAM-A-Gesamtpunktzahl; n (%)	≥ 20; < 20	X	X
Einnahme von Antidepressiva jemals; n (%)	Ja; Nein	X	X
Begleitende Verfahren - Psychotherapie; n (%)	Ja; Nein	--	X
X: Charakteristikum wurde erhoben; --: Charakteristikum wurde nicht erhoben. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.			

4.2.5.2.3 Patientenrelevante Endpunkte

Für die Bewertung des medizinischen Zusatznutzens von Zuranolon im vorliegenden Anwendungsgebiet werden die Endpunkte der im SGB V § 35b und der AM-NutzenV definierten patientenrelevanten Endpunktkategorien Mortalität, Morbidität und Sicherheit auf Grundlage einer Meta-Analyse basierend auf den Studien SKYLARK und ROBIN herangezogen.

Tabelle 4-6: Übersicht über patientenrelevante Endpunkte und Analysen

Kategorie	Endpunkt	Operationalisierung	Meta-Analyse	ROBIN	SKYLARK
Mortalität	**Todesfälle**	Anzahl der Todesfälle im Rahmen der Sicherheitsendpunkte	nein	Sowohl in der ROBIN-Studie als auch in der SKYLARK-Studie gab es keine Todesfälle.	
		Anzahl der Patientinnen mit	Abschnitt	Anhang 4-	

Kategorie	Endpunkt	Operationalisierung	Meta-Analyse	ROBIN	SKYLARK
Morbidität	**HAMD-17**	Ansprechen	4.3.1.3.1.2.1	H	Anhang 4-I
		Zeit bis zum ersten Ansprechen (+ Kaplan-Meier-Plot)			
		Anzahl der Patientinnen mit Remission			
		Zeit bis zur ersten Remission (+ Kaplan-Meier-Plot)			
	MADRS	Anzahl der Patientinnen mit Ansprechen	Abschnitt 4.3.1.3.1.2.2	Anhang 4-H	Anhang 4-I
		Zeit bis zum ersten Ansprechen (+ Kaplan-Meier-Plot)			
		Anzahl der Patientinnen mit Remission			
		Zeit bis zur ersten Remission (+ Kaplan-Meier-Plot)			
	EPDS	Anzahl der Patientinnen mit Ansprechen	Abschnitt 4.3.1.3.1.2.3	Anhang 4-H	Anhang 4-I
		Zeit bis zum ersten Ansprechen(+ Kaplan-Meier-Plot)			
		Veränderung des EPDS-Gesamtwertes im Vergleich zu Baseline (+ Grafik über Zeit)	Anhang 4-G		
		Anzahl der Patientinnen mit Verschlechterung	Anhang 4-G		
		Zeit bis zur ersten Verschlechterung (+ Kaplan-Meier-Plot)	Anhang 4-G		
	PHQ-9	Anzahl der Patientinnen mit Ansprechen	Abschnitt 4.3.1.3.1.2.4	Anhang 4-H	Anhang 4-I
		Zeit bis zum ersten Ansprechen (+ Kaplan-Meier-Plot)			
		Veränderung des PHQ-9-Gesamtwertes im Vergleich zu Baseline (+ Grafik über Zeit)	Anhang 4-G		
		Anzahl der Patientinnen mit Verschlechterung	Anhang 4-G		
		Zeit bis zur ersten Verschlechterung (+ Kaplan-Meier-Plot)	Anhang 4-G		
	HAM-A	Anzahl der Patientinnen mit Ansprechen	Abschnitt 4.3.1.3.1.2.5	Anhang 4-H	Anhang 4-I
		Zeit bis zum ersten Ansprechen (+ Kaplan-Meier-Plot)			

Kategorie	Endpunkt	Operationalisierung	Meta-Analyse	ROBIN	SKYLARK
Lebens-qualität	**SF-36**	Anzahl der Patientinnen mit Ansprechen im PCS	nein	Abschnitt 4.3.1.3.1.3.1	nein
		Zeit bis zum ersten Ansprechen im PCS (+ Kaplan-Meier-Plot)			
		Anzahl der Patientinnen mit Ansprechen im MCS			
		Zeit bis zum ersten Ansprechen im MCS (+ Kaplan-Meier-Plot)			
		Anzahl der Patientinnen mit Ansprechen in folgenden Subskalen nach SF-36: • körperliche Schmerzen • soziale Funktionsfähigkeit • psychisches Wohlbefinden			
		Zeit bis zum ersten Ansprechen in folgenden Subskalen nach SF-36 (+ Kaplan-Meier-Plot) • körperliche Schmerzen • soziale Funktionsfähigkeit • psychisches Wohlbefinden			
		Veränderung des PCS-Gesamtwertes (+ Grafik über Zeit)	nein	Anhang 4-H	nein
		Veränderung des MCS-Gesamtwertes (+ Grafik über Zeit)			
		Veränderung jeder einzelnen Dimension (+ Grafik über Zeit)			
		Anzahl der Patientinnen mit Verschlechterung im PCS			
		Zeit bis zur ersten Verschlechterung im PCS (+ Kaplan-Meier-Plot)			
		Anzahl der Patientinnen mit Verschlechterung im MCS			
		Zeit bis zur ersten Verschlechterung im MCS (+ Kaplan-Meier-Plot)			
		Anzahl der Patientinnen mit Verschlechterung in folgenden Subskalen nach SF-36:			

Kategorie	Endpunkt	Operationalisierung	Meta-Analyse	ROBIN	SKYLARK
		• körperliche Schmerzen • soziale Funktionsfähigkeit • psychisches Wohlbefinden			
		Zeit bis zur ersten Verschlechterung in folgenden Subskalen nach SF-36 (+ Kaplan-Meier-Plot) • körperliche Schmerzen • soziale Funktionsfähigkeit • psychisches Wohlbefinden			
	BIMF	Anzahl der Patientinnen mit Ansprechen im BIMF Gesamtwert	nein	Abschnitt 4.3.1.3.1.3.2	nein
		Zeit bis zum ersten Ansprechen im BIMF Gesamtwert (+ Kaplan-Meier-Plot)			
		Anzahl der Patientinnen mit Ansprechen auf den folgenden Subskalen des BIMF: • Kompetenz der Mutter • Bedürfnisse der Mutter			
		Zeit bis zum ersten Ansprechen auf den folgenden Subskalen des BIMF (+ Kaplan-Meier-Plot): • Kompetenz der Mutter • Bedürfnisse der Mutter			
		Veränderung des BIMF Gesamtwertes im Vergleich zu Baseline (+ Grafik über Zeit).	nein	Anhang 4-H	nein
		Veränderung auf den Subskalen „Kompetenz der Mutter“ und „Bedürfnisse der Mutter“ im Vergleich zu Baseline (+ Grafik über Zeit)			
		Anzahl der Patientinnen mit Verschlechterung im BIMF Gesamtwert			
		Zeit bis zur ersten Verschlechterung im BIMF Gesamtwert (+ Kaplan-Meier-Plot)			
		Anzahl der Patientinnen mit Verschlechterung auf den			

Kategorie	Endpunkt	Operationalisierung	Meta-Analyse	ROBIN	SKYLARK
		folgenden Subskalen des BIMF: • Kompetenz der Mutter • Bedürfnisse der Mutter			
		Zeit bis zur ersten Verschlechterung auf den folgenden Subskalen des BIMF (+ Kaplan-Meier-Plot): • Kompetenz der Mutter • Bedürfnisse der Mutter			
Sicherheit	C-SSRS	Anzahl der Patientinnen mit mindestens einer „Ja“ Antwort nach Baseline	nein	Abschnitt 4.3.1.3.1.4.1	Abschnitt 4.3.1.3.1.4.1
	UE	Anzahl der Patientinnen mit mindestens einem UE	Abschnitt 4.3.1.3.1.4.2	Anhang 4-H	Anhang 4-I
		Jegliche UE nach SOC und PT			
	Schwere UE	Anzahl der Patientinnen mit mindestens einem schweren UE	Abschnitt 4.3.1.3.1.4.3	Anhang 4-H	Anhang 4-I
		schwere UE nach SOC und PT			
	SUE	Anzahl der Patientinnen mit mindestens einem SUE	Abschnitt 4.3.1.3.1.4.4	Anhang 4-H	Anhang 4-I
		SUE nach SOC und PT			
	Therapie-abbruch aufgrund von UE	Anzahl der Patientinnen mit Therapieabbruch aufgrund von UE	Abschnitt 4.3.1.3.1.4.5	Anhang 4-H	Anhang 4-I
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.					

4.2.5.2.4 Patientenrelevanz, Validität und Operationalisierung der Endpunkte

4.2.5.2.4.1 Mortalität

Die Mortalität wird im vorliegenden Dossier anhand der Anzahl der während der Studie aufgetretenen Todesfälle dargestellt, welche über den Studienzeitraum im Rahmen der Sicherheitsanalysen erfasst wurden. Als ein in der Verfahrensordnung definierter patientenrelevanter und objektiver Endpunkt ist die Mortalität von besonderer Bedeutung. Gemäß § 2 Satz 3 der AM-NutzenV gilt der Nutzen eines Arzneimittels als nachgewiesen, wenn ein patientenrelevanter therapeutischer Effekt, insbesondere in Bezug auf die Verlängerung des Überlebens, vorliegt.

4.2.5.2.4.2 Morbidität

HAMD-17

Die HAMD-17 besteht aus 17 Items und wird verwendet, um das Vorhandensein und die Schwere von verschiedenen Symptomen der Depression zu evaluieren. Sie wurde durch

behandelnde Ärzt*innen als klinisches Interview zur Fremdbeurteilung der Schwere psychischer sowie somatischer Depressionssymptome bei depressiven Patient*innen entwickelt. Die 17 Items umfassen 17 unterschiedliche Symptome wie „Depressive Stimmung“, „Suizidalität“ oder „Gewichtsverlust“. Jedes Item kann einen Wert von entweder 0 – 2 oder 0 – 4 annehmen (Gesamtspanne 0 – 52), wobei höhere Werte eine höhere Symptomschwere bedeuten [21]. Der resultierende HAMD-17-Gesamtwert wird wie folgt interpretiert: Ein Wert von ≤ 8 weist auf keine Depression oder eine remittierte Depression hin, 9–16 entspricht einem leichten depressiven Syndrom, 17 – 24 einem mittelgradigen depressiven Syndrom und ein Wert von ≥ 25 einem schweren depressiven Syndrom [13]. Weitere Details zur Operationalisierung und Auswertung des Endpunktes sind in Abschnitt 4.3.1.3.1.2.1 zu finden. Die HAMD-17 lässt sich als ein Instrument einordnen, das schwerwiegende Symptome in den Vordergrund stellt. Zwar enthält die Skala auch Items zu somatischen Aspekten wie Schlafstörungen oder Appetitverlust, doch liegt ihr Schwerpunkt auf psychopathologisch zentralen und potenziell gefährlichen Dimensionen wie ausgeprägter Depressivität, Schuldgefühlen, Suizidgedanken sowie psychomotorischen Veränderungen.

Patientenrelevanz und Validität

Die Nationale Versorgungsleitlinie (NVL) weist die HAMD-Skala als bewährte Skala zur Beurteilung der Verbesserung der depressiven Symptomatik aus [13; 22]. Sie wurde bereits 1960 von Max Hamilton eingeführt und in dem Zuge erstmals auf ihre Validität überprüft [21]. Auch bei nachfolgenden Untersuchungen wies die HAMD-17 eine gute inhaltliche Validität sowie eine gute konvergente Validität im Sinne von Korrelationen mit parallelen Depressionstests auf. Die interne Reliabilität beträgt 0,64 (Cronbach’s Alpha) und die Interrater-Reliabilität (intra-class-correlation) 0,78 [23]. Darüber hinaus konnte auch eine gute Validität der HAMD-Skala für Personen mit PPD und eine konsistente Reliabilität über diverse Populationen hinweg nachgewiesen werden [24-26].

Auch das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) hat bei seiner Nutzenbewertung von SSRI und Selektiven Serotonin-Noradrenalin-Wiederaufnahmehemmern (SSNRI) bei Patientinnen mit Depressionen die HAMD-17 als relevantes Zielkriterium herangezogen. Der Schwellenwert der HAMD-17 lag hier bei ≤ 7 für eine Remission wie bereits 1991 von Frank *et al*. vorgeschlagen, während das Ansprechen über eine 50 %-Reduktion der HAMD-17-Gesamtwert von Baseline definiert wurde [27; 28]. Darüber hinaus hat der G-BA die HAMD-17 als Standard-Messinstrument im Kontext der Beratung zu Zuranolon im Anwendungsgebiet PPD bezeichnet [29].

MADRS

Die MADRS ist ein validiertes Instrument zur Messung von Symptomveränderungen bei Depressionen. Sie wird häufig eingesetzt, um den Schweregrad depressiver Erkrankungen zu quantifizieren. Die Erhebung erfolgt durch unabhängige Prüfärztinnen und Prüfärzte auf Basis eines semi-strukturierten Interviews und umfasst 10 Items wie z. B. „beobachtbare Traurigkeit“, „Konzentrationsschwierigkeiten“ oder „Suizidgedanken“. Jedes Item wird entsprechend seiner Ausprägung von 0 (nicht vorhandene oder normale Symptomatik) bis 6 Punkte (schwerwiegende oder kontinuierlich vorhandene Symptomatik) bewertet. Höhere

Scores stehen für eine schwerwiegendere depressive Erkrankung, wobei der Gesamtwert maximal 60 Punkte beträgt [30]. Die MADRS konzentriert sich stark auf zentrale psychopathologische Merkmale wie depressive Grundstimmung, Traurigkeit, Gefühllosigkeit, pessimistische Gedanken und Suizidalität. Leichte oder unspezifische körperliche Symptome sind in der MADRS kaum enthalten. Dadurch wird vor allem die Intensität der „schweren" depressiven Symptomatik gut abgebildet. In der aktuellen NVL werden folgende Schwellenwerte zur Beurteilung der Schwere genannt: Ein MADRS-Score von 0 – 6 deutet darauf hin, dass kein depressives Syndrom oder eine Remission vorliegt, 7 – 19 entspricht einem leichten depressiven Syndrom, 20 – 34 einem mittelgradigen depressiven Syndrom, und ein Score von ≥ 35 weist auf ein schweres depressives Syndrom hin [13]. Weitere Details zur Operationalisierung und Auswertung des Endpunktes sind in Abschnitt 4.3.1.3.1.2.2 zu finden.

Patientenrelevanz und Validität

Die MADRS wurde nach der HAMD-17 entwickelt, ist besonders einfach aufgebaut und zielt darauf ab die Veränderung der depressiven Symptomatik bei der Behandlung mit Antidepressiva sensitiv zu messen [30]. Ihr Haupteinsatzbereich liegt in klinischen Studien, in denen sie zunehmend an Bedeutung gewinnt und als Outcome-Kriterium empfohlen wird [31].

Die Skala zeichnet sich durch ihre hohe Änderungssensitivität, eine starke Interrater-Reliabilität sowie eine hohe Korrelation mit anderen etablierten Messinstrumenten der Depression, wie der HAMD, aus [32-34]. Die Reliabilität und Validität der Schwellenwerte zur Diagnose und Bewertung des Schweregrads einer Depression mittels MADRS sind ebenfalls gegeben [13; 35-38]. Als validiertes und etabliertes Instrument zur Einschätzung der depressiven Symptomatik wurde die MADRS sowohl in einer Nutzenbewertung des IQWiG zur Bewertung in Deutschland zugelassener SSRI und SSNRI bei Patienten mit Depression verwendet als auch in beiden Nutzenbewertungsverfahren für den Wirkstoff Esketamin herangezogen [27; 39; 40].

EPDS

Die EPDS wurde speziell entwickelt, um PPD festzustellen und das Ansprechen auf Behandlungen oder die Schwere der Symptome über die Zeit abzubilden. Der selbstberichtete Fragebogen schließt körperliche Symptome aus und umfasst 10 Items zur Gefühlslage der Patientin. Die Patientinnen können bei jedem Item anhand von 4 Antwortmöglichkeiten bewerten, wie sie sich in den letzten 7 Tagen gefühlt haben. Jede Antwort hat einen Punktewert von 0 – 3 und die Werte der 10 Items werden zu einem Gesamtwert von 0 – 30 summiert. Je höher der Gesamtwert, desto schwerwiegender die Symptomatik der PPD [41]. Weitere Details zur Operationalisierung und Auswertung des Endpunktes sind in Abschnitt 4.3.1.3.1.2.3 zu finden.

Patientenrelevanz und Validität

Die EPDS ist ein selbstberichteter Endpunkt und ihre Validität und Reliabilität wurde in der Vergangenheit bereits vielfach untersucht und für gut befunden [42]. Auch der G-BA bewertet die EPDS als reliabel und valide und empfiehlt ihren Einsatz zur Identifikation der PPD und zur Kontrolle von Therapieeffekten [29].

PHQ-9

Der PHQ wurde zur Erfassung psychologischer Funktionen entwickelt. Sein Depressionsmodul, der PHQ-9, kann isoliert verwendet werden und eignet sich zur Selbstbeurteilung der Schwere depressiver Symptome. Der Fragebogen setzt sich aus 9 Items zusammen, die Symptome wie Anhedonie, psychomotorische Veränderungen oder Suizidgedanken abfragen und jeweils mit 0 – 3 Punkten bewertet werden können. Der Gesamtwert wird durch die Summe der Punkte gebildet, je höher der Score, desto schwerer die Depression [43]. Weitere Details zur Operationalisierung und Auswertung des Endpunktes sind in Abschnitt 4.3.1.3.1.2.4 zu finden.

Patientenrelevanz und Validität

Der PHQ-9 gilt als valides und reliables Instrument zur Erfassung der Schwere von Symptomen einer Depression [44]. Insbesondere ab einem Gesamtwert von ≥ 10 Punkten konnte eine Maximierung der Genauigkeit und Spezifität festgestellt werden [45]. Aus diesem Grund liegt die Relevanz des PHQ-9 darin, dass besonders stark ausgeprägte Depressionen erkannt und somit bestmöglich therapiert werden können. Im Beratungsgespräch stufte auch der G-BA den PHQ-9 als reliables und valides Instrument ein, dessen Stärke im Screening der PPD liegt und der darüber hinaus auch für die Kontrolle von Therapieeffekten im Rahmen einer Studie als geeignet erscheint [29]. In der Vergangenheit wurde der PHQ-9 im Rahmen der Neubewertung von Esketamin in Kombination mit SSRI oder SNRI zur Behandlung therapieresistenter MDD bereits als patientenrelevant anerkannt und in der Nutzenbewertung berücksichtigt [40].

HAM-A

Die HAM-A bewertet anhand von 14 Items die Schwere von Angstzuständen [46-49]. Jedes der 14 Items wird durch eine Reihe von Symptomen definiert und misst sowohl psychische Angst (geistige Erregung und psychische Belastung) als auch somatische Angst (körperliche Beschwerden im Zusammenhang mit Angstzuständen) der vergangenen Woche. Die Bewertung der HAM-A erfolgt mittels eines Punktsystems, indem Werte von 0 (nicht vorhanden) bis 4 (sehr schwer) zugewiesen werden. Der Gesamtwert wird durch die Summe der einzelnen Antworten gebildet und reicht von 0 – 56, wobei < 17 einen leichten Schweregrad, 18 – 24, einen leichten bis mittelschweren Schweregrad und 25 – 30 einen mittelschwerer bis schwerer Schweregrad anzeigt. Weitere Details zur Operationalisierung und Auswertung des Endpunktes sind in Abschnitt 4.3.1.3.1.2.5 zu finden.

Patientenrelevanz und Validität

Die Skala wird als reliabel und valide beschrieben [27; 31]. In einer Nutzenbewertung des IQWiG zur Bewertung in Deutschland zugelassener SSRI und SSNRI bei Patienten mit Depression, wurde die HAM-A verwendet um Angstsymptome zu evaluieren [27]. Auch im Zuge des vergangenen Beratungsgesprächs stufte der G-BA die HAM-A als Standardinstrument ein [29].

4.2.5.2.4.3 Lebensqualität

SF-36

Der Short Form 36 Health Survey (SF-36) ist ein aus 36 Fragen bestehender, multidimensionaler, validierter und generischer Lebensqualitätsfragebogen. Der SF-36 besteht aus 8 Domänen die wiederum in unterschiedlichen Summenscores, dem körperlichen Summenscore (Physical Component Summary; PCS) und psychischen Summenscore (Mental Component Summary; MCS) zusammengefasst werden. Weitere Details zur Operationalisierung und Auswertung des Endpunktes sind in Abschnitt 4.3.1.3.1.3.1 zu finden.

Patientenrelevanz und Validität

Die gesundheitsbezogene Lebensqualität der Patientinnen wird anhand des SF-36 adäquat operationalisiert und ist entsprechend § 2 Satz 3 AM-NutzenV als patientenrelevant zu betrachten. In der Vergangenheit wurde der SF-36 vom G-BA in vielen Verfahren bereits als patientenrelevant anerkannt und unter anderem zur Bestimmung des Zusatznutzens von Esketamin zur Behandlung therapieresistenter MDD herangezogen[40].

Der SF-36 ist ein standardisierter Fragebogen mit hoher interner und externer Validität. Er gehört zu den am weitesten verbreiteten Instrumenten zur Messung der generellen gesundheitsbezogenen Lebensqualität in klinischen Studien und gilt als etabliertes und validiertes Instrument zur Erhebung der generischen Lebensqualität. Der SF-36 wurde für den Einsatz in zahlreichen Ländern und Indikationen validiert und ist in den entsprechenden Sprachen verfügbar[50-52].

BIMF

Der BIMF ist ein 20 Item umfassender Fragebogen, der zur Selbsteinschätzung der Mutterrolle/ mütterlichen Funktionsfähigkeit für den Zeitraum der ersten 12 Monate nach der Geburt entworfen wurde. Er deckt eine breite Palette funktionaler Bereiche ab, darunter Selbstfürsorge, Säuglingspflege, Mutter-Kind-Interaktion, psychisches Wohlbefinden der Mutter, soziale Unterstützung, Management und Anpassung. Jedes Item wird auf einer Skala von 0 (stimme überhaupt nicht zu) bis 6 (stimme voll und ganz zu) bewertet. Der Gesamtwert des BIMF wird als Summe der 20 Einzelbewertungen berechnet und reicht von 0 – 120 Punkten. Niedrigere Scores weisen auf eine eingeschränkte Funktion in der Mutterrolle hin [53]. Weitere Details zur Operationalisierung und Auswertung des Endpunktes sind in Abschnitt 4.3.1.3.1.3.2 zu finden.

Patientenrelevanz und Validität

Der funktionale Status von Müttern ist in den ersten 12 Monaten nach der Geburt entscheidend, da diese Zeit eine wichtige Phase für Mutter und Kind darstellt. Mütter übernehmen meist die Hauptverantwortung für die Pflege des Kindes, einschließlich Aufgaben wie Füttern, Windelwechsel und Arztbesuche. Die Qualität der Mutter-Kind-Interaktion beeinflusst die Entwicklung des Säuglings maßgeblich. Eine hohe mütterliche Funktionsfähigkeit korreliert mit positiven Entwicklungsergebnissen, während eine eingeschränkte Funktionsfähigkeit die optimale Entwicklung des Kindes beeinträchtigen könnte [53]. Die PPD erschwert die Erfüllung der Aufgaben, die mit einer Mutterschaft einhergehen maßgeblich mit den

entsprechenden Konsequenzen für die Mutter-Kind-Beziehung und die Entwicklung des Säuglings. Zur Entwicklung des BIMF wurden über 30 Mütter, die innerhalb des vorherigen Jahres ein Kind geboren haben, interviewt und anschließend ein Komitee mit Expertinnen und Experten hinzugezogen. Im Nachgang wurde die Reliabilität und Validität des BIMF psychometrisch in einem klinischen Setting mit PPD-Patientinnen getestet. Schließlich konnte festgestellt werden, dass der BIMF reliable und valide ist [53].

4.2.5.2.4.4 Sicherheit

C-SSRS

Die Suizidalität wurde mittels C-SSRS erfasst. Die C-SSRS wurde durch die Patientinnen berichtet und ist ein etabliertes Fragebogeninstrument zur standardisierten Untersuchung von suizidalen Gedanken und Verhalten. Der Fragebogen besteht aus zwei Domänen, Suizidgedanken und suizidales Verhalten mit jeweils 5 Fragen und jeweils zwei Antwortmöglichkeiten (ja/nein). Eine zusätzliche "Ja/Nein"-Frage wird verwendet, um zu erfassen, ob die Person bei Post-Baseline-Visiten selbstverletzendes Verhalten ohne suizidale Absicht zeigte. Für die Studienvisiten nach Baseline wurde eine „Since last visit version“ der C-SSRS verwendet.

Patientenrelevanz und Validität

Suizidalität und selbstverletzendes Verhalten sind patientenrelevante Endpunkte. Die C-SSRS ist ein standardisiertes Instrument zur Bewertung von Selbstmordverhalten und -vorstellungen und sie wird u. a. zur strengen Überwachung der Suizidalität in klinischen Studien eingesetzt. Die C-SSRS hat sich in mehreren Untersuchungen als valides Instrument zur Erfassung des Suizidalität erwiesen [54-56].

UE

Die Sicherheit und Verträglichkeit von Zuranolon wurden anhand der Häufigkeit des Auftretens und des Schweregrads von unerwünschten Ereignissen beurteilt.

Patientenrelevanz und Validität

Die Beurteilung der Sicherheit eines Medikaments ist ein grundlegender Bestandteil der Nutzenbewertung und gemäß AM-NutzenV § 2 Satz 3 patientenrelevant, da ein unmittelbarer Einfluss auf den Gesundheitszustand der Patientinnen besteht.

Alle UE und SUE wurden gemäß MedDRA (Medical Dictionary for Regulatory Activities) Terminologie nach SOC und bevorzugtem Term PT klassifiziert. Die Erhebung und Dokumentation der UE wurden demnach nach allgemein anerkannten internationalen und unabhängigen Standards vorgenommen.

4.2.5.3 Meta-Analysen

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse quantitativ zusammengefasst werden, wenn die Studien aus medizinischen (zum Beispiel Patientengruppen) und methodischen (zum Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde beziehungsweise warum einzelne Studien gegebenenfalls nicht in die Meta-Analyse einbezogen wurden. Für Meta-Analysen soll die im Folgenden beschriebene Methodik eingesetzt werden.

Für die statistische Auswertung sollen primär die Ergebnisse aus ITT-Analysen, so wie sie in den vorliegenden Dokumenten beschrieben sind, verwendet werden. Die Meta-Analysen sollen in der Regel auf Basis von Modellen mit zufälligen Effekten nach der Knapp-Hartung-Methode mit der Paule-Mandel-Methode zur Heterogenitätsschätzung[5] erfolgen. Im Fall von sehr wenigen Studien ist die Heterogenität nicht verlässlich schätzbar. Liegen daher weniger als fünf Studien vor, ist auch die Anwendung eines Modells mit festem Effekt oder eine qualitative Zusammenfassung in Betracht zu ziehen. Kontextabhängig können auch alternative Verfahren wie zum Beispiel Bayes'sche Verfahren, Methoden aus dem Bereich der generalisierten linearen Modelle oder das Beta-Binomialmodell[6,7] in Erwägung gezogen werden.

Falls die für eine Meta-Analyse 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 Meta-Analysen primär sowohl anhand des Odds Ratios als auch des Relativen Risikos 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 verwendet[8] 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 geeigneter statistischer Maße auf Vorliegen von Heterogenität[9, 5] erfolgen. Die Heterogenitätsmaße sind unabhängig von dem Ergebnis der

[5] Veroniki AA, Jackson D, Viechtbauer W, Bender R, Knapp G, Kuss O et al. Recommendations for quantifying the uncertainty in the summary intervention effect and estimating the between-study heterogeneity variance in random-effects meta-analysis. Cochrane Database Syst Rev 2015: 25-27.
[6] Lilienthal J, Sturtz S, Schurmann C et al. Bayesian random-effects meta-analysis with empirical heterogeneity priors for application in health technology assessment with very few studies. Res Syn Methods 2024; 15: 275-287.
[7] IQWiG. Allgemeine Methoden, Version 7.0 vom 19.09.2023 [online]. IQWiG, Köln,2023. https://www.iqwig.de/de/methoden/methodenpapier.3020.html.
[8] 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.
[9] Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003;327(7414):557-560.

Untersuchung auf Heterogenität immer anzugeben. Ist die Heterogenität der Studienergebnisse nicht bedeutsam (zum Beispiel p-Wert für Heterogenitätsstatistik ≥ 0,05), soll der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt werden. Bei bedeutsamer Heterogenität sollen die Ergebnisse nur in begründeten 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 Meta-Analysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.

Sofern Informationen zur eingesetzten Methodik für Meta-Analysen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Prüfung der Studien vor der Meta-Analyse

Zur Bewertung der Studienpopulationen wurden die demografischen und krankheitsspezifischen Charakteristika zu Baseline beider Studien in Abschnitt 4.3.1.2.1 detailliert dargestellt und anschließend miteinander verglichen. Potenzielle Unterschiede werden im selben Abschnitt diskutiert.

Methodik der Meta-Analyse und Lebensqualität (nur RCT ROBIN)

Für die Erstellung des vorliegenden Dossiers standen Daten auf Patientenebene zur Verfügung. Meta-Analysen, die auf individuellen Patientendaten (IPD) basieren, gelten laut der Cochrane Collaboration als Goldstandard für die Zusammenfassung klinischer Daten [57]. Auch das IQWiG hebt in seinen aktuellen Methoden hervor, dass IPD-basierte Meta-Analysen eine höhere Aussagekraft besitzen [58]. Aufgrund der sehr ähnlichen Studienprotokolle der relevanten Studien können systematische Unterschiede in den Patientencharakteristika, Interventionen oder Endpunkten als Ursache für Heterogenität ausgeschlossen werden.

Heterogenitätstest

Zur Beurteilung der Heterogenität der Ergebnisse zwischen den Studien wird ein Interaktionstest herangezogen. In der Meta-Analyse (Heterogenitätstest) gibt es drei Arten von Analysen: MMRM, GEE und CoxPH). MMRM- und GEE-Modelle haben zwei Interaktionsterme – „Interaktion zwischen Studie und Behandlung“ und „Interaktion zwischen Studie, Zeitpunkt und Behandlung“. Der p-Wert ist der Durchschnitt von der Interaktion zwischen Behandlung und Studie über verschiedene Zeitpunkte hinweg.

Kontinuierliche Variablen und Time-to event (TTE)-Analysen

Die Analysen der kontinuierlichen Variablen wurde für patientenberichtete Endpunkte wie EPDS, PHQ-9, SF-36 und BIMF berechnet.

- MMRM für EPDS und PHQ-9*:* Dieses Modell berücksichtigt die Einnahme von Antidepressiva zu Baseline, den Baselinewert der Variablen, die Studie, die Behandlung, den Zeitpunkt der Beurteilung, die Interaktion zwischen Zeitpunkt und Studie sowie die Interaktion zwischen Zeitpunkt und Behandlung als feste Effekte mit unstrukturierter Kovarianzstruktur.

- MMRM für SF-36 und BIMF aus RCT ROBIN: Dieses Modell berücksichtigt die Einnahme von Antidepressiva zu Baseline, den Baselinewert der Variablen, die Behandlung, den Zeitpunkt der Beurteilung und die Interaktion zwischen Zeitpunkt und Behandlung als feste Effekte mit unstrukturierter Kovarianzstruktur.

Die Analyse der Zeit bis zum Ansprechen wurde für die Endpunkte HAMD-17, MADRS, EPDS, PHQ-9, HAM-A, SF-36 und BIMF berechnet. Zu den verwendeten statistischen Methoden gehören:

- Kaplan-Meier Methode: Der geschätzte Anteil der Probanden, die ein Ansprechen erzielten, wurde unter Verwendung der Kaplan-Meier-Produkt-Limit-Methode berechnet.

- CoxPH Meta-Analyse: Hazard Ratios (HRs), 95 %-Konfidenzintervalle (KIs) und p-Werte wurden aus einem Cox-Proportional-Hazards-Modell abgeleitet. Dieses Modell wurde hinsichtlich der Einnahme von Antidepressiva zu Baseline, der Baselinewerte, der Studie, der Behandlungsgruppe und bei Subgruppenanalysen der Subgruppe und der Interaktion zwischen Subgruppe und Behandlung angepasst.

- CoxPH für SF-36 und BIMF aus RCT ROBIN: HRs, 95 %-KI und p-Werte wurden aus einem CoxPH-Modell abgeleitet. Dieses Modell wurde hinsichtlich der Einnahme von Antidepressiva zu Baseline, der Baselinewerte und der Behandlungsgruppe.

- Zensurregeln: Die Probanden wurden am Tag ihrer letzten verfügbaren Bewertung zensiert, wenn das Ereignis von Interesse nicht beobachtet wurde.

Responder-Analysen:

Die Responder-Analysen für binäre Variablen wurden für die Endpunkte HAMD-17, MADRS, EPDS, PHQ-9, SF-36 und BIMF sowie Sicherheit berechnet, bei denen binäre Responder-Ergebnisse zu erwarten waren, Die verwendeten statistischen Methoden umfassen:

- GEE: Wird für binäre Antwortdaten verwendet und berücksichtigt Kovariablen wie die Einnahme von Antidepressiva zu Baseline, Baselinewerte, Studie, Behandlung, Zeitpunkt der Bewertung, die Interaktion zwischen Zeitpunkt und Studie sowie die Interaktion zwischen Zeitpunkt und Behandlung mit unstrukturierter Kovarianzstruktur.

- GEE für SF-36 und BIMF aus RCT ROBIN: Wird für binäre Antwortdaten verwendet und berücksichtigt Kovariablen wie die Einnahme von Antidepressiva zu Baseline, Baselinewerte, Behandlung, Zeitpunkt der Bewertung und die Interaktion zwischen Zeitpunkt und Behandlung mit unstrukturierter Kovarianzstruktur.

- Link-Funktionen und Verteilung: „Log"-Link-Funktion mit binomialer Verteilung und „Log"-Link-Funktion mit einer Poisson-Verteilung und „Log"-Link-Funktion mit binomialer Verteilung und unter Ausschluss des Terms „Interaktion zwischen Studie und Zeitpunkt und Behandlung".

- Umgang mit fehlenden Daten: Fehlende Werte bei den Besuchen wurden als „keine Antwort" imputiert. Für verschiedene Endpunkte galten unterschiedliche Definitionen hinsichtlich der Anzahl der akzeptierten fehlenden Werte und des Umgangs mit den fehlenden Daten. Diese Regeln kamen nur für den Endpunkt PHQ-9 zum Einsatz und sind in der Operationalisierung (Tabelle 4-43) detailliert beschrieben.

- Analyse der UE: Die Analyse der UE erfolgt analog zu den Wirksamkeitsanalysen.

Darstellung der Ergebnisse

Im Gegensatz zu Meta-Analysen mit aggregierten Daten werden die Ergebnisse einer IPD-Meta-Analyse üblicherweise nicht in Form von Forest-Plots dargestellt. Ebenso wird die zusätzliche Berechnung von Heterogenitätsmaßen, wie I^2 oder Cochran's Q, in der IPD-Meta-Analyse als nicht sinnvoll erachtet.

4.2.5.4 Sensitivitätsanalysen

Zur Einschätzung der Robustheit der Ergebnisse sollen Sensitivitätsanalysen hinsichtlich methodischer Faktoren durchgeführt werden. 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.

Sofern Informationen zu durchgeführten Sensitivitätsanalysen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Für die Einzelstudien ROBIN und SKYLARK wurden Sensitivitätsanalysen *a priori* definiert [14; 15]. Für die Bewertung des Zusatznutzens wird eine Meta-Analyse beider Studien herangezogen. Hierfür waren keine Sensitivitätsanalysen geplant. Somit liegen keine bewertungsrelevanten Sensitivitätsanalysen vor.

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 (zum Beispiel 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- beziehungsweise 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 Meta-Analysen unter Kategorisierung der Studien bezüglich der möglichen Effektmodifikatoren, sie sind deshalb zu bevorzugen. Es sollen, soweit sinnvoll, folgende Faktoren bezüglich einer möglichen Effektmodifikation berücksichtigt werden:

– Geschlecht
– Alter
– Krankheitsschwere beziehungsweise –stadium
– Zentrums- und Ländereffekte

Sofern für die Krankheitsschwere mehrere Definitionen beziehungsweise Operationalisierungen vorliegen, ist eine Auswahl zu begründen. Sollten sich aus den verfügbaren Informationen Anzeichen für weitere mögliche Effektmodifikatoren ergeben, sollten diese ebenfalls begründet einbezogen werden.

Die Ergebnisse von Subgruppenanalysen zu den im Studienprotokoll festgelegten Stratifikationsfaktoren sind immer darzustellen (zu ergänzenden Kriterien zur Darstellung siehe Abschnitt 4.3.1.3.2).

Grundsätzlich soll für die Definition beziehungsweise Operationalisierung der Subgruppen einschließlich der Trennwerte auf a priori geplante und in Studienunterlagen festgelegte Subgruppenanalysen zurückgegriffen werden.

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 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. Begründen Sie die durchgeführten Subgruppenanalysen beziehungsweise 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.

Sofern Informationen zu durchgeführten Subgruppenanalysen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Um die Abhängigkeit der Ergebnisse von bestimmten Patientenmerkmalen und weiteren Faktoren zu analysieren, wurden Subgruppenanalysen durchgeführt. Entsprechend der Modulvorlage werden die Subgruppen Alter, Krankheitsschwere bzw. -stadium dargestellt. Zusätzlich werden vollständige Subgruppen berücksichtigt, die in beiden Studien konsistent erfasst wurden. Im vorliegenden Dossier werden folgende Subgruppen, mit ihren Kategorien entsprechend der Präspezifizierung im SAP betrachtet:

Tabelle 4-7: Prädefinierte Subgruppen und zugehörige Kategorien der Studien ROBIN und SKYLARK

Subgruppe	**Kategorien**
Alter	18 – 24 Jahre; 25 – 45 Jahre
Schwere der Erkrankung – definiert durch Ausgangs-HAMD-17-Gesamtpunktzahl	< 28; ≥ 28 (basierend auf dem Median)
Abstammung	Schwarz oder Afroamerikanisch; Weiß
Antidepressiva-Nutzung zu Baseline	Ja; Nein
BMI – Kategorie (kg/m2)	≤ 24,9; 25 – 29,9; ≥ 30
Beginn der PPD	3. Trimester; Nach der Geburt
PPD in der Familienanamnese	Ja; Nein
PPD in der Anamnese	Erste Episode; Wiederkehrende Episode
Depression mit erhöhter Angst, basierend auf der HAM-A-Gesamtpunktzahl	< 20; ≥ 20
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.	

Das Geschlecht konnte nicht als Subgruppe betrachtet werden, da die Teilnehmenden gemäß Einschlusskriterien weiblich sein mussten. Außerdem konnten Ländereffekte nicht über entsprechende Subgruppen ermittelt werden, da die überwiegende Mehrheit der Patientinnen beider Studien aus den USA stammt.

Im Falle einer potenziellen Effektmodifikation ($p < 0,05$) der Meta-Analyse, wurden Subgruppenanalysen im Modul 4 dargestellt und anschließend diskutiert. Eine vollständige Darstellung aller Subgruppenanalysen, auch solcher mit nicht signifikantem Interaktionsterm, befindet sich in für die Meta-Analyse in Anhang 4-G. Die Interaktionstests und Subgruppenanalysen für die die Studien ROBIN und SKYLARK jeweils in Anhang 4-H und Anhang 4-I.

4.2.5.6 Indirekte Vergleiche

In den letzten Jahren wurden zahlreiche Methoden zur Durchführung indirekter Vergleiche entwickelt. Es besteht dabei internationaler Konsens, dass Vergleiche einzelner Behandlungsgruppen aus verschiedenen Studien ohne Bezug zu einem gemeinsamen Komparator (häufig als nicht adjustierte indirekte Vergleiche bezeichnet) regelhaft keine valide Analysemethode darstellen[10]. Eine Ausnahme stellt das Vorliegen eines dramatischen Effekts dar. An Stelle von nicht adjustierten indirekten Vergleichen sollen je nach Datenlage einfache adjustierte indirekte Vergleiche nach Bucher et al. (1997)[11] oder komplexere Netzwerk-Meta-Analysen (auch als „MTC Meta-Analysen" oder „Multiple Treatment Meta-Analysen" bezeichnet) für den simultanen Vergleich von mehr als zwei Therapien unter Berücksichtigung sowohl direkter als auch indirekter Vergleiche berechnet werden. Gängige Verfahren für Netzwerk-Meta-Analysen sind hierbei Bayes´sche Methoden nach Lu und Ades (2004)[12] sowie frequentistische Methoden nach Rücker (2012)[13].

Zur Durchführung frequentistischer Netzwerk-Meta-Analysen hat sich seit einiger Zeit das Programm netmeta[14] etabliert. Wie in paarweisen Meta-Analysen sollte auch bei Netzwerk-Meta-Analysen standardmäßig ein Modell mit zufälligen Effekten gewählt werden.

Alle Verfahren für indirekte Vergleiche gehen im Prinzip von den gleichen zentralen Annahmen aus. Hierbei handelt es sich um die Annahmen der Ähnlichkeit der eingeschlossenen Studien, der Homogenität der paarweisen Vergleiche und der Konsistenz zwischen direkter und indirekter Evidenz innerhalb des zu analysierenden Netzwerkes. Als Inkonsistenz wird dabei die Diskrepanz zwischen dem Ergebnis eines direkten und eines oder mehreren indirekten Vergleichen verstanden, die nur durch Zufall erklärbar ist[15].

Das Ergebnis eines indirekten Vergleichs kann maßgeblich von der Auswahl des Brückenkomparators beziehungsweise der Brückenkomparatoren abhängen. Als Brückenkomparatoren sind dabei insbesondere Interventionen zu berücksichtigen, für die sowohl zum bewertenden Arzneimittel als auch zur zweckmäßigen Vergleichstherapie mindestens eine direkt vergleichende Studie vorliegt (Brückenkomparatoren ersten Grades).

[10] Bender R, Schwenke C, Schmoor C, Hauschke D. Stellenwert von Ergebnissen aus indirekten Vergleichen - Gemeinsame Stellungnahme von IQWiG, GMDS und IBS-DR [online]. [Zugriff: 31.10.2016]. URL: http://www.gmds.de/pdf/publikationen/stellungnahmen/120202_IQWIG_GMDS_IBS_DR.pdf.
[11] Bucher HC, Guyatt GH, Griffith LE, Walter SD. The results of direct and indirect treatment comparisons in meta-analysis of randomized controlled trials. J Clin Epidemiol 1997; 50(6): 683-691.
[12] Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004; 23(20): 3105-3124.
[13] Rücker G. Network meta-analysis, electrical networks and graph theory. Res Synth Methods 2012; 3(4): 312-324.
[14] Balduzzi S, Rücker G, Nikolakopoulou A, Papakonstantinou T, Salanti G, Efthimiou O, Schwarzer G. netmeta: An R Package for Network Meta-Analysis Using Frequentist Methods.
Journal of Statistical Software 2023; 106(2):1-40.
[15] Schöttker B, Lühmann D, Boulkhemair D, Raspe H. Indirekte Vergleiche von Therapieverfahren. Schriftenreihe Health Technology Assessment Band 88, DIMDI, Köln, 2009.

Insgesamt ist es notwendig, die zugrunde liegende Methodik für alle relevanten Endpunkte genau und reproduzierbar zu beschreiben und die zentralen Annahmen zu untersuchen[16, 17, 18]

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

- *Benennung aller potentiellen Brückenkomparatoren ersten Grades und gegebenenfalls Begründung für die Auswahl.*
- *Genaue Spezifikation des statistischen Modells inklusive aller Modellannahmen. Bei Verwendung eines Bayes'schen Modells sind dabei auch die angenommenen A-priori-Verteilungen (falls informative Verteilungen verwendet werden, mit Begründung), die Anzahl der Markov-Ketten, die Art der Untersuchung der Konvergenz der Markov-Ketten und deren Startwerte und Länge zu spezifizieren.*
- *Art der Prüfung der Ähnlichkeit der eingeschlossenen Studien.*
- *Art der Prüfung der Homogenität der Ergebnisse direkter paarweiser Vergleiche.*
- *Art der Prüfung der Konsistenzannahme im Netzwerk.*
- *Bilden Sie den Code des Computerprogramms inklusive der einzulesenden Daten in lesbarer Form ab und geben Sie an, welche Software Sie zur Berechnung eingesetzt haben (gegebenenfalls inklusive Spezifizierung von Modulen, Prozeduren, Packages et cetera; siehe auch Modul 5 zur Ablage des Programmcodes).*
- *Art und Umfang von Sensitivitätsanalysen.*

Sofern Informationen zur zugrunde liegenden Methodik des indirekten Vergleichs im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

[16] 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.
[17] Song F, Xiong T, Parekh-Bhurke S, Loke YK, Sutton AJ, Eastwood AJ et al. Inconsistency between direct and indirect comparisons of competing interventions: meta-epidemiological study BMJ 2011; 343 :d4909
[18] Donegan S, Williamson P, D'Alessandro U, Tudur Smith C. Assessing key assumptions of network meta-analysis: a review of methods. Res Synth Methods 2013; 4(4): 291-323.

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 4.3.2 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)

Falls für die Bewertung des Zusatznutzens mehrere Komparatoren (zum Beispiel Wirkstoffe) herangezogen werden, sind die Aussagen zum Zusatznutzen primär gegenüber der Gesamtheit der gewählten Komparatoren durchzuführen (zum Beispiel basierend auf Meta-Analysen unter gemeinsamer Betrachtung aller direkt vergleichender Studien). Spezifische methodische Argumente, die gegen eine gemeinsame Analyse sprechen (zum Beispiel statistische oder inhaltliche Heterogenität), sind davon unbenommen. Eine zusammenfassende Aussage zum Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie ist in jedem Fall erforderlich.

Sofern Angaben zum medizinischen Nutzen und zum medizinischen Zusatznutzen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist in dem jeweiligen Abschnitt des vorliegenden Dossiers darauf zu verweisen.

Die Verweise sind dabei bis zur untersten vorhandenen Gliederungsebene und auf Abschnittsebene zu spezifizieren. Bei Verweisen auf Tabellen oder Abbildungen ist zusätzlich die jeweilige Tabellen- beziehungsweise Abbildungsnummerierung anzugeben.

Sofern Angaben zum medizinischen Nutzen und zum medizinischen Zusatznutzen teilweise oder vollständig nicht im EU-Dossier vorgelegt wurden, so sind diese Angaben in den betroffenen Abschnitten des Moduls 4 jeweils zu ergänzen.

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

Das Datum der Recherche soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen. Liegt die Recherche im EU-Dossier mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt, ist die Recherche zu aktualisieren. Die aktualisierte Recherche ist im Dossier darzustellen.

In den Fällen, in denen keine neuen Quellen identifiziert werden, kann auf das Rechercheergebnis im EU-Dossier verwiesen werden. Sofern neue Quellen identifiziert werden, sind die neuen Erkenntnisse entsprechend der Vorgaben aufzuarbeiten.

4.3.1.1.1 Studien des pharmazeutischen Unternehmers

Nachfolgend sollen alle Studien (RCT), die an die Zulassungsbehörde übermittelt wurden (Zulassungsstudien), sowie alle Studien (RCT), für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, benannt werden. Beachten Sie dabei folgende Konkretisierungen:

- *Es sollen alle RCT, die der Zulassungsbehörde im Zulassungsdossier übermittelt wurden und deren Studienberichte im Abschnitt 5.3.5 des Zulassungsdossiers enthalten sind, aufgeführt werden. Darüber hinaus sollen alle RCT, für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, aufgeführt werden.*

- *Benennen Sie in der nachfolgenden Tabelle nur solche RCT, 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“, Angabe über die Beteiligung (Sponsor ja/nein), Studienstatus (abgeschlossen, abgebrochen, laufend), Studiendauer, Angabe zu geplanten und durchgeführten Datenschnitten und Therapiearme. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile.

Tabelle 4-8: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel

Studie	**Zulassungsstudie (ja/nein)**	**Sponsor (ja/nein)**	**Status (abgeschlossen / abgebrochen / laufend)**	**Studiendauer ggf. Datenschnitt**	**Therapiearme**
ROBIN (217-PPD-201)	ja	ja	abgeschlossen	2 Wochen RCT 45 Tage verblindete Nachbeobachtung	Zuranolon + PiT PiT
SKYLARK (217-PPD-301)	ja	ja	abgeschlossen	2 Wochen RCT 45 Tage verblindete Nachbeobachtung	Zuranolon +PiT PiT

Geben Sie an, welchen Stand die Information in Tabelle 4-8 hat, das heißt. zu welchem Datum der Studienstatus abgebildet wird. Das Datum des Studienstatus soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Sofern Angaben zu Studien des pharmazeutischen Unternehmers im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Die Information hat den Stand vom 16.04.2026.

Geben Sie in der nachfolgenden Tabelle an, welche der in Tabelle 4-8 genannten Studien nicht für die Nutzenbewertung herangezogen wurden. Begründen Sie dabei jeweils die Nichtberücksichtigung. Fügen Sie für jede Studie eine neue Zeile ein.

Sofern Ausschlussgründe für Studien des pharmazeutischen Unternehmers im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-9: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	**Begründung für die Nichtberücksichtigung der Studie**
Nicht zutreffend.	

4.3.1.1.2 Studien aus der bibliografischen Recherche

Beschreiben Sie nachfolgend das Ergebnis der bibliografischen Recherche. Illustrieren Sie den Selektionsprozess und das Ergebnis der Selektion mit einem Flussdiagramm. Geben Sie dabei an, wie viele Treffer sich insgesamt (das heißt über alle durchsuchten Datenbanken) aus der bibliografischen Recherche 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 Dokumente unter Nennung des Ausschlussgrunds in Anhang 4-C.

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

Geben Sie im Flussdiagramm auch das Datum der Recherche an. Die Recherche soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

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

Sofern Angaben zu Studien aus der bibliographischen Recherche im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

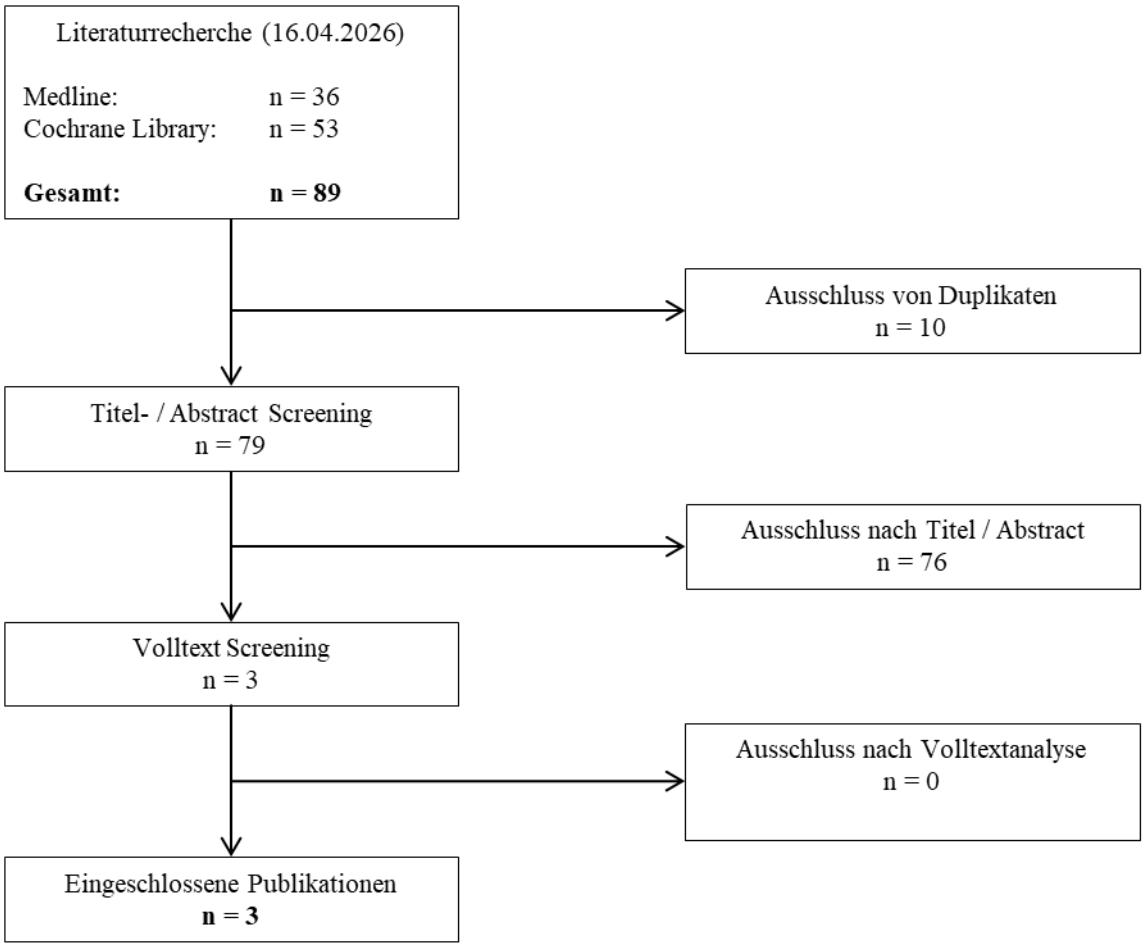


Abbildung 4-1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel

Die bibliografische Literaturrecherche ergab insgesamt 89 Treffer. Nach Ausschluss von 10 Duplikaten wurden die verbleibenden 79 Publikationen von zwei Personen unabhängig voneinander basierend auf den in Abschnitt 4.2.2 definierten Ein- und Ausschlusskriterien nach Relevanz selektiert. Nach dem Screening von Titel und Abstract wurden 76 Publikationen ausgeschlossen. Es wurden 3 Publikationen im Volltext gesichtet und als relevant eingeschlossen. Zwei Publikationen können der Studie ROBIN zugeordnet werden [59; 60] und eine der Studie SKYLARK [61].

4.3.1.1.3 Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken

Beschreiben Sie in der nachfolgenden Tabelle alle relevanten Studien, die durch die Suche in Studienregistern/Studienergebnisdatenbanken identifiziert wurden. Geben Sie dabei an, in welchem Studienregister/Studienergebnisdatenbank die Studie identifiziert wurde und welche Dokumente dort zur Studie jeweils hinterlegt sind (zum Beispiel Studienregistereintrag, Bericht über Studienergebnisse et cetera). Geben Sie auch an, ob die Studie in der Liste der Studien des pharmazeutischen Unternehmers enthalten ist (siehe Tabelle 4-8) und ob die Studie auch durch die bibliografische Recherche identifiziert wurde. Fügen Sie für jede Studie eine neue Zeile ein. Listen Sie die ausgeschlossenen Studien unter Nennung des Ausschlussgrunds in Anhang 4-D.

[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 4-10: Relevante Studien (auch laufende Studien) aus der Suche in Studienregistern/ Studienergebnisdatenbanken – RCT mit dem zu bewertenden Arzneimittel

Studie	Identifikationsorte (Name des Studienregisters/ der Studienergebnisdatenbank und Angabe der Zitate[a])	Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja/nein)	Studie durch bibliografische Literaturrecherche identifiziert (ja/nein)	Status (abgeschlossen/ abgebrochen/ laufend)
ROBIN (217-PPD-201)	Clinicaltrials.gov: NCT02978326 [62]	ja	ja	abgeschlossen
SKYLARK (217-PPD-301)	Clinicaltrials.gov: NCT04442503 [63] EU-CTR: 2020-001424-34 [64]	ja	ja	abgeschlossen
a: Zitat des Studienregistereintrags, die Studienregisternummer (NCT-Nummer, EudraCT-Nummer) sowie, falls vorhanden, der im Studienregister/in der Studienergebnisdatenbank aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.				

Geben Sie an, welchen Stand die Information in Tabelle 4-10 hat, das heißt zu welchem Datum die Recherche durchgeführt wurde. Das Datum der Recherche soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Sofern Angaben zu Studien aus der Suche in Studienregistern/Studienergebnisdatenbanken im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Die Information hat den Stand vom 16.04.2026.

4.3.1.1.4 Studien aus der Suche auf der Internetseite des Gemeinsamen Bundesausschusses

Beschreiben Sie in der nachfolgenden Tabelle alle relevanten Studien, die durch die Sichtung der Internetseite des Gemeinsamen Bundesausschusses identifiziert wurden. Geben Sie dabei an, welche Dokumente dort hinterlegt sind (zum Beispiel Dossier eines anderen pharmazeutischen Unternehmers, IQWiG Nutzenbewertung). Geben Sie auch an, ob die Studie in der Liste der Studien des pharmazeutischen Unternehmers enthalten ist (siehe Tabelle 4-8) und ob die Studie auch durch die bibliografische Recherche beziehungsweise Suche in Studienregistern/Studienergebnisdatenbank identifiziert 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 4-11: Relevante Studien aus der Suche auf der Internetseite des Gemeinsamen Bundesausschusses – RCT mit dem zu bewertenden Arzneimittel

Studie	**Relevante Quellen**[a]	**Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja/nein)**	**Studie durch bibliografische Literaturrecherche identifiziert (ja/nein)**	**Studie durch Suche in Studienregistern / Studienergebnis datenbanken identifiziert (ja/nein)**
Zu den identifizierten Studien konnten keine relevanten Quellen identifiziert werden.				
a: Quellen aus der Suche auf der Internetseite des G-BA				

Geben Sie an, welchen Stand die Information in Tabelle 4-11 hat, das heißt. zu welchem Datum die Recherche durchgeführt wurde. Das Datum der Recherche soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Sofern Angaben zu Studien aus der Suche auf der Internetseite des Gemeinsamen Bundesausschusses im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Die Information hat den Stand vom 16.04.2026.

4.3.1.1.5 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, 4.3.1.1.3 und 4.3.1.1.4) resultierenden Pool relevanter Studien (exklusive laufender Studien) für das zu bewertende Arzneimittel, auch im direkten Vergleich zur zweckmäßigen Vergleichstherapie. Alle durch die vorhergehenden Schritte identifizierten und in der Tabelle genannten Quellen der relevanten Studien sollen für die Bewertung dieser Studien herangezogen werden.

Folgende Informationen sind in der Tabelle darzulegen: Studienbezeichnung, Studienkategorie und verfügbare Quellen. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Hierbei sollen die Studien durch Zwischenzeilenüberschriften gegebenenfalls 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.

Sofern Angaben zum resultierenden Studienpool aus RCT mit dem zu bewertenden Arzneimittel im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-12: Studienpool – RCT mit dem zu bewertenden Arzneimittel

Studie	Studienkategorie			verfügbare Quellen[a]		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	gesponserte Studie[b] (ja/nein)	Studie Dritter (ja/nein)	Studienberichte (ja/nein [Zitat])	Register-einträge[c] (ja/nein [Zitat])	Publikation und sonstige Quellen[d] (ja/nein [Zitat])
ggf. Zwischenüberschrift zur Strukturierung des Studienpools						
kontrolliert, zweckmäßige Vergleichstherapie(n)						
ROBIN (217-PPD-201)	ja	ja	nein	ja [15]	ja [62]	ja [59; 60]
SKYLARK (217-PPD-301)	ja	ja	nein	ja [14]	ja [63; 64]	ja [61]

a: Bei Angabe „ja“ sind jeweils die Zitate der Quelle(n) (z. B. Publikationen, Studienberichte, Studienregister-einträge) mit anzugeben, und zwar als Verweis auf die in Abschnitt 4.6 genannte Referenzliste.
b: Studie, für die der Unternehmer Sponsor war.
c: Zitat der Studienregistereinträge.
d: Sonstige Quellen: Dokumente aus der Suche auf der Internetseite des Gemeinsamen Bundesausschusses.

In den Populationen der Studien ROBIN und SKYLARK befinden sich nur Patientinnen mit **mittelschwerer und schwerer PPD**. Daher sind die identifizierten Studien für die Nutzenbewertung von Zuranolon in der Patientenpopulation b) geeignet.

Für die Patientenpopulation a) Erwachsene Patientinnen nach der Geburt mit einer PPD mit **leichter** depressiver Symptomatik konnte keine Studie mit Zuranolon identifiziert werden.

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. Falls Teilpopulationen berücksichtigt werden, ist die Charakterisierung der Studienpopulation auch für diese Teilpopulation durchzuführen. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile. Geben Sie bei den Datenschnitten auch den Anlass des Datenschnittes an. Fügen Sie für jede Studie eine neue Zeile ein.

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

Sofern Informationen zum Studiendesign und zur Studienpopulation im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-13: 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/ Datenschnitte** <ggf. Run-in, Behandlung, Nachbeobachtung>	**Ort und Zeitraum der Durchführung**	**Primärer Endpunkt; patientenrelevante sekundäre Endpunkte**
ROBIN (217-PPD-201)	RCT, doppelblind, parallel	Patientinnen mit einer PPD mit mittelschwerer bis schwerer depressiver Symptomatik	Zuranolon + PiT (n = 77) PiT (n = 76)	Screening: ≤ 28 Tage Behandlung: 14 Tage Follow-Up: bis Tag 45 Datenschnitt: 13.02.2019	USA 01/2017 – 12/2018	**Primärer Endpunkt:** Veränderung des HAMD-Gesamtwerts vom Baselinewert bis Tag 15. **Weitere Endpunkte** - HAMD - MADRS - EPDS - PHQ-9 - HAM-A - SF-36 - BIMF - Sicherheit
SKYLARK (217-PPD-301)	RCT, doppelblind, parallel	Patientinnen mit einer PPD mit mittelschwerer bis schwerer depressiver Symptomatik	Zuranolon + PiT (n = 99) PiT (n = 101)	Screening: ≤ 28 Tage Behandlung: 14 Tage Follow-Up: bis Tag 45 Datenschnitt: 16.05.2022	USA, Spanien, UK. 06/2020 – 04/2022	**Primärer Endpunkt:** Veränderung des HAMD-Gesamtwerts vom Baselinewert bis Tag 15. **Weitere Endpunkte** - HAMD - MADRS - EPDS - PHQ-9

Studie	**Studiendesign** <RCT, doppelblind/einfach, verblindet/offen, parallel/cross-over etc.>	**Population** <relevante Charakteristika, z. B. Schweregrad>	**Interventionen (Zahl der randomisierten Patienten)**	**Studiendauer/ Datenschnitte** <ggf. Run-in, Behandlung, Nachbeobachtung>	**Ort und Zeitraum der Durchführung**	**Primärer Endpunkt; patientenrelevante sekundäre Endpunkte**
						- HAM-A - Sicherheit
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.						

Tabelle 4-14: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel

Studie	**Zuranolon + PiT**	**PiT**	***ggf. weitere Spalten mit Behandlungscharakteristika*** ***z. B. Vorbehandlung, Behandlung in der Run-in-Phase etc.***
ROBIN (217-PPD-201)	Zuranolon; 30 mg, abends, oral in Form von Kapseln PiT: allen nach Protokoll nicht ausgeschlossenen Therapieoptionen	PiT, bestehend aus einem Placebo abends, oral in Form von Kapseln, um die Zuranolon-Gabe zu verblinden sowie allen nach Protokoll nicht ausgeschlossenen Therapieoptionen	
SKYLARK (217-PPD-301)	Zuranolon; 50 mg oder 40 mg (Dosisreduktion nur bei Unverträglichkeit), abends, oral in Form von Kapseln PiT: allen nach Protokoll nicht ausgeschlossenen Therapieoptionen	PiT, bestehend aus einem Placebo abends, oral in Form von Kapseln, um die Zuranolon-Gabe zu verblinden sowie allen nach Protokoll nicht ausgeschlossenen Therapieoptionen	
Hinweis: Die in der ROBIN verabreichte Dosis von 30 mg Zuranolon hat durch eine veränderte Kapselformulierung eine höhere Bioverfügbarkeit Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.			

Tabelle 4-15: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel

Variable	Statistische Größe oder Kategorie	SKYLARK			ROBIN			POOLED		
		Zuranolon + PiT (N=99)	PiT (N=101)	Gesamt (N=200)	Zuranolon + PiT (N=77)	PiT (N=76)	Gesamt (N=153)	Zuranolon + PiT (N=176)	PiT (N=177)	Gesamt (N=353)
Demografische Charakteristika										
Alter zum Zeitpunkt der Einwilligung (Jahre)	n	99	101	200	77	76	153	176	177	353
	Mittelwert (SD)	29,9 (5,88)	30,9 (5,93)	30,4 (5,91)	29,2 (5,39)	27,3 (5,29)	28,3 (5,41)	29,6 (5,67)	29,4 (5,92)	29,5 (5,79)
	Median	30,0	31,0	31,0	29,0	27,5	28,0	29,0	29,0	29,0
	Q1; Q3	25,0; 34,0	26,0; 36,0	26,0; 35,0	25,0; 33,0	23,0; 31,5	24,0; 33,0	25,0; 33,0	25,0; 34,0	25,0; 33,0
	Min; Max	19; 44	19; 43	19; 44	18; 44	18; 38	18; 44	18; 44	18; 43	18; 44
Alter zum Zeitpunkt der Einwilligung (Jahre) - Kategorie; n (%)	n [1]	99	101	200	77	76	153	176	177	353
	18 – 24 Jahre	19 (19,2)	13 (12,9)	32 (16,0)	18 (23,4)	26 (34,2)	44 (28,8)	37 (21,0)	39 (22,0)	76 (21,5)
	25 – 45 Jahre	80 (80,8)	88 (87,1)	168 (84,0)	59 (76,6)	50 (65,8)	109 (71,2)	139 (79,0)	138 (78,0)	277 (78,5)
Ethnizität; n (%)	n [1]	99	101	200	77	76	153	176	177	353
	Nicht hispanisch oder lateinamerikanisch	64 (64,6)	58 (57,4)	122 (61,0)	60 (77,9)	58 (76,3)	118 (77,1)	124 (70,5)	116 (65,5)	240 (68,0)

Variable	Statistische Größe oder Kategorie	SKYLARK			ROBIN			POOLED		
		Zuranolon + PiT (N=99)	PiT (N=101)	Gesamt (N=200)	Zuranolon + PiT (N=77)	PiT (N=76)	Gesamt (N=153)	Zuranolon + PiT (N=176)	PiT (N=177)	Gesamt (N=353)
	Hispanisch oder lateinamerikanisch	33 (33,3)	43 (42,6)	76 (38,0)	17 (22,1)	18 (23,7)	35 (22,9)	50 (28,4)	61 (34,5)	111 (31,4)
	Nicht angegeben	2 (2,0)	0	2 (1,0)	N/A	N/A	N/A	2 (1,1)	0	2 (0,6)
Abstammung; n (%)	n [1]	99	101	200	77	76	153	176	177	353
	Weiß	68 (68,7)	70 (69,3)	138 (69,0)	45 (58,4)	40 (52,6)	85 (55,6)	113 (64,2)	110 (62,1)	223 (63,2)
	Schwarz oder Afroamerikanisch	26 (26,3)	20 (19,8)	46 (23,0)	31 (40,3)	33 (43,4)	64 (41,8)	57 (32,4)	53 (29,9)	110 (31,2)
	Mehr als eine Ethnie	3 (3,0)	2 (2,0)	5 (2,5)	0	1 (1,3)	1 (0,7)	3 (1,7)	3 (1,7)	6 (1,7)
	Andere	0	4 (4,0)	4 (2,0)	0	0	0	0	4 (2,3)	4 (1,1)
	Amerikanischer Ureinwohner oder Ureinwohner Alaskas	0	3 (3,0)	3 (1,5)	0	0	0	0	3 (1,7)	3 (0,8)
	Asiatisch	1 (1,0)	1 (1,0)	2 (1,0)	1 (1,3)	1 (1,3)	2 (1,3)	2 (1,1)	2 (1,1)	4 (1,1)
	Hawaiianischer Ureinwohner oder anderer Pazifikinsulaner	0	0	0	0	1 (1,3)	1 (0,7)	0	1 (0,6)	1 (0,3)

Variable	Statistische Größe oder Kategorie	SKYLARK			ROBIN			POOLED		
		Zuranolon + PiT (N=99)	PiT (N=101)	Gesamt (N=200)	Zuranolon + PiT (N=77)	PiT (N=76)	Gesamt (N=153)	Zuranolon + PiT (N=176)	PiT (N=177)	Gesamt (N=353)
	Nicht angegeben	1 (1,0)	1 (1,0)	2 (1,0)	0	0	0	1 (0,6)	1 (0,6)	2 (0,6)
Abstammung - Kategorie; n (%)	n [1]	99	101	200	77	76	153	176	177	353
	Schwarz oder Afroamerikanisch	26 (26,3)	20 (19,8)	46 (23,0)	31 (40,3)	33 (43,4)	64 (41,8)	57 (32,4)	53 (29,9)	110 (31,2)
	Weiß	68 (68,7)	70 (69,3)	138 (69,0)	45 (58,4)	40 (52,6)	85 (55,6)	113 (64,2)	110 (62,1)	223 (63,2)
	Andere	5 (5,1)	11 (10,9)	16 (8,0)	1 (1,3)	3 (3,9)	4 (2,6)	6 (3,4)	14 (7,9)	20 (5,7)
Bildungsniveau; n (%)	n [1]	96	99	195	N/A	N/A	N/A	96	99	195
	≤ 12. Klasse; kein Abschluss	4 (4,2)	4 (4,0)	8 (4,1)	N/A	N/A	N/A	4 (4,2)	4 (4,0)	8 (4,1)
	Abschluss der 12. Klasse oder GED	31 (32,3)	33 (33,3)	64 (32,8)	N/A	N/A	N/A	31 (32,3)	33 (33,3)	64 (32,8)
	Einige College-Kurse; aber kein Abschluss	22 (22,9)	25 (25,3)	47 (24,1)	N/A	N/A	N/A	22 (22,9)	25 (25,3)	47 (24,1)
	Beruflicher Associate-Abschluss	2 (2,1)	3 (3,0)	5 (2,6)	N/A	N/A	N/A	2 (2,1)	3 (3,0)	5 (2,6)
	Akademischer Associate-Abschluss	6 (6,3)	6 (6,1)	12 (6,2)	N/A	N/A	N/A	6 (6,3)	6 (6,1)	12 (6,2)

Variable	Statistische Größe oder Kategorie	SKYLARK			ROBIN			POOLED		
		Zuranolon + PiT (N=99)	PiT (N=101)	Gesamt (N=200)	Zuranolon + PiT (N=77)	PiT (N=76)	Gesamt (N=153)	Zuranolon + PiT (N=176)	PiT (N=177)	Gesamt (N=353)
	Bachelor-Abschluss	24 (25,0)	18 (18,2)	42 (21,5)	N/A	N/A	N/A	24 (25,0)	18 (18,2)	42 (21,5)
	Master-Abschluss	5 (5,2)	9 (9,1)	14 (7,2)	N/A	N/A	N/A	5 (5,2)	9 (9,1)	14 (7,2)
	Beruflicher Abschluss	2 (2,1)	1 (1,0)	3 (1,5)	N/A	N/A	N/A	2 (2,1)	1 (1,0)	3 (1,5)
	Doktorgrad	0	0	0	N/A	N/A	N/A	0	0	0
Familien-/Ehestand; n (%)	n [1]	96	99	195	N/A	N/A	N/A	96	99	195
	Verheiratet	37 (38,5)	51 (51,5)	88 (45,1)	N/A	N/A	N/A	37 (38,5)	51 (51,5)	88 (45,1)
	Nie verheiratet	40 (41,7)	39 (39,4)	79 (40,5)	N/A	N/A	N/A	40 (41,7)	39 (39,4)	79 (40,5)
	Geschieden	7 (7,3)	3 (3,0)	10 (5,1)	N/A	N/A	N/A	7 (7,3)	3 (3,0)	10 (5,1)
	Lebenspartner	11 (11,5)	6 (6,1)	17 (8,7)	N/A	N/A	N/A	11 (11,5)	6 (6,1)	17 (8,7)
	Ledig	1 (1,0)	0	1 (0,5)	N/A	N/A	N/A	1 (1,0)	0	1 (0,5)
Beschäftigungsstatus ; n (%)	n [1]	96	99	195	N/A	N/A	N/A	96	99	195
	Vollzeit (≥ 35 h / Woche)	33 (34,4)	25 (25,3)	58 (29,7)	N/A	N/A	N/A	33 (34,4)	25 (25,3)	58 (29,7)
	Teilzeit (< 35 h / Woche)	7 (7,3)	12 (12,1)	19 (9,7)	N/A	N/A	N/A	7 (7,3)	12 (12,1)	19 (9,7)

Variable	Statistische Größe oder Kategorie	SKYLARK			ROBIN			POOLED		
		Zuranolon + PiT (N=99)	PiT (N=101)	Gesamt (N=200)	Zuranolon + PiT (N=77)	PiT (N=76)	Gesamt (N=153)	Zuranolon + PiT (N=176)	PiT (N=177)	Gesamt (N=353)
	Im Ruhestand	1 (1,0)	0	1 (0,5)	N/A	N/A	N/A	1 (1,0)	0	1 (0,5)
	Arbeitslos	44 (45,8)	54 (54,5)	98 (50,3)	N/A	N/A	N/A	44 (45,8)	54 (54,5)	98 (50,3)
	Andere	11 (11,5)	8 (8,1)	19 (9,7)	N/A	N/A	N/A	11 (11,5)	8 (8,1)	19 (9,7)
Land; n (%)	n [1]	99	101	200	77	76	153	176	177	353
	USA	96 (97,0)	99 (98,0)	195 (97,5)	77 (100)	76 (100)	153 (100)	173 (98,3)	175 (98,9)	348 (98,6)
	Rest der Welt	3 (3,0)	2 (2,0)	5 (2,5)	0	0	0	3 (1,7)	2 (1,1)	5 (1,4)
Körpergröße (cm)	n	97	98	195	76	75	151	173	173	346
	Mittelwert (SD)	163,19 (7,679)	162,87 (6,482)	163,03 (7,086)	164,98 (7,429)	162,44 (7,130)	163,72 (7,369)	163,98 (7,601)	162,68 (6,754)	163,33 (7,209)
	Median	163,30	162,85	163,00	164,85	162,56	163,00	164,50	162,60	163,00
	Q1; Q3	157,50; 167,60	158,00; 167,50	157,50; 167,60	160,02; 170,00	157,50; 167,60	159,00; 169,00	158,50; 169,00	158,00; 167,50	158,50; 167,64
	Min; Max	144,8; 182,9	149,9; 181,6	144,8; 182,9	147,3; 185,4	147,3; 181,0	147,3; 185,4	144,8; 185,4	147,3; 181,6	144,8; 185,4
Gewicht (kg)	n	98	98	196	76	75	151	174	173	347

Variable	Statistische Größe oder Kategorie	SKYLARK			ROBIN			POOLED		
		Zuranolon + PiT (N=99)	PiT (N=101)	Gesamt (N=200)	Zuranolon + PiT (N=77)	PiT (N=76)	Gesamt (N=153)	Zuranolon + PiT (N=176)	PiT (N=177)	Gesamt (N=353)
	Mittelwert (SD)	82,29 (18,701)	78,97 (19,493)	80,63 (19,125)	85,06 (19,104)	79,94 (23,546)	82,52 (21,508)	83,50 (18,874)	79,39 (21,285)	81,45 (20,188)
	Median	80,05	74,45	78,70	85,95	77,40	81,10	81,60	76,50	79,60
	Q1; Q3	68,10; 96,20	64,90; 91,70	66,75; 94,10	70,35; 97,18	62,30; 96,00	64,86; 97,07	70,00; 96,40	64,20; 93,00	66,30; 95,00
	Min; Max	47,7; 142,9	48,6; 137,0	47,7; 142,9	49,5; 144,2	44,5; 172,4	44,5; 172,4	47,7; 144,2	44,5; 172,4	44,5; 172,4
BMI (kg/m²)	n	97	98	195	76	75	151	173	173	346
	Mittelwert (SD)	30,86 (6,331)	29,61 (6,261)	30,23 (6,311)	31,14 (6,218)	30,17 (8,087)	30,66 (7,199)	30,99 (6,265)	29,85 (7,094)	30,42 (6,707)
	Median	31,07	28,76	29,69	30,65	28,50	30,00	30,82	28,70	29,81
	Q1; Q3	26,49; 34,81	25,38; 33,46	25,49; 34,58	26,45; 34,60	23,40; 36,30	25,40; 35,10	26,49; 34,71	25,00; 34,47	25,48; 34,70
	Min; Max	19,0; 44,8	18,7; 44,8	18,7; 44,8	19,6; 47,2	16,5; 56,1	16,5; 56,1	19,0; 47,2	16,5; 56,1	16,5; 56,1
BMI (kg/m²) - Kategorie; n (%)	n [1]	97	98	195	76	75	151	173	173	346
	≤18,4	0	0	0	0	2 (2,7)	2 (1,3)	0	2 (1,2)	2 (0,6)

Variable	Statistische Größe oder Kategorie	SKYLARK			ROBIN			POOLED		
		Zuranolon + PiT (N=99)	PiT (N=101)	Gesamt (N=200)	Zuranolon + PiT (N=77)	PiT (N=76)	Gesamt (N=153)	Zuranolon + PiT (N=176)	PiT (N=177)	Gesamt (N=353)
	18,5 – 24,9	20 (20,6)	22 (22,4)	42 (21,5)	13 (17,1)	19 (25,3)	32 (21,2)	33 (19,1)	41 (23,7)	74 (21,4)
	25 – 29,9	24 (24,7)	34 (34,7)	58 (29,7)	21 (27,6)	20 (26,7)	41 (27,2)	45 (26,0)	54 (31,2)	99 (28,6)
	≥ 30	53 (54,6)	42 (42,9)	95 (48,7)	42 (55,3)	34 (45,3)	76 (50,3)	95 (54,9)	76 (43,9)	171 (49,4)
BMI (kg/m^2) - Kategorie 2; n (%)	n [1]	97	98	195	76	75	151	173	173	346
	≤ 24,9	20 (20,6)	22 (22,4)	42 (21,5)	13 (17,1)	21 (28,0)	34 (22,5)	33 (19,1)	43 (24,9)	76 (22,0)
	25 – 29,9	24 (24,7)	34 (34,7)	58 (29,7)	21 (27,6)	20 (26,7)	41 (27,2)	45 (26,0)	54 (31,2)	99 (28,6)
	≥ 30	53 (54,6)	42 (42,9)	95 (48,7)	42 (55,3)	34 (45,3)	76 (50,3)	95 (54,9)	76 (43,9)	171 (49,4)
Krankheitsspezifische Charakteristika										
Antidepressiva-Nutzung zu Baseline; n (%)	n [1]	99	101	200	77	76	153	176	177	353
	Ja	15 (15,2)	15 (14,9)	30 (15,0)	16 (20,8)	13 (17,1)	29 (19,0)	31 (17,6)	28 (15,8)	59 (16,7)
	Nein	84 (84,8)	86 (85,1)	170 (85,0)	61 (79,2)	63 (82,9)	124 (81,0)	145 (82,4)	149 (84,2)	294 (83,3)
Ausgangs-HAMD-17-Gesamtpunktzahl - Kategorie; n (%)	n [1]	98	98	196	76	75	151	174	173	347
	< 28 (Median des HAMD-17-Gesamtwertes zu	33 (33,7)	32 (32,7)	65 (33,2)	30 (39,5)	22 (29,3)	52 (34,4)	63 (36,2)	54 (31,2)	117 (33,7)

Variable	Statistische Größe oder Kategorie	SKYLARK			ROBIN			POOLED		
		Zuranolon + PiT (N=99)	PiT (N=101)	Gesamt (N=200)	Zuranolon + PiT (N=77)	PiT (N=76)	Gesamt (N=153)	Zuranolon + PiT (N=176)	PiT (N=177)	Gesamt (N=353)
	Baseline)									
	≥ 28 (Median des HAMD-17-Gesamtwertes zu Baseline)	65 (66,3)	66 (67,3)	131 (66,8)	46 (60,5)	53 (70,7)	99 (65,6)	111 (63,8)	119 (68,8)	230 (66,3)
Beginn der PPD; n (%)	n [1]	99	101	200	77	76	153	176	177	353
	3. Trimester	35 (35,4)	32 (31,7)	67 (33,5)	32 (41,6)	32 (42,1)	64 (41,8)	67 (38,1)	64 (36,2)	131 (37,1)
	Nach der Geburt	64 (64,6)	69 (68,3)	133 (66,5)	45 (58,4)	44 (57,9)	89 (58,2)	109 (61,9)	113 (63,8)	222 (62,9)
PPD in der Anamnese; n (%)	n [1]	99	101	200	77	76	153	176	177	353
	Erste Episode	82 (82,8)	90 (89,1)	172 (86,0)	67 (87,0)	60 (78,9)	127 (83,0)	149 (84,7)	150 (84,7)	299 (84,7)
	Wiederkehrende Episode	17 (17,2)	11 (10,9)	28 (14,0)	10 (13,0)	16 (21,1)	26 (17,0)	27 (15,3)	27 (15,3)	54 (15,3)
PPD in der Familienanamnese; n (%)	n [1]	99	101	200	77	76	153	176	177	353
	Ja	21 (21,2)	16 (15,8)	37 (18,5)	10 (13,0)	10 (13,2)	20 (13,1)	31 (17,6)	26 (14,7)	57 (16,1)
	Nein/Unbekannt	78 (78,8)	85 (84,2)	163 (81,5)	67 (87,0)	66 (86,8)	133 (86,9)	145 (82,4)	151 (85,3)	296 (83,9)

Variable	Statistische Größe oder Kategorie	SKYLARK			ROBIN			POOLED		
		Zuranolon + PiT (N=99)	PiT (N=101)	Gesamt (N=200)	Zuranolon + PiT (N=77)	PiT (N=76)	Gesamt (N=153)	Zuranolon + PiT (N=176)	PiT (N=177)	Gesamt (N=353)
COVID-19-in der Anamnese; n (%) [2]	n [1]	99	101	200	N/A	N/A	N/A	99	101	200
	COVID-19	1 (1,0)	0	1 (0,5)	N/A	N/A	N/A	1 (1,0)	0	1 (0,5)
	Vermutetes COVID-19	0	0	0	N/A	N/A	N/A	0	0	0
	Patientenisolation	0	0	0	N/A	N/A	N/A	0	0	0
	SARS-CoV-2-Test positiv	3 (3,0)	6 (5,9)	9 (4,5)	N/A	N/A	N/A	3 (3,0)	6 (5,9)	9 (4,5)
	Krankenhausaufenthalt	0	0	0	N/A	N/A	N/A	0	0	0
	Nicht betroffen	95 (96,0)	95 (94,1)	190 (95,0)	N/A	N/A	N/A	95 (96,0)	95 (94,1)	190 (95,0)
Depression mit erhöhter Angst; basierend auf der HAMD-17-Angst-Subskala; n (%)	n [1]	98	98	196	76	75	151	174	173	347
	≥ 7 (Normalisiert: ≥ 39)	94 (95,9)	96 (98,0)	190 (96,9)	74 (97,4)	72 (96,0)	146 (96,7)	168 (96,6)	168 (97,1)	336 (96,8)
	<7 (Normalisiert: < 39)	4 (4,1)	2 (2,0)	6 (3,1)	2 (2,6)	3 (4,0)	5 (3,3)	6 (3,4)	5 (2,9)	11 (3,2)
Depression mit	n [1]	98	98	196	76	75	151	174	173	347

Variable	Statistische Größe oder Kategorie	SKYLARK			ROBIN			POOLED		
		Zuranolon + PiT (N=99)	PiT (N=101)	Gesamt (N=200)	Zuranolon + PiT (N=77)	PiT (N=76)	Gesamt (N=153)	Zuranolon + PiT (N=176)	PiT (N=177)	Gesamt (N=353)
erhöhter Angst; basierend auf der HAM-A-Gesamtpunktzahl; n (%)	≥ 20	74 (75,5)	77 (78,6)	151 (77,0)	66 (86,8)	67 (89,3)	133 (88,1)	140 (80,5)	144 (83,2)	284 (81,8)
	<20	24 (24,5)	21 (21,4)	45 (23,0)	10 (13,2)	8 (10,7)	18 (11,9)	34 (19,5)	29 (16,8)	63 (18,2)
Einnahme von Antidepressiva jemals; n (%)	n [1]	99	101	200	77	76	153	176	177	353
	Ja	36 (36,4)	33 (32,7)	69 (34,5)	24 (31,2)	20 (26,3)	44 (28,8)	60 (34,1)	53 (29,9)	113 (32,0)
	Nein	63 (63,6)	68 (67,3)	131 (65,5)	53 (68,8)	56 (73,7)	109 (71,2)	116 (65,9)	124 (70,1)	240 (68,0)
Begleitende Verfahren - Psychotherapie; n (%)	n [1]	99	101	200	N/A	N/A	N/A	99	101	200
	Ja	2 (2,0)	1 (1,0)	3 (1,5)	N/A	N/A	N/A	2 (2,0)	1 (1,0)	3 (1,5)
	Nein	97 (98,0)	100 (99,0)	197 (98,5)	N/A	N/A	N/A	97 (98,0)	100 (99,0)	197 (98,5)

1: Wird als Nenner des Prozentsatzes für den jeweiligen Abschnitt verwendet.
2: Bildungsniveau, Familien/-Ehestand, Beschäftigungsstatus, COVID-19 in der Anamnese und Psychotherapie wurden nur in der Studie SKYLARK erhoben.

Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.

Beschreiben Sie die Studien zusammenfassend. In der Beschreibung der Studien sollten Informationen zur Behandlungsdauer sowie zu geplanter und tatsächlicher Beobachtungsdauer enthalten sein. Sofern sich die Beobachtungsdauer zwischen den relevanten Endpunkten unterscheidet, sind diese unterschiedlichen Beobachtungsdauern endpunktbezogen anzugeben.

Machen Sie Angaben zu verabreichten Folgetherapien nach Abbruch der Studienmedikation (bei onkologischen Fragestellungen zusätzlich auch separate Angaben zur ersten Folgetherapie).

Beschreiben Sie zudem, ob und aus welchem Anlass verschiedene Datenschnitte durchgeführt wurden oder noch geplant sind. Geben Sie dabei auch an, ob diese Datenschnitte jeweils vorab (das heißt im statistischen Analyseplan) geplant waren. In der Regel ist nur die Darstellung von a priori geplanten oder von Zulassungsbehörden geforderten Datenschnitten erforderlich. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

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

Sofern Informationen zu den eingeschlossenen Studien im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Studiendesign

Studiendesign – ROBIN

Die ROBIN-Studie wurde als multizentrische, randomisierte, doppelblinde, kontrollierte Parallelgruppenstudie konzipiert. Das Zuteilungsverhältnis der Randomisierung betrug 1 : 1, wobei die Teilnehmerinnen entweder Zuranolon+PiT oder Placebo+PiT erhielten.

Die Studie bestand aus einer Screening-Periode (vor Amendment 4: von bis zu 14 Tagen; nach Amendment 4: von bis zu 28 Tagen), einer 14-tägigen doppelblinden Behandlungsperiode und einer Follow-Up-Periode bis Tag 45. Während der Screening-Periode wurde festgestellt ob die Teilnehmerinnen gemäß DSM-5 SCIDI eine PPD aufweisen. Außerdem wurden die Teilnehmerinnen hinsichtlich Ein- und Ausschlusskriterien überprüft.

Wenn die Teilnehmerinnen die Kriterien der Screening-Periode erfüllten, konnten sie in die Behandlungsperiode übergehen und wurden 1 : 1 in den Zuranolon + PiT-Arm oder PiT-Arm randomisiert. Die orale Einnahme von 30 mg Zuranolon oder Placebo erfolgte einmal täglich über einen Zeitraum von 14 Tagen. In diesem Zeitraum wurden Wirksamkeits- und Sicherheitsendpunkte regelmäßig gemäß dem Zeitplan erhoben.

Während der Follow-Up Periode, die bis Tag 45 nach Beginn der Verabreichung des Studienmedikaments andauerte, wurden die Teilnehmerinnen weiterhin beobachtet.

Studiendesign – SKYLARK

Die SKYLARK-Studie wurde als randomisierte, doppel-blinde, kontrollierte, parallel-gruppierte Phase 3 Studie konzipiert, um die Wirksamkeit und Sicherheit von Zuranolon bei Erwachsenen mit schwergradiger PPD zu evaluieren.

Die Studie bestand aus einer Screening-Periode von bis zu 28 Tagen, einer 14-tägigen doppelblinden Behandlungsperiode und einer Follow-Up-Periode bis Tag 45.

Während der Screening-Periode wurde die Diagnose Depression mittels dem DSM-5 SCID-5-CT gestellt und die Teilnehmerinnen wurden auf die entsprechenden Ein- und Ausschlusskriterien überprüft. Anschließend wurden die Patientinnen in einem 1 : 1 Verhältnis in die Zuranolon + PiT-Gruppe oder PiT-Gruppe randomisiert.

Während der Behandlungsperiode nahmen die Patientinnen für 14 Tage jeden Abend ihre Studienmedikation ein, solange keine dosislimitierenden Bedenken bezüglich der Sicherheit oder Verträglichkeit bestanden. Teilnehmerinnen, die 50 mg nicht vertrugen, erhielten 40 mg für den Rest der Behandlungsperiode.

Während der Follow-Up Periode, die bis Tag 45 nach Beginn der Verabreichung des Studienmedikaments andauerte, wurden die Teilnehmerinnen weiterhin beobachtet.

Datenschnitte

Sowohl für die ROBIN-Studie als auch für die SKYLARK-Studie gab es nur einen Datenschnitt und keine Interimsanalysen. Der einzige und finale Datenschnitt für die ROBIN Analyse ist datiert auf den 13.02.2019, während der einzige und finale Datenschnitt der SKYLARK am 16.05.2022 stattfand.

Charakterisierung der Studienpopulationen

Eine detaillierte Charakterisierung der einzelnen Studienpopulationen sowie der gepoolten Population lässt sich der Tabelle 4-15 entnehmen.

ROBIN

Innerhalb der ROBIN-Studie haben die Patientinnen in den beiden Behandlungsarmen vergleichbare demografische Charakteristika. Die krankheitsspezifischen Charakteristika der Patientinnen sind ebenso überwiegend vergleichbar zwischen den beiden Behandlungsarmen. Lediglich bei der HAMD-17 und einer PPD in der Anamnese scheinen die Patientinnen im PiT-Arm mit einem höheren Anteil von Patientinnen mit HAMD-17 ≥ 28 und wiederkehrender PPD zu Baseline etwas stärker betroffen zu sein als die im Zuranolon + PiT-Arm.

SKYLARK

Die Patientinnen der SKYLARK-Studie weisen zu Baseline vergleichbare demografische Kennwerte zwischen den beiden Behandlungsarmen auf. In Hinblick auf die krankheitsspezifischen Charakteristika sind die Patientinnen der beiden Behandlungsarme überwiegend vergleichbar. Ein kleiner Unterschied zeigt sich für den Anteil mit PPD in der Anamnese, sowie einer PPD in der Familienanamnese, da hier ein etwas größerer Anteil der

Patientinnen im Zuranolon + PiT-Arm bereits in der Vergangenheit unter einer PPD-Episode litt und zusätzlich häufiger eine PPD im Familienumkreis vorlag als im PiT-Arm.

Meta-Analyse

Die Studien ROBIN und SKYLARK sind insgesamt hinreichend vergleichbar, um gemeinsam im Rahmen einer Meta-Analyse ausgewertet zu werden. Hierfür wurden die relevanten Studien hinsichtlich ihrer Vergleichbarkeit in Bezug auf Fragestellung, Studiendesign, Studienpopulation, Kontrollarm und Endpunkte untersucht. Die Studien ROBIN und SKYLARK weisen ähnliche Protokolle und Einschlusskriterien auf und vergleichen die Sicherheit und Wirksamkeit von Zuranolon in Kombination mit einer patientenindividuellen Therapie (PiT) gegenüber einem PiT-Vergleichsarm. Die Effektschätzer beider Studien zeigen in dieselbe Richtung, und die Heterogenitätstests waren unauffällig, was die Eignung für eine gemeinsame Effektschätzung unterstützt.

Die Ein- und Ausschlusskriterien der beiden Studien waren sehr ähnlich, wobei die Ausschlusskriterien der SKYLARK etwas detaillierter sind, da sie die Folgestudie ist. Dementsprechend sind die Merkmale der Patientinnen wie zu erwarten zu Baseline sehr ähnlich. Die Unterschiede in der Abstammung, mit einem höheren Anteil an schwarzen oder afroamerikanischen Personen in der ROBIN-Studie ist in der Lage der Studienzentren begründet. Die ROBIN-Studie wurde ausschließlich in US-amerikanischen Studienzentren durchgeführt, während die SKYLARK-Studie über US-Studienzentren hinaus auch an Zentren in UK und Spanien durchgeführt wurde. Während der Anteil der schwarzen und afroamerikanischen Person in der ROBIN-Studie höher ist, sind in der SKYLARK-Studie mehr Patientinnen mit einer hispanischen oder lateinamerikanischen Ethnizität eingeschlossen worden. Darüber hinaus gleichen sich die bereits beschriebenen leichten Unterschiede bezüglich einer PPD in der Anamnese in den jeweiligen Behandlungsarmen durch das Zusammenführen der Studiendaten aus.

Hinsichtlich der PPD-Erkrankung lässt sich die gepoolte Population wie folgt beschreiben: Zu Baseline nahmen 17 % der Frauen Antidepressiva ein, während insgesamt 33 % im Verlauf ihres Lebens bereits eine Behandlung mit diesen Medikamenten erfahren hatten. Dementsprechend ist davon auszugehen, dass ein Drittel der Patientinnen bereits in der Vergangenheit an einer psychischen Erkrankung bzw. Depression erkrankt war. Der Beginn der PPD variierte innerhalb der Population: Bei grob einem Drittel der Patientinnen traten die Symptome bereits im 3. Trimester der Schwangerschaft auf, während die Mehrheit der Patientinnen erst nach der Geburt erkrankte. Die Mehrheit der Frauen (85 %) litt erstmalig an einer PPD und 84 % wiesen keine familiäre PPD auf. Zudem zeigte sich eine hohe Prävalenz von Angststörungen, die bei 97 % der Frauen über den HAMD-17 und bei 82 % über den HAM-A dokumentiert wurde.

Auch hinsichtlich der Studienmedikation bestehen keine methodischen Einschränkungen für die Durchführung einer Metaanalyse. Die in der ROBIN verabreichte Dosis von 30 mg Zuranolon+PiT weist durch eine veränderte Kapselformulierung eine höhere Bioverfügbarkeit auf. Insgesamt zeigt sich, dass ROBIN und SKYLARK sowohl im Sinne der Charakteristika

der Studienpopulationen als auch methodisch hinreichend vergleichbar sind, um sie in einer Meta-Analyse gemeinsam auszuwerten.

Schweregrad der Erkrankung

In beiden Studien wurden Patientinnen nur dann eingeschlossen, wenn sie einen HAMD-17-Gesamtwert von ≥ 26 Punkten aufweisen konnten, der auf dieser Skala einer schweren Depression entspricht. Die HAMD-17 ist ein etabliertes Instrument zur Fremdbeurteilung depressiver Symptome, das somatische Aspekte wie Schlafstörungen stärker gewichtet. Auf Nachfrage der European Medicines Agency (EMA) wurden zusätzlich Auswertungen nach dem MADRS vorgelegt, der stärker auf emotionale Kernsymptome fokussiert und auch zur Einteilung der Schwere der Erkrankung genutzt werden kann [65]. Die Auswertung nach MADRS zeigte, dass Patientinnen mit mittelschweren und schweren Symptomen eingeschlossen wurden (siehe Tabelle 4-16 und Abbildung 4-2). Somit können Patientinnen nach HAMD-17 als „schwer depressiv" eingestuft werden, während sie nach MADRS im Bereich „mittelschwer" eingeordnet werden. Diese Diskrepanz ist in der klinischen Praxis bekannt und spiegelt die unterschiedlichen Schwerpunkte der Skalen und auch das Krankheitskontinuum der PPD, mit einem fließenden Übergang von mittelschwerer zu schwerer Symptomatik, wider [66; 67]. Die Auswahl des Tests und die Interpretation der Ergebnisse sollten stets im Kontext des gesamten klinischen Bildes erfolgen, auch im Hinblick darauf, dass die HAMD-17 überwiegend in klinischen Studien hinzugezogen wird, während die Bestimmung der Schwere der Erkrankung im Klinikalltag auch über andere Instrumente wie den MADRS oder im Gespräch erfolgen kann. Somit können die MADRS-Auswertungen ergänzend zu den HAMD-Gesamtwerten hinzugezogen werden und zeigen, dass die Studie Patientinnen mit einem Spektrum mittelschwerer bis schwerer Symptomatik eingeschlossen hat, was die klinische Realität besser abbildet.

Tabelle 4-16: Verteilung der Patienten nach Schwere der Erkrankung mittels MADRS-Gesamtwert zu Baseline in der ITT-Population der Studien ROBIN und SKYLARK

Variable	**Statistische Größe oder Kategorie**	**ROBIN & SKYLARK**		
		Zuranolon + PiT (N = 176)	**PiT (N = 177)**	**Gesamt (N = 150)**
Baseline MADRS- Wert	**n**	**174**	**172**	**346**
20 – 34 (moderat)	n (%)	81 (46,6)	71 (41,3)	152 (43,9)
≥ 35 (schwer)	n (%)	93 (53,4)	101 (58,7)	194 (56,1)
Quelle: [68] Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.				

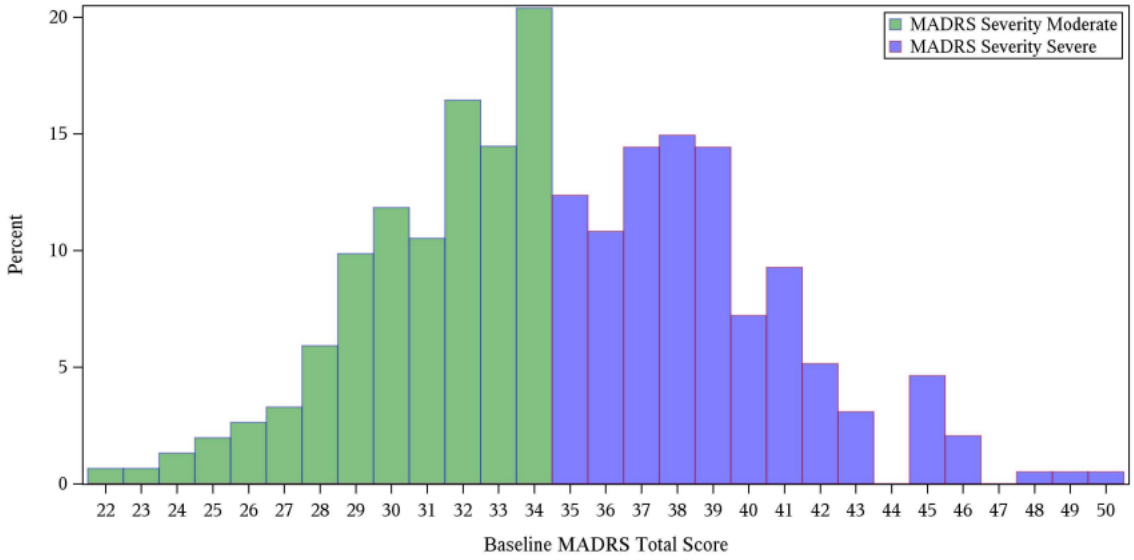


Abbildung 4-2: Histogramm der MADRS-Gesamtwerte zu Baseline nach Schweregrad
Quelle: [68] basierend auf der ITT-Population der Studien ROBIN und SKYLARK

Behandlungsdauer

In beiden Studien war eine 14-tägige Behandlungsdauer geplant und tatsächlich wurden Patientinnen die an der ROBIN-Studie (Part B) teilnahmen im Schnitt 13,6 Tage (siehe Tabelle 4-17) mit der Studienmedikation behandelt, während Patientinnen in der SKYLARK-Studie durchschnittlich 13,3 Tage die Studienmedikation erhielten.

Tabelle 4-17: Tatsächliche Behandlungsdosis und -dauer in der Studie ROBIN

Variable/Statistik	PiT (N = 73)	Zuranolon + PiT (N = 78)	Gesamt (N = 151)
Tolerierte Dosis			
20 mg	NA	3 (3,8 %)	3 (2,0 %)
30 mg	NA	74 (94,9 %)	74 (49,0 %)
Absetzen von 20 mg aufgrund von Unverträglichkeit	NA	0	0
Absetzen von 30 mg aufgrund von Unverträglichkeit	NA	1 (1,3%)	1 (0,7%)
Gesamte Exposition (Tage)			
MW (SD)	13,8 (1,42)	13,4 (2,77)	13,6 (2,23)
Median	14,0	14,0	14,0
Min, Max	2; 15	1; 17	1; 17
Quelle: CSR [15] Tabelle 14.1.5; basierend auf dem Safety Set (Part B) Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.			

Tabelle 4-18: Tatsächliche Behandlungsdosis und -dauer in der Studie SKYLARK

Variable/Statistik	PiT (N = 98)	Zuranolon + PiT (N = 98)	Gesamt (n = 196)
Anzahl der Patientinnen mit einer Dosis-Adjustierung			
Dosis reduziert	1 (1,0 %)	16 (16,3 %)	17 (8,7 %)
Dosis unterbrochen	0	0	0
Dosis eingestellt	0	2 (2,0 %)	2 (1,0 %)
Gesamte Exposition (Tage)			
MW (SD)	13,4 (2,32)	13,3 (2,19)	13,3 (2,25)
Median	14,0	14,0	14,0
Min, Max	2; 15	3; 15	2; 15
Quelle: CSR [14] Tabelle 14.1.4.10; basierend auf dem Safety Set. Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.			

Absetzen der Studienmedikation und Folgetherapien

Fehlende Werte aufgrund von Studienabbrüchen werden für die Betrachtung der Morbidität sowie gesundheitsbezogenen Lebensqualität adäquat ersetzt (vgl. Abschnitt 4.2.5.3). In der Studie ROBIN setzten 3 Patientinnen (2 %) die Studienmedikation vorzeitig ab [Abschnitt 12.1,15]; in der Studie SKYLARK waren es 16 Patientinnen (8,2 %) [Tabelle 7,14]. Im

Rahmen der Studien waren keine Folgetherapien vorgesehen und wurden entsprechend nicht dokumentiert.

Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext

Die PPD ist eine komplexe Erkrankung, die potenziell jede Frau im Zeitraum um die Geburt betreffen kann. Neben dem Zusammenhang mit der hormonellen Umstellung sind besonders die Frauen gefährdet, bei denen Risikofaktoren wie bestehende Angststörungen, fehlende soziale Unterstützung oder belastende Partnerbeziehungen vorliegen. Diese Faktoren verdeutlichen, dass die Entstehung einer PPD nicht ausschließlich von biologischen Faktoren abhängt, sondern auch durch psychosoziale Einflüsse geprägt wird, die oft schwer zu erfassen und systematisch zu dokumentieren sind. Die unzureichende Datenlage zur PPD in Deutschland und Europa stellt eine Herausforderung dar, denn selbst wenn es ein homogenes Profil der PPD-Patientin gäbe, dieses derzeit nicht durch valide Daten abgebildet werden kann.

Das durchschnittliche Alter der Patientinnen in der Meta-Analyse beträgt 29,5 Jahre und stimmt mit dem durchschnittlichen Alter der Mütter in Deutschland bei der Geburt ihres ersten Kindes überein [69]. Die Diagnose nach SCID der wiederum auf den Diagnosekriterien des DSM-5 beruht, wird auch in Deutschland empfohlen und eingesetzt [22]. Darüber hinaus kann aufgrund der unzureichenden Datenlage zur durchschnittlichen deutschen PPD-Population kein Vergleich mit der Studienpopulation angestellt werden. Eine detaillierte Beschreibung der Studienpopulation kann der Tabelle 4-15 und dem Abschnitt „Charakterisierung der Patientenpopulation“ entnommen werden.

In der europäischen Leitlinie wird außerdem hervorgehoben, dass sich der Entwicklungszustand des Landes auch auf die Entwicklung der PPD auswirkt [70]. Hier erscheint eine Übertragbarkeit der Daten aus dem US-amerikanischen, bzw. spanischen oder britischen Kontext als gegeben. Auch in Hinblick auf die Behandlungsempfehlungen im Hinblick auf den Schweregrad der PPD, stimmen die europäische und die amerikanische Leitlinie weitestgehend überein [70; 71].

4.3.1.2.2 Verzerrungspotenzial auf Studienebene

Bewerten Sie das Verzerrungspotenzial der RCT auf Studienebene mithilfe des Bewertungsbogens in Anhang 4-F. 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-F.

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

Studie	Adäquate Erzeugung der Randomisierungs-sequenz	Verdeckung der Gruppenzuteilung	Verblindung		Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial auf Studienebene
			Patient	Behandelnde Personen			
<Studie 1>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>	<hoch / niedrig>
ROBIN	Ja	Ja	Ja	Ja	Ja	Ja	niedrig
SKYLARK	Ja	Ja	Ja	Ja	Ja	Ja	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Sofern eine Bewertung des Verzerrungspotenzials auf Studienebene im EU-Dossier hinterlegt ist und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen. Die obenstehende Tabelle 4-9 ist im Dossier unabhängig von den Angaben im EU-Dossier auszufüllen.

Die vorliegende Bewertung des Verzerrungspotentials begründet sich jeweils auf der Auswertung der Studienmethodik in Anhang 4-E sowie auf der in Anhang 4-F dargestellten Bewertung des Verzerrungspotentials der beiden Studien ROBIN und SKYLARK.

Verzerrungspotential der Studie ROBIN

Bei der Studie ROBIN handelt es sich um eine multizentrische, randomisierte, kontrollierte und doppelt verblindete Phase 3 Studie im Parallelgruppendesign. Den Teilnehmerinnen, die die Einschlusskriterien erfüllten, wurde vor der Protokollversion 8.0 nacheinander eine Randomisierungsnummer aus dem Randomisierungsschema zugewiesen, das dem entblindeten pharmazeutischen Personal zur Verfügung gestellt wurde. Ab der Protokollversion 8.0 wurde die Randomisierung zentral über eine IRT durchgeführt.

In der Studie ROBIN fand eine doppelte Verblindung statt, dementsprechend waren Patientinnen, Kliniker*innen und das Studienteam über die Zuteilung zur Behandlung verblindet. Vor der Protokolländerung 7, Version 8.0, waren das Apothekenpersonal, das für die Abgabe des Studienmedikaments verantwortlich war, über das Randomisierungsschema informiert. Unter Protokolländerung 7, Version 8.0 und danach, wurde die Randomisierung zentral über ein IRT-System durchgeführt, und alle Mitarbeitenden vor Ort, einschließlich des Apothekenpersonals, waren verblindet; entblindetes pharmazeutisches Personal, das vor Änderung 7 für die Abgabe des Studienmedikaments verantwortlich war, waren nach der Einführung des IRT-Systems nicht mehr an der Durchführung der Studie beteiligt.

Nur im Falle eines medizinischen Notfalls durfte das zuständige Apothekenpersonal den tatsächlichen Inhalt des Studienmedikaments dem Prüfpersonal offenlegen. Es trat im Rahmen der Studie kein medizinischer Notfall ein.

Die in der Studie ROBIN erhobenen Endpunkte, sowie die durchzuführenden Analysen werden wie im Studienprotokoll und im statistischen Analyseplan präspezifiziert erhoben. *Post-hoc* definierte Analysen wurden ergänzt, um den Anforderungen der Nutzenbewertung zu entsprechen.

Das ITT-Prinzip wurde in der Studie adäquat umgesetzt. Es liegen keine sonstigen Aspekte vor, die das Verzerrungspotential beeinflussen können. Abschließend wird das Verzerrungspotential somit als niedrig eingestuft.

Verzerrungspotential der Studien SKYLARK

Bei der zulassungsrelevanten Studie SKYLARK handelt es sich um eine multizentrische, randomisierte, kontrollierte und doppelt verblindete Phase 3 Studie im Parallelgruppendesign.

Die Randomisierung erfolgt zentral über ein IRT-System. Jeder Teilnehmerin wurde eine ID und eine zugehörige Medikamenten-ID vermittelt. Somit ist die adäquate Erzeugung der Randomisierungssequenz sowie eine verdeckte Gruppenzuteilung gewährleistet.

Bei der Studie SKYLARK handelt es sich um eine doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam bis zur Datenbanksperre verblindet gegenüber der Behandlungszuweisung. Das Randomisierungsschema und die IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung vom Prüfpersonal über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing. Es fand während der Studie keine Aufhebung der Verblindung statt.

Die in der Studie SKYLARK erhobenen Endpunkte, sowie die durchzuführenden Analysen werden wie im Studienprotokoll und im statistischen Analyseplan präspezifiziert erhoben. *Post-hoc* definierte Analysen wurden ergänzt, um den Anforderungen der Nutzenbewertung zu entsprechen.

Das ITT-Prinzip wurde in der Studie adäquat umgesetzt. Es liegen keine sonstigen Aspekte vor, die das Verzerrungspotential beeinflussen können. Abschließend wird das Verzerrungspotential somit als niedrig eingestuft.

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.

Sofern ein Überblick über die patientenrelevanten Endpunkte, auf denen die Bewertung des medizinischen Nutzens und Zusatznutzens beruht, im EU-Dossier hinterlegt ist und dieser Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

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

Endpunkt	ROBIN	SKYLARK	Meta-Analyse
Mortalität			
Todesfälle	●	●	●
Morbidität			
HAMD-17	●	●	●
MADRS	●	●	●
EPDS	●	●	●
PHQ-9	●	●	●
HAM-A	●	●	●
CGI-I	●	●	●
CGI-S	●	●	●
Gesundheitsbezogene Lebensqualität			
SF-36	●	-	-
BIMF	●	-	-
Sicherheit			
C-SSRS	●	●	●
UE	●	●	●
Schwere UE	●	●	●
SUE	●	●	●
Therapieabbruch aufgrund eines UE	●	●	●

●: Endpunkt wurde in der Studie erhoben; -: Endpunkt wurde in der Studie nicht erhoben.
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.

4.3.1.3.1.1 Mortalität – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst drei 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 Meta-Analyse 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 inklusive Angaben zur Häufigkeit von und zum Umgang mit nicht oder nicht vollständig beobachteten Patienten (bei Verlaufs-beobachtungen pro Messzeitpunkt)
- dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
- bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
- bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion
- entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen gegebenenfalls adjustiert wurde.

Unterschiedliche Beobachtungszeiten zwischen den Behandlungsgruppen sollen durch adäquate Analysen (zum Beispiel Überlebenszeitanalysen) adressiert werden, und zwar für alle Endpunkte (einschließlich UE nach den nachfolgend genannten Kriterien), für die eine solche Analyse aufgrund deutlich unterschiedlicher Beobachtungszeiten erforderlich ist.

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko sowie zum Anteil der zensierten Patienten im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden. Dabei ist für jeden Endpunkt, für den eine solche Analyse durchgeführt wird, eine separate Kaplan-Meier-Kurve darzustellen.

Zu mit Skalen erhobenen patientenberichteten Endpunkten (zum Beispiel zur gesundheitsbezogenen Lebensqualität oder zu Symptomen) sind immer auch die Werte im Studienverlauf anzugeben, auch als grafische Darstellung, sowie eine Auswertung, die die über den Studienverlauf ermittelten Informationen vollständig berücksichtigt (zum Beispiel als Symptomlast über die Zeit, geschätzt mittels MMRM-Analyse [falls aufgrund der Datenlage geeignet]).

Bei MMRM-Analysen handelt es sich um komplexe Auswertungen, die verschiedene Effekte schätzen können. Daher sollten Hypothesen zum Behandlungseffekt und die Modelle möglichst präspezifiziert und Abweichungen davon beschrieben und diskutiert werden. Es sollte beschrieben werden, welche Variable modelliert wird (der erhobene Wert selbst oder die Veränderung gegenüber dem Ausgangswert [change-from-baseline]). Auch sollte das jeweilige Auswertungsmodell exakt beschrieben werden (feste und zufällige Effekte, bei den entsprechenden Variablen auch die verwendete Skalierung beziehungsweise Merkmals-

Kategorisierung, die Kovarianzstruktur und Interaktionsterme). Es muss erkenntlich sein, ob ein Unterschied zwischen den Behandlungsgruppen zu einem festen Zeitpunkt oder als gemittelter Wert über mehrere Zeitpunkte (und welche) geschätzt wurde. Darüber hinaus muss erkenntlich sein, wie viele und welche Erhebungszeitpunkte in das MMRM eingehen (mit Angabe des im Modell verwendeten Variablennamen für die Zeitpunkte, zum Beispiel Visite). Es sollten Angaben zur Anzahl der in die MMRM Auswertung für den Effektschätzer eingegangenen Patientinnen und Patienten und dem Anteil der fehlenden Werte pro Arm und pro Erhebungszeitpunkte gemacht werden. Falls die Erhebungszeitpunkte ab Randomisierung nicht zu gleichen parallelen Zeitpunkten in beiden Studienarmen stattfanden, sondern zum Beispiel infolge von unterschiedlichen Therapieregimen/Therapiezyklendauern zeitlich versetzt stattfanden, kann dies die Validität der Analyse anhand des MMRM-Modell einschränken, welches möglichst parallele Messzeitpunkte voraussetzt. Daher sollten Angaben dazu gemacht werden, wie die Erhebungszeitpunkte den Studienvisiten zugeordnet wurden und wie im Modell damit umgegangen wurde. Sensitivitätsanalysen sollten beschrieben und dargestellt werden.

Die Auswertung von Responderanalysen mittels klinischer Relevanzschwellen bei komplexen Skalen soll nach dem folgenden Vorgehen erfolgen:

1. Falls in einer Studie Responderanalysen unter Verwendung einer MID präspezifiziert sind und das Responsekriterium mindestens 15 % der Skalenspannweite des verwendeten Erhebungsinstruments entspricht, sind diese Responderanalysen für die Bewertung darzustellen.

2. Falls präspezifiziert Responsekriterien im Sinne einer MID unterhalb von 15 % der Skalenspannweite liegen, bestehen in diesen Fällen und solchen, in denen gar keine Responsekriterien präspezifiziert wurden, aber stattdessen Analysen kontinuierlicher Daten zur Verfügung stehen, verschiedene Möglichkeiten. Entweder können post hoc spezifizierte Analysen mit einem Responsekriterium von genau 15 % der Skalenspannweite dargestellt werden. Alternativ können Analysen der kontinuierlichen Daten dargestellt werden, für die Relevanzbewertung ist dabei auf ein allgemeines statistisches Maß in Form von standardisierten Mittelwertdifferenzen (SMDs, in Form von Hedges' g) zurückzugreifen. Dabei ist eine Irrelevanzschwelle als Intervall von -0,2 bis 0,2 zu verwenden: Liegt das zum Effektschätzer korrespondierende Konfidenzintervall vollständig außerhalb dieses Irrelevanzbereichs, wird davon ausgegangen, dass die Effektstärke nicht in einem sicher irrelevanten Bereich liegt. Dies soll gewährleisten, dass der Effekt hinreichend sicher mindestens als klein angesehen werden kann.

3. Liegen sowohl geeignete Responderanalysen (Responsekriterium präspezifiziert mindestens 15 % der Skalenspannweite oder post hoc genau 15 % der Skalenspannweite) als auch Analysen stetiger Daten vor, sind die Responderanalysen darzustellen.

Zu UE sind folgende Auswertungen vorzulegen:

1. Gesamtrate UE,

2. Gesamtrate schwerwiegender UE (SUE),

3. Gesamtrate der Abbrüche wegen UE,

4. Gesamtrate schwerer UE, sofern dies in der/den relevante/n Studie/n erhoben wurde (zum Beispiel gemäß CTCAE ≥ 3 und/oder einer anderen etablierten beziehungsweise validierten indikationsspezifischen Klassifikation).

5. zu den unter 1, 2 und 4 genannten Kategorien (UE ohne weitere Differenzierung, SUE, UE differenziert nach Schweregrad) soll zusätzlich zu den Gesamtraten die Darstellung nach Organsystemen und Einzelereignissen (als SOCs und PT) jeweils nach folgenden Kriterien erfolgen:

   – UE (unabhängig vom Schweregrad): Ereignisse, die bei mindestens 10 % der Patienten in einem Studienarm aufgetreten sind
   – Schwere UE (zum Beispiel CTCAE-Grad ≥ 3) und SUE: Ereignisse, die bei mindestens 5% der Patienten in einem Studienarm aufgetreten sind
   – zusätzlich für alle Ereignisse unabhängig vom Schweregrad: Ereignisse, die bei mindestens 10 Patienten UND bei mindestens 1 % der Patienten in einem Studienarm aufgetreten sind.

Es ist zu beachten, dass bei der Berechnung der Häufigkeiten für die SOC alle PT, auch solche mit einer Häufigkeit unterhalb der vorzulegenden Grenzen, berücksichtigt werden.

6. zu Kategorie 3: Die Abbruchgründe auf SOC/PT-Ebene müssen vollständig, jedoch nur deskriptiv dargestellt werden.

Sofern bei der Erhebung unerwünschter Ereignisse erkrankungsbezogenen Ereignisse (zum Beispiel Progression, Exazerbation) berücksichtigt werden (diese Ereignisse also in die UE-Erhebung eingehen), sollen für die Gesamtraten (UE, schwere UE, SUE und Abbruch wegen UE) zusätzliche UE-Analysen durchgeführt werden, bei denen diese Ereignisse unberücksichtigt bleiben. Alle Auswertungen zu UE können auch in einem separaten Anhang des vorliegenden Modul 4 dargestellt werden. Dabei kann die Ausgabe der Statistik-Software unverändert verwendet werden, sofern diese alle notwendigen Angaben enthält. Eine Darstellung ausschließlich in Modul 5 ist nicht ausreichend. Davon unbenommen sind die Gesamtraten (UE, schwere UE, SUE und Abbrüche wegen UE), sowie die für die Gesamtaussage zum Zusatznutzen herangezogenen Ergebnisse im vorliegenden Abschnitt darzustellen.

Auswertungen zu den im Abschnitt 4.3.1.2.1 aufgeführten Datenschnitten sollen vollständig, das heißt für alle erhobenen relevanten Endpunkte, durchgeführt und vorgelegt werden. Das gilt auch dann, wenn ein Datenschnitt ursprünglich nur zur Auswertung einzelner Endpunkte geplant war. Auf die Darstellung der Ergebnisse einzelner Endpunkte eines Datenschnitts beziehungsweise eines gesamten Datenschnitts kann verzichtet werden, wenn hierdurch kein wesentlicher Informationsgewinn gegenüber einem anderen Datenschnitt zu erwarten ist (zum Beispiel wenn die Nachbeobachtung zu einem Endpunkt bereits zum vorhergehenden

Datenschnitt nahezu vollständig war oder ein Datenschnitt in unmittelbarer zeitlicher Nähe zu einem anderen Datenschnitt liegt).

Falls für die Auswertung eine andere Population als die ITT-Population herangezogen wird, soll diese benannt (zum Beispiel Safety-Population) und definiert werden.

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (zum Beispiel Patientengruppen) und methodischen (zum Beispiel Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde beziehungsweise warum einzelne Studien gegebenenfalls nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen 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). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (zum Beispiel Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

4.3.1.3.1.1.1 Todesfälle – RCT

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

Sofern Informationen zur Operationalisierung des jeweiligen Endpunkts im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-21: Operationalisierung von Todesfällen

Studie	Operationalisierung
ROBIN	Die Anzahl der Todesfälle wurde im Rahmen der Sicherheitsendpunkte erhoben.
SKYLARK	Die Anzahl der Todesfälle wurde im Rahmen der Sicherheitsendpunkte erhoben.
Meta-Analyse	Sowohl in der ROBIN-Studie als auch in der SKYLARK-Studie gab es keine Todesfälle. Dementsprechend wurde keine Meta-Analyse durchgeführt.
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. 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-F.

Tabelle 4-22: Bewertung des Verzerrungspotenzials für Todesfälle in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
ROBIN	niedrig	ja	ja	ja	ja	niedrig
SKYLARK	niedrig	ja	ja	ja	ja	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Sofern eine Bewertung des Verzerrungspotenzials für den jeweiligen Endpunkt im EU-Dossier hinterlegt ist und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen. Die obenstehende Tabelle 4-12 ist im Dossier unabhängig von den Angaben im EU-Dossier auszufüllen.

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam bis zur Datenbanksperre verblindet gegenüber der Behandlungszuweisung.

In den Studien ROBIN und SKYLARK erfolgte die Erhebung des Endpunkts „Todesfälle“ im Rahmen der schwerwiegenden UE. Es wurden objektiv jegliche UE der Patientinnen, unabhängig davon, ob das Ereignis in kausalem Zusammenhang mit der Verabreichung der Studienmedikation stand, erhoben. Die Erhebung und Klassifikation der UE erfolgten auf Basis allgemein anerkannter internationaler Standards. Somit ist eine ergebnisunabhängige Erhebung gewährleistet.

Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

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

Sofern Ergebnisse für den jeweiligen Endpunkt im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-23: Ergebnisse für Todesfälle aus RCT mit dem zu bewertenden Arzneimittel

Studie	
ROBIN	Es trat kein Todesfall auf.
SKYLARK	Es trat kein Todesfall auf.
Quellen: [15] Tabelle 14.3.2.2; [14] Tabelle 35	

In beiden Studien wurden keine Todesfälle beobachtet.

Übertragbarkeit auf den deutschen Versorgungskontext

Wie bereits in Abschnitt 4.3.1.2.1 ausgeführt, wird von einer Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ausgegangen.

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.

4.3.1.3.1.2 Morbidität – RCT

4.3.1.3.1.2.1 HAMD-17 – RCT

Tabelle 4-24: Operationalisierung von HAMD-17

Studie	Operationalisierung
ROBIN	Die HAMD-17 wird verwendet, um das Vorhandensein und die Schwere von verschiedenen Symptomen der Depression zu evaluieren und quantifizieren. Der HAMD-17-Fragebogen zur Fremdbeurteilung besteht aus 17 Items mit folgenden 17 unterschiedlichen Symptomen: - Depressive Stimmung, - Schuldgefühle, - Suizidalität, - Einschlafstörungen, - Durchschlafstörungen, - Schlafstörungen am Morgen, - Beeinträchtigungen von Arbeit und Tätigkeiten/Aktivitäten, - Psychomotorische Hemmung/Verlangsamung, - Psychomotorische Unruhe/Erregung, - Psychische Angstsymptome, - Körperliche Angstsymptome, - Gastrointestinale Symptome/Appetitverlust, - Allgemeine körperliche Symptome, - Genitalsymptome/Libidoverlust, - Hypochondrie, - Gewichtsverlust, - Mangelnde Krankheitseinsicht. Jedes Item kann einen Wert von entweder 0 – 2 oder 0 – 4 annehmen (Gesamtspanne 0 – 52), wobei höhere Werte eine höhere Symptomschwere bedeuten. <u>Erhebende Person und Erhebungszeitpunkte</u> Die HAMD-17 wurde von Kliniker*innen zu folgenden Zeitpunkten erhoben: - Vor Amendment 4: Screening, vor der Verabreichung an Tag 1 und 12 ± 1 Stunde nach der abendlichen Verabreichung an Tag 2, 3, 4, 5, 6, 8, 15, 21 und 45. - Nach Amendment 4: zu den Visiten beim Screening, vor der Dosierung an Tag 1 und 12 ±1 Stunde nach der abendlichen Verabreichung an Tag 3, 8, 15; und morgens an Tag 21 und 45. <u>Schwellenwerte</u> Das Ansprechen nach HAMD-17 wurde im Statistischen Analyseplan als eine Reduktion von ≥ 50 % gegenüber dem Gesamtwert zu Baseline prädefiniert. Die Remission nach HAMD-17 wurde im Statistischen Analyseplan mit einem Gesamtwert ≤ 7 prädefiniert. <u>Dargestellte Analysen:</u> Folgende Analysen werden für die ITT Population der Studie ROBIN in Anhang 4-H dargestellt: - Anzahl der Patientinnen mit HAMD-17-Ansprechen - Zeit bis zum ersten HAMD-17-Ansprechen (inkl. Kaplan-Meier-Plot) - Anzahl der Patientinnen mit HAMD-17-Remission - Zeit bis zur ersten HAMD-17-Remission (inkl. Kaplan-Meier-Plot)

Studie	Operationalisierung
SKY-LARK	Die Definition des Endpunkts HAMD-17 entspricht der Definition in der Studie ROBIN. Erhebende Person und Erhebungszeitpunkte Die HAMD-17 wurde von trainierten und zertifizierten Kliniker*innen zu folgenden Zeitpunkten erhoben: - Zum Screening und zu den Visiten an Tag 1, 3, 8, 15, 21, 28 und 45. Jeweils so früh wie möglich zu jeder Visite. Schwellenwerte Die Definition der Schwellenwerte des Endpunkts HAMD-17 entspricht der Definition der Studie ROBIN. Dargestellte Analysen: Folgende Analysen werden für die ITT-Population der Studie SKYLARK in Anhang 4-Idargestellt: - Siehe ROBIN
Meta-Analyse	Die Definition des Endpunkts HAMD-17 entspricht der Definition in der Studie ROBIN. Aufgrund den leicht abweichenden Erhebungszeitpunkten wurden die Daten der folgenden Erhebungszeitpunkte gepoolt: Tag 3, 8, 15, 21 und Tag 45. Dargestellte Analysen: Folgende Analysen werden für die gepoolte ITT-Population beider Studien nachfolgend dargestellt: - Anzahl der Patientinnen mit HAMD-17-Ansprechen - Zeit bis zum ersten HAMD-17-Ansprechen (inkl. Kaplan-Meier-Plot) - Anzahl der Patientinnen mit HAMD-17-Remission - Zeit bis zur ersten HAMD-17-Remission (inkl. Kaplan-Meier-Plot) Dazugehörige Statistische Methoden: Eine ausführliche Beschreibung der Methodik, die auf dem SKYLARK SAP beruht, findet sich in Abschnitt 4.2.5 bzw. in den Fußnoten. Zudem wurden Heterogenitätstests durchgeführt, die *post-hoc* definiert wurden und deren methodische Grundlage im gleichen Abschnitt zu finden sind.
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.	

Tabelle 4-25: Bewertung des Verzerrungspotenzials für HAMD-17 in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
ROBIN	niedrig	ja	ja	ja	ja	niedrig
SKYLARK	niedrig	ja	ja	ja	ja	niedrig

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das

Studienteam verblindet gegenüber der Behandlungszuweisung. Der Endpunkt wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Es wird nicht von den präspezifizierten Methoden und Definitionen abgewichen. Das ITT-Prinzip wurde adäquat umgesetzt.

Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Tabelle 4-26: Rücklaufrate für HAMD-17 aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

HAMD-17-Rücklaufrate	**Zuranolon+PiT N=176**	**PiT N=177**
Tag 3	172 (98 %)	170 (96 %)
Tag 8	168 (95 %)	169 (95 %)
Tag 15	167 (95 %)	163 (92 %)
Tag 21	158 (90 %)	156 (88 %)
Tag 45	157 (89 %)	154 (87 %)
Werte sind aus den MMRM Analysen entnommen (s. Anhang 4-G) Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.		

Die Rücklaufraten für die HAMD-17 lagen zu jedem Zeitpunkt über 87 %.

Tabelle 4-27: Ergebnisse für HAMD-17-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

HAMD-17-Ansprechen (50 %-Reduktion von Baseline)	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	173
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	56 (32,2)	32 (18,5)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (1,1)	3 (1,7)
RR	1,723	
95 %-KI für das RR	[1,186; 2,503]	
p-Wert für das RR	0,0043	
OR	2,147	

HAMD-17-Ansprechen (50 %-Reduktion von Baseline)	**Zuranolon+PiT**	**PiT**
95 %-KI für das OR	[1,301; 3,541]	
p-Wert für das OR	0,0028	
ARR	0,137	
95 %-KI für das ARR	[0,050; 0,224]	
p-Wert für das ARR	0,0020	
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	96 (55,2)	57 (32,9)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (3,4)	4 (2,3)
RR	1,655	
95 %-KI für das RR	[1,299; 2,110]	
p-Wert für das RR	0,0000	
OR	2,595	
95 %-KI für das OR	[1,667; 4,040]	
p-Wert für das OR	0,0000	
ARR	0,226	
95 %-KI für das ARR	[0,127; 0,326]	
p-Wert für das ARR	0,0000	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	106 (60,9)	70 (40,5)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	7 (4,0)	10 (5,8)
RR	1,507	
95 %-KI für das RR	[1,219; 1,863]	
p-Wert für das RR	0,0002	
OR	2,338	
95 %-KI für das OR	[1,510; 3,620]	
p-Wert für das OR	0,0001	
ARR	0,207	

HAMD-17-Ansprechen (50 %-Reduktion von Baseline)	**Zuranolon+PiT**	**PiT**
95 %-KI für das ARR	[0,106; 0,309]	
p-Wert für das ARR	0,0001	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	103 (59,2)	76 (43,9)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	16 (9,2)	17 (9,8)
RR	1,342	
95 %-KI für das RR	[1,098; 1,641]	
p-Wert für das RR	0,0041	
OR	1,904	
95 %-KI für das OR	[1,229; 2,951]	
p-Wert für das OR	0,0040	
ARR	0,155	
95 %-KI für das ARR	[0,054; 0,257]	
p-Wert für das ARR	0,0027	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	107 (61,5)	85 (49,1)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	17 (9,8)	19 (11,0)
RR	1,269	
95 %-KI für das RR	[1,051; 1,532]	
p-Wert für das RR	0,0130	
OR	1,679	
95 %-KI für das OR	[1,092; 2,581]	
p-Wert für das OR	0,0182	
ARR	0,129	
95 %-KI für das ARR	[0,026; 0,231]	
p-Wert für das ARR	0,0140	
Heterogenitätstest (p-Wert)	0,6264	

HAMD-17-Ansprechen (50 %-Reduktion von Baseline)	Zuranolon+PiT	PiT
[1] Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert des HAMD-17, der Studie, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Studie sowie der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells. [2] Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Information: p-Werte, die hier mit „0,0000" angegeben sind, entsprechen p < 0,0001 Quelle: [72]		

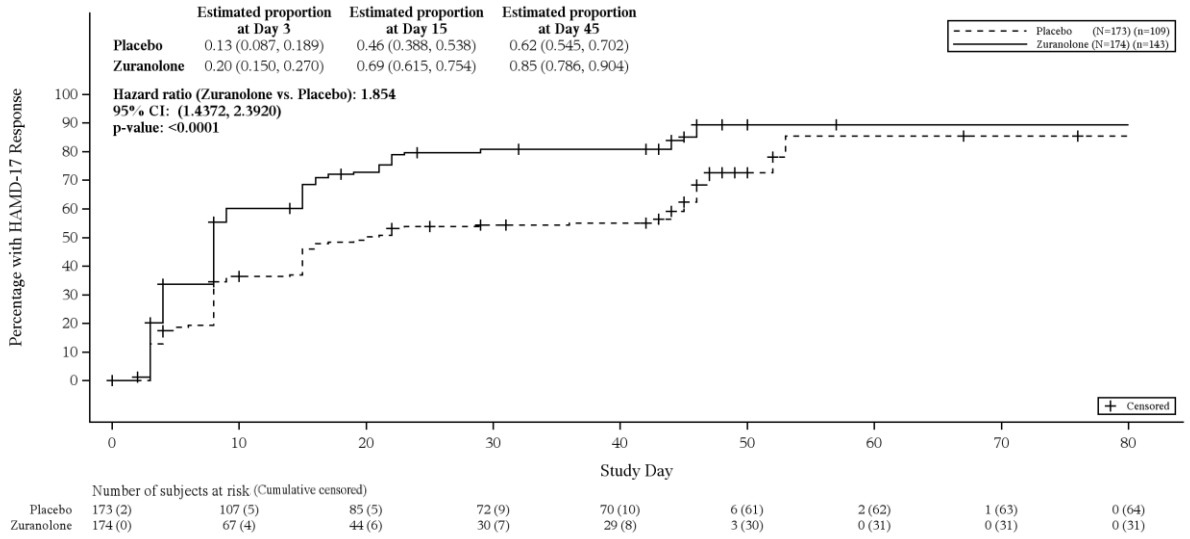
Im Zuranolon+PiT-Arm konnten zu allen Zeitpunkten statistisch signifikant mehr Patientinnen eine Verbesserung um ≥ 50 % gegenüber dem Gesamtwert zu Baseline auf der HAMD-17-Skala erreichen als unter der Vergleichstherapie.

Tabelle 4-28: Ergebnisse für Zeit bis zum ersten HAMD-17-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten HAMD-17-Ansprechen	Zuranolon+PiT	PiT
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	173
Anzahl der Patientinnen mit Antwort, n (%) [2]	143 (82,2)	109 (63,0)
Zeit bis zum ersten HAMD-17-Ansprechen (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	4,0 [3,0; 4,0]	8,0 [5,0; 8,0]
50. Perzentil - Median (95 %-KI)	8,0 [8,0; 9,0]	20,0 [15,0; 44,0]
75. Perzentil - Q3 (95 %-KI)	21,0 [15,0; 44,0]	52,0 [46,0; NB]
HR - Hazard Ratio [4]	1,854	
95 %-KI für HR	[1,4372; 2,3920]	
p-Wert für HR	< 0,0001	
Heterogenitätstest (p-Wert)	0,8211	

Zeit bis zum ersten HAMD-17-Ansprechen	Zuranolon+PiT	PiT
Das Ansprechen nach HAMD-17 ist definiert als Reduktion von ≥ 50 % gegenüber dem Gesamtwert zu Baseline. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren HAMD-17-Bewertung zensiert. [1] Anzahl der Patientinnen mit nicht fehlenden HAMD-17-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit einer HAMD-17-Ansprechrate. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. [3] Kaplan-Meier-Schätzer. [4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Studie, Behandlung, Baselinewert des HAMD-17-Gesamtwerts und Antidepressiva zu Baseline (Ja oder Nein). Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Patientinnen im Zuranolon+PiT-Arm sprachen statistisch signifikant schneller (8 vs. 20 Tage im Median) auf der HAMD-17-Skala an als Patientinnen im Vergleichsarm.



[1] Patientinnen mit nicht fehlenden HAMD-17-Gesamtwerten zu Baseline sind in die Analyse einbezogen. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Das Ansprechen nach HAMD-17 ist definiert als Reduktion von ≥ 50 % gegenüber dem Gesamtwert zu Baseline. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren HAMD-17-Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Behandlung, Baselinewert des HAMD-17, Antidepressiva zu Baseline (Ja oder Nein) und Studie angepasst ist.

Abbildung 4-3: Kaplan-Meier-Plot für Zeit bis zum ersten HAMD-17-Ansprechen

Tabelle 4-29: Ergebnisse für HAMD-17-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

HAMD-17-Remission (Gesamtwert ≤ 7)	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	173
Tag 3		
Anzahl der Patientinnen mit Remission, n (%) [2]	22 (12,6)	9 (5,2)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (1,1)	3 (1,7)
RR	2,414	
95 %-KI für das RR	[1,151; 5,064]	
p-Wert für das RR	0,0197	
OR	2,623	
95 %-KI für das OR	[1,165; 5,908]	
p-Wert für das OR	0,0199	
ARR	0,064	
95 %-KI für das ARR	[0,006; 0,122]	
p-Wert für das ARR	0,0306	
Tag 8		
Anzahl der Patientinnen mit Remission, n (%) [2]	41 (23,6)	25 (14,5)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (3,4)	4 (2,3)
RR	1,599	
95 %-KI für das RR	[1,022; 2,500]	
p-Wert für das RR	0,0398	
OR	1,821	
95 %-KI für das OR	[1,041; 3,187]	
p-Wert für das OR	0,0357	
ARR	0,085	
95 %-KI für das ARR	[0,006; 0,164]	
p-Wert für das ARR	0,0351	

HAMD-17-Remission (Gesamtwert ≤ 7)	**Zuranolon+PiT**	**PiT**
Tag 15		
Anzahl der Patientinnen mit Remission, n (%) [2]	58 (33,3)	32 (18,5)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	7 (4,0)	10 (5,8)
RR	1,773	
95 %-KI für das RR	[1,222; 2,574]	
p-Wert für das RR	0,0026	
OR	2,203	
95 %-KI für das OR	[1,330; 3,649]	
p-Wert für das OR	0,0022	
ARR	0,145	
95 %-KI für das ARR	[0,056; 0,234]	
p-Wert für das ARR	0,0014	
Tag 21		
Anzahl der Patientinnen mit Remission, n (%) [2]	56 (32,2)	38 (22,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	16 (9,2)	17 (9,8)
RR	1,441	
95 %-KI für das RR	[1,017; 2,042]	
p-Wert für das RR	0,0401	
OR	1,676	
95 %-KI für das OR	[1,030; 2,727]	
p-Wert für das OR	0,0377	
ARR	0,097	
95 %-KI für das ARR	[0,007; 0,188]	
p-Wert für das ARR	0,0356	
Tag 45		
Anzahl der Patientinnen mit Remission, n (%) [2]	76 (43,7)	46 (26,6)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	17 (9,8)	19 (11,0)

HAMD-17-Remission (Gesamtwert ≤ 7)	**Zuranolon+PiT**	**PiT**
RR	1,634	
95 %-KI für das RR	[1,213; 2,200]	
p-Wert für das RR	0,0012	
OR	2,128	
95 %-KI für das OR	[1,350; 3,353]	
p-Wert für das OR	0,0011	
ARR	0,167	
95 %-KI für das ARR	[0,069; 0,265]	
p-Wert für das ARR	0,0009	
Heterogenitätstest (p-Wert)	0,4869	
[1] Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert des HAMD-17, der Studie, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Studie sowie der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells. [2] Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Im Zuranolon+PiT-Arm konnte zu allen Zeitpunkten statistisch signifikant mehr Patientinnen eine Remission auf der HAMD-17-Skala erreichen als unter der Vergleichstherapie.

Tabelle 4-30: Ergebnisse für Zeit bis zur ersten HAMD-17-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

HAMD-17-Remission (Gesamtwert ≤ 7)	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	173
Anzahl der Patientinnen mit Antwort, n (%) [2]	99 (56,9)	66 (38,2)
Zeit bis zur ersten HAMD-17-Remission (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	9,0 [8,0; 15,0]	21,0 [15,0; 24,0]
50. Perzentil - Median (95 %-KI)	29,0 [21,0; 45,0]	50,0 [48,00; NB]

HAMD-17-Remission (Gesamtwert ≤ 7)	Zuranolon+PiT	PiT
75. Perzentil - Q3 (95 %-KI)	NB [51,0; NB]	NB [NB; NB]
HR - Hazard Ratio [4]	1,753	
95 %-KI für HR	[1,2818; 2,3962]	
p-Wert für HR	0,0004	
Heterogenitätstest (p-Wert)	0,7996	

Die HAMD-17-Remission ist definiert als HAMD-17 Gesamtwert ≤ 7. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren HAMD-17-Bewertung zensiert.

[1] Anzahl der Patientinnen mit nicht fehlenden HAMD-17-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

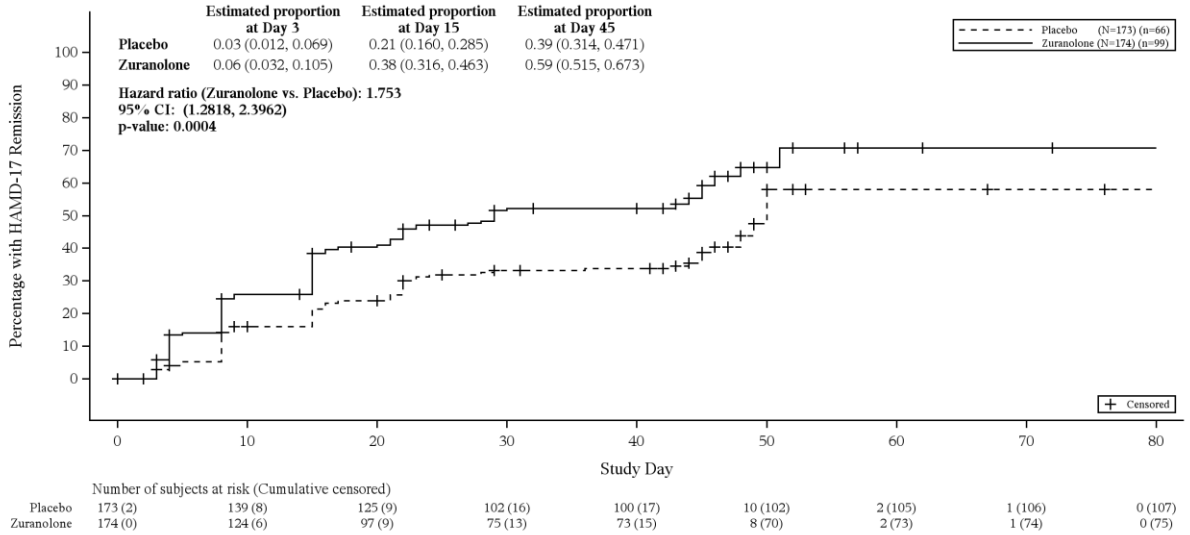
[2] Anzahl der Patientinnen mit einer HAMD-17-Remission. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen

[3] Kaplan-Meier-Schätzer.

[4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Studie, Behandlung, Baselinewert des HAMD-17-Gesamtwerts und Antidepressiva zu Baseline (Ja oder Nein)
Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Patientinnen im Zuranolon+PiT-Arm erreichten statistisch signifikant schneller (29 vs. 50 Tage im Median) eine Remission auf der HAMD-17-Skala als Patientinnen im Vergleichsarm.



[1] Patientinnen mit nicht fehlenden HAMD-17-Gesamtwerten zu Baseline sind in die Analyse einbezogen. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Die HAMD-17-Remission ist definiert als HAMD-17 Gesamtwert ≤ 7. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren HAMD-17-Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-

Modell, das für Behandlung, Baselinewert des HAMD-17, Antidepressiva zu Baseline (Ja oder Nein) und Studie angepasst ist.

Abbildung 4-4: Kaplan-Meier-Plot für Zeit bis zur ersten HAMD-17-Remission

Übertragbarkeit auf den deutschen Versorgungskontext

Wie bereits in Abschnitt 4.3.1.2.1 ausgeführt, wird von einer Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext ausgegangen.

4.3.1.3.1.2.2 MADRS – RCT

Tabelle 4-31: Operationalisierung von MADRS

Studie	Operationalisierung
ROBIN	Die MADRS Erhebung erfolgt durch unabhängige Prüfärzte auf Basis eines semi-strukturierten Interviews umfasst die folgenden 10 Items: 1. beobachtbare Traurigkeit, 2. berichtete Traurigkeit, 3. innere Anspannung, 4. Schlaflosigkeit, 5. Appetitverlust, 6. Konzentrationsschwierigkeiten, 7. Niedergeschlagenheit, 8. Gefühl der Gefühllosigkeit, 9. pessimistische Gedanken, 10. Suizidgedanken. Jedes Item kann einen Wert zwischen 0 – 6 haben (Gesamtbereich 0 – 60), wobei höhere Werte eine stärkere Symptomausprägung anzeigen. Befragende und Erhebungszeitpunkte Die MADRS wurde von dem Prüfpersonal zu folgenden Zeitpunkten durchgeführt: Während der Screening-Besuche, vor der Dosierung am Tag 1, 12 Stunden ± 1 Stunde nach der abendlichen Verabreichung an den Tagen 2 (nur vor Amendment 4), 3, 8 und 15; und morgens an den Tagen 21 und 45. Schwellenwert Das Ansprechen nach MADRS wurde im statistischen Analyseplan als eine Reduktion von ≥ 50 % gegenüber dem Baselinewert definiert. Die MADRS-Remission wurde im statistischen Analyseplan als eine Gesamtpunktzahl von ≤ 10 Punkten definiert. Dargestellte Analysen: Die folgenden Analysen werden für die ITT-Population der ROBIN-Studie in Anhang 4-H dargestellt: - Anzahl der Patientinnen mit MADRS-Ansprechen - Zeit bis zum ersten MADRS-Ansprechen (inkl. Kaplan-Meier-Plot) - Anzahl der Patientinnen mit MADRS-Remission - Zeit bis zur ersten MADRS-Remission (inkl. Kaplan-Meier-Plot)
SKY-LARK	Die Definition des MADRS-Endpunkts entspricht der Definition in der ROBIN-Studie. Befrager und Befragungszeitpunkte Die MADRS wurde vom Prüfpersonal zu folgenden Zeitpunkten durchgeführt: - Bei der Voruntersuchung und bei den Besuchen an den Tagen 1, 8, 15, 28 und 45. Schwellenwerte Die Definition der Schwellenwerte für den MADRS-Endpunkt entspricht der Definition in der ROBIN-Studie. Dargestellte Analysen: Die folgenden Analysen werden für die ITT-Population der SKYLARK-Studie in Anhang 4-I dargestellt: Siehe ROBIN.

Meta-Analyse	Die Definition des Endpunkts MADRS entspricht der Definition in der Studie ROBIN. Aufgrund den leicht abweichenden Erhebungszeitpunkten wurden die die Daten der folgenden Erhebungszeitpunkte gepoolt: Tag 8, 15 und Tag 45. Dargestellte Analysen: Folgende Analysen werden für die gepoolte ITT-Population beider Studien nachfolgend dargestellt: - Anzahl der Patientinnen mit MADRS-Ansprechen - Zeit bis zum ersten MADRS-Ansprechen (inkl. Kaplan-Meier-Plot) - Anzahl der Patientinnen mit MADRS-Remission - Zeit bis zur ersten MADRS-Remission (inkl. Kaplan-Meier-Plot) Dazugehörige Statistische Methoden: Eine ausführliche Beschreibung der Methodik, die auf dem SKYLARK SAP beruht, findet sich in Abschnitt 4.2.5 bzw. in den Fußnoten. Zudem wurden Heterogenitätstests durchgeführt, die *post-hoc* definiert wurden und deren methodische Grundlage im gleichen Abschnitt zu finden sind.
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.	

Tabelle 4-32: Bewertung des Verzerrungspotenzials für MADRS in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
ROBIN	niedrig	ja	ja	ja	ja	niedrig
SKYLARK	niedrig	ja	ja	ja	ja	niedrig

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der Endpunkt wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Es wird nicht von den präspezifizierten Methoden und Definitionen abgewichen. Das ITT-Prinzip wurde adäquat umgesetzt.

Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Tabelle 4-33: Rücklaufrate für MADRS aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

MADRS-Rücklaufrate	**Zuranolon+PiT N=176**	**PiT N=177**
Tag 8	168 (95 %)	168 (95 %)
Tag 15	166 (94 %)	162 (92 %)
Tag 45	157 (89 %)	153 (86 %)
Werte sind aus den MMRM Analysen entnommen (s. Anhang 4-G) Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.		

Die Rücklaufraten für die MADRS lagen zu jedem Zeitpunkt über 86 %.

Tabelle 4-34: Ergebnisse für MADRS-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

MADRS-Ansprechen (≥ 50 %-Reduktion von Baseline)	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	172
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	92 (52,9)	62 (36,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (3,4)	4 (2,3)
RR	1,423	
95 %-KI für das RR	[1,125; 1,799]	
p-Wert für das RR	0,0032	
OR	2,046	
95 %-KI für das OR	[1,316; 3,182]	
p-Wert für das OR	0,0015	
ARR	0,170	
95 %-KI für das ARR	[0,069; 0,271]	
p-Wert für das ARR	0,0010	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	106 (60,9)	68 (39,5)

MADRS-Ansprechen (≥ 50 %-Reduktion von Baseline)	**Zuranolon+PiT**	**PiT**
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	8 (4,6)	10 (5,8)
RR	1,528	
95 %-KI für das RR	[1,233; 1,894]	
p-Wert für das RR	0,0001	
OR	2,424	
95 %-KI für das OR	[1,565; 3,754]	
p-Wert für das OR	0,0001	
ARR	0,215	
95 %-KI für das ARR	[0,113; 0,316]	
p-Wert für das ARR	0,0000	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	109 (62,6)	82 (47,7)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	17 (9,8)	19 (11,0)
RR	1,315	
95 %-KI für das RR	[1,086; 1,592]	
p-Wert für das RR	0,0050	
OR	1,850	
95 %-KI für das OR	[1,202; 2,848]	
p-Wert für das OR	0,0052	
ARR	0,151	
95 %-KI für das ARR	[0,048; 0,254]	
p-Wert für das ARR	0,0041	
Heterogenitätstest (p-Wert)	0,6198	

MADRS-Ansprechen (≥ 50 %-Reduktion von Baseline)	Zuranolon+PiT	PiT
[1] Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert des MADRS, der Studie, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Studie sowie der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells. [2] Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

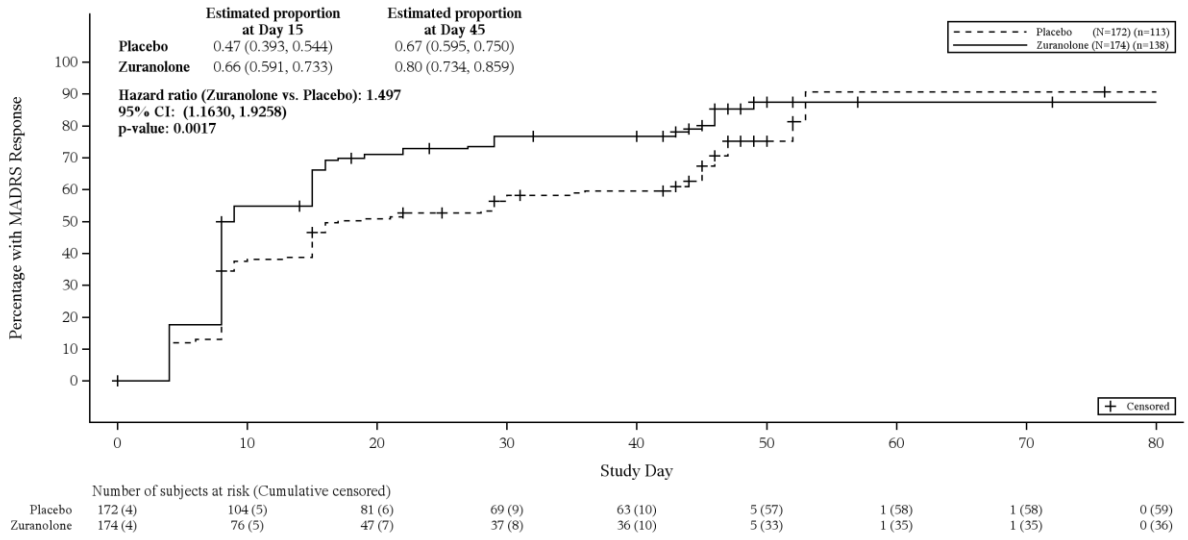
Im Zuranolon+PiT-Arm konnte zu allen Zeitpunkten statistisch signifikant mehr Patientinnen eine Verbesserung von ≥ 50 % gegenüber dem Gesamtwert zu Baseline auf der MADRS-Skala erreichen als unter der Vergleichstherapie.

Tabelle 4-35: Ergebnisse für Zeit bis zum ersten MADRS-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten MADRS-Ansprechen	Zuranolon+PiT	PiT
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	172
Anzahl der Patientinnen mit Antwort, n (%) [2]	138 (79,3)	113 (65,7)
Zeit bis zum ersten MADRS-Ansprechen (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	8,0 [NB; NB]	8,0 [NB; NB]
50. Perzentil - Median (95 %-KI)	8,5 [8,0; 15,0]	17,0 [15,0; 30,0]
75. Perzentil - Q3 (95 %-KI)	29,0 [16,0; 46,0]	47,0 [45,0; NB]
HR - Hazard Ratio [4]	1,497	
95 %-KI für HR	[1,1630; 1,9258]	
p-Wert für HR	0,0017	
Heterogenitätstest (p-Wert)	0,7628	

Zeit bis zum ersten MADRS-Ansprechen	Zuranolon+PiT	PiT
Das Ansprechen nach MADRS ist als Reduktion von ≥ 50 % gegenüber dem Baselinewert definiert. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren MADRS-Bewertung zensiert [1] Anzahl der Patientinnen mit nicht fehlenden MADRS-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit einer MADRS-Ansprechrate. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. [3] Kaplan-Meier-Schätzer. [4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Studie, Behandlung, Baselinewert des MADRS-Gesamtwerts und Antidepressiva zu Baseline (Ja oder Nein). Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Patientinnen im Zuranolon+PiT-Arm sprachen statistisch signifikant schneller (8,5 vs. 17 Tage im Median) auf der MADRS-Skala an als Patientinnen im Vergleichsarm.



[1] Patientinnen mit nicht fehlenden MADRS-Gesamtwerten zu Baseline sind in die Analyse einbezogen. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Das Ansprechen nach MADRS ist definiert als Reduktion von ≥ 50 % gegenüber dem Gesamtwert zu Baseline. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren MADRS-Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Behandlung, Baselinewert des MADRS, Antidepressiva zu Baseline (Ja oder Nein) und Studie angepasst ist.

Abbildung 4-5: Kaplan-Meier-Plot für Zeit bis zum ersten MADRS-Ansprechen

Tabelle 4-36: Ergebnisse für MADRS-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

MADRS-Remission (Gesamtwert ≤ 10)	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	172
Tag 8		
Anzahl der Patientinnen mit Remission, n (%) [2]	50 (28,7)	38 (22,1)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (3,4)	4 (2,3)
RR	1,261	
95 %-KI für das RR	[0,879; 1,809]	
p-Wert für das RR	0,2071	
OR	1,401	
95 %-KI für das OR	[0,853; 2,302]	
p-Wert für das OR	0,1832	
ARR	0,064	
95 %-KI für das ARR	[-0,026; 0,154]	
p-Wert für das ARR	0,1640	
Tag 15		
Anzahl der Patientinnen mit Remission, n (%) [2]	78 (44,8)	45 (26,2)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	8 (4,6)	10 (5,8)
RR	1,690	
95 %-KI für das RR	[1,258; 2,271]	
p-Wert für das RR	0,0005	
OR	2,277	
95 %-KI für das OR	[1,444; 3,592]	
p-Wert für das OR	0,0004	
ARR	0,187	
95 %-KI für das ARR	[0,089; 0,284]	
p-Wert für das ARR	0,0002	

MADRS-Remission (Gesamtwert ≤ 10)	Zuranolon+PiT	PiT
Tag 45		
Anzahl der Patientinnen mit Remission, n (%) [2]	83 (47,7)	58 (33,7)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	17 (9,8)	19 (11,0)
RR	1,409	
95 %-KI für das RR	[1,089; 1,823]	
p-Wert für das RR	0,0091	
OR	1,776	
95 %-KI für das OR	[1,147; 2,751]	
p-Wert für das OR	0,0100	
ARR	0,138	
95 %-KI für das ARR	[0,036; 0,240]	
p-Wert für das ARR	0,0078	
Heterogenitätstest (p-Wert)	0,6219	
[1] Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert des MADRS, der Studie, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Studie sowie der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells. [2] Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Im Zuranolon+PiT-Arm konnten zu Tag 15 und Tag 45 statistisch signifikant mehr Patientinnen eine Remission auf der MADRS-Skala erreichen als unter der Vergleichstherapie.

Tabelle 4-37: Ergebnisse für Zeit bis zur ersten MADRS-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Zeit bis zur ersten MADRS-Remission (Gesamtwert ≤ 10)	Zuranolon+PiT	PiT
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	172
Anzahl der Patienten mit Antwort, n (%) [2]	117 (67,2)	83 (48,3)

Zeit bis zur ersten MADRS-Remission (Gesamtwert ≤ 10)	**Zuranolon+PiT**	**PiT**
Zeit bis zur ersten MADRS-Remission (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	8,0 [8,0; 15,0]	15,0 [8,0; 16,0]
50. Perzentil - Median (95 %-KI)	16,0 [15,0; 28,0]	45,0 [30,0; NB]
75. Perzentil - Q3 (95 %-KI)	46,0 [45,0; NB]	NB [52,0; NB]
HR - Hazard Ratio [4]	1,683	
95 %-KI für HR	[1,2676; 2,2350]	
p-Wert für HR	0,0003	
Heterogenitätstest (p-Wert)	0,3723	

MADRS-Remission ist definiert als ein MADRS-Gesamtwert von ≤ 10. Die Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen werden am Tag der letzten verfügbaren Bewertung zensiert, wenn das interessierende Ereignis nicht beobachtet wird.
[1] Anzahl der Patientinnen mit nicht fehlenden MADRS-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit einer MADRS-Ansprechrate. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

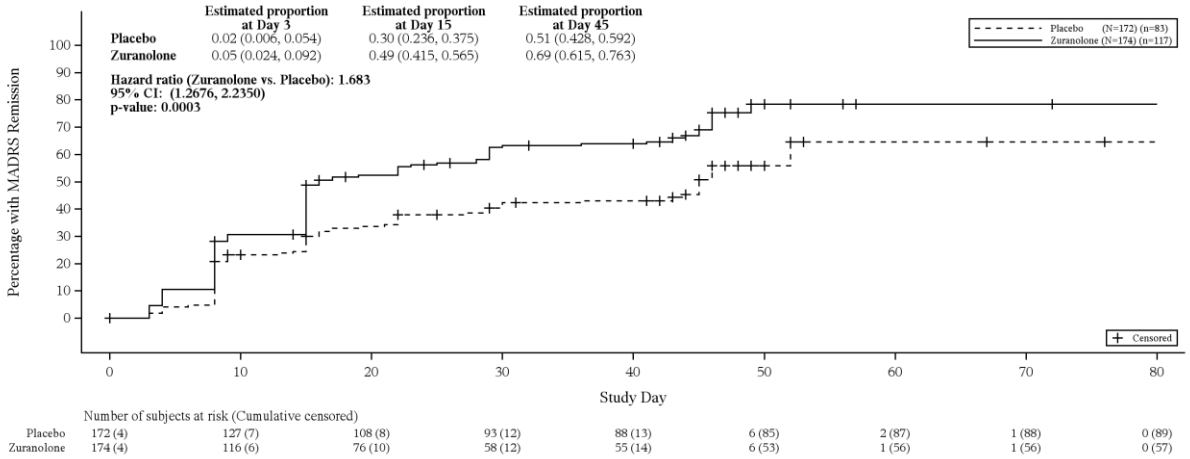
[3] Kaplan-Meier-Schätzer

[4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Studie, Behandlung, Baselinewert des MADRS-Gesamtwerts und Antidepressiva zu Baseline (Ja oder Nein)

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Patientinnen im Zuranolon+PiT-Arm erreichten statistisch signifikant schneller (16 vs. 45 Tage im Median) eine Remission auf der MADRS-Skala als Patientinnen im Vergleichsarm.



[1] Patientinnen mit nicht fehlenden MADRS-Gesamtwerten zu Baseline sind in die Analyse einbezogen. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Die MADRS-Remission ist definiert als MADRS-Gesamtwert ≤ 10. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren MADRS-Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Behandlung, Baselinewert des MADRS, Antidepressiva zu Baseline (Ja oder Nein) und Studie angepasst ist.

Abbildung 4-6: Kaplan-Meier-Plot für Zeit bis zur ersten MADRS-Remission

4.3.1.3.1.2.3 EPDS – RCT

Tabelle 4-38: Operationalisierung von EPDS

Studie	Operationalisierung
ROBIN	Der EPDS ist ein patientenberichteter Fragebogen, der körperliche Symptome ausschließt und 10 Fragen zum emotionalen Zustand der Patientin umfasst, darunter: 1. … konnte ich lachen und das Leben von der sonnigen Seite sehen. 2. … konnte ich mich so richtig auf etwas freuen. 3. … fühlte ich mich unnötigerweise schuldig, wenn etwas schief lief. 4. … war ich ängstlich und besorgt aus nichtigen Gründen. 5. … erschrak ich leicht bzw. reagierte panisch aus unerfindlichen Gründen. 6. … überforderten mich verschiedenste Umstände. 7. … war ich so unglücklich, dass ich nicht schlafen konnte. 8. … habe ich mich traurig und schlecht gefühlt. 9. … war ich so unglücklich, dass ich geweint habe. 10. … überkam mich der Gedanke, mir selbst Schaden zuzufügen. Für jede Frage gibt es vier mögliche Antworten (z. B. „Ja, meistens“, „Ja, manchmal“, „Nein, nicht sehr oft“, „Nein, überhaupt nicht“), um die Häufigkeit des Gefühls anzugeben. Jede Antwort wird mit 0 – 3 Punkten bewertet, und die Punkte für die 10 Fragen werden addiert, um eine Gesamtpunktzahl von 0 – 30 zu erhalten. Je höher die Gesamtpunktzahl, desto schwerer sind die Symptome der PPD. Erhebende Person und Erhebungszeitpunkte Der Fragebogen zur EPDS wurde von den Patientinnen zu folgenden Zeitpunkten ausgefüllt: - Zu den Visiten beim Screening; vor der Verabreichung am Tag 1; 12 ± 1 Stunde nach der abendlichen Verabreichung an den Tagen 3 (nur Amendment 4), 8 und 15; sowie morgens an den Tagen 21 und 45. Schwellenwerte Responder-Analysen wurden *post-hoc* definiert und durchgeführt: - Das Ansprechen/ die Verschlechterung nach MADRS wurde gemäß IQWiG-Methoden als eine Reduktion/ Zunahme um ≥ 15 % der Skalenspannweite, in diesem Fall um ≥ 4,5 Punkte, festgelegt. Dargestellte Analysen: Folgende Analysen werden für die ITT Population der Studie ROBIN in Anhang 4-H dargestellt: - Veränderung des EPDS-Gesamtwertes im Vergleich zu Baseline (inkl. grafische Darstellung über die Zeit). - Anzahl der Patientinnen mit EPDS-Ansprechen - Zeit bis zum ersten EPDS-Ansprechen (inkl. Kaplan-Meier-Plot) - Anzahl der Patientinnen mit EPDS-Verschlechterung - Zeit bis zur ersten EPDS-Verschlechterung (inkl. Kaplan-Meier-Plot)
SKY-LARK	Die Definition des Endpunkts EPDS entspricht der Definition in der Studie ROBIN. Erhebende Person und Erhebungszeitpunkte Der Fragebogen zur EPDS wurde von den Patientinnen zu folgenden Zeitpunkten ausgefüllt: - Zum Screening und zu den Visiten an Tag 1, 3, 8, 15, 21, 28 und 45. Schwellenwerte

	Responder-Analysen wurden, wie in der Beschreibung für ROBIN *post-hoc* definiert und durchgeführt. Dargestellte Analysen: Folgende Analysen werden für die ITT-Population der Studie SKYLARK in Anhang 4-I dargestellt: - Siehe ROBIN.
Meta-Analyse	Die Definition des Endpunkts EPDS entspricht der Definition in der Studie ROBIN. Aufgrund den leicht abweichenden Erhebungszeitpunkten wurden die die Daten der folgenden Erhebungszeitpunkte gepoolt: Tag 3, 8, 15, 21 und Tag 45. Schwellenwerte Da keine Responder Analysen in den jeweiligen Studienunterlagen von ROBIN oder SKYLARK prädefiniert waren, werden folgende *post-hoc* Responderanalysen durchgeführt. Dargestellte Analysen: Folgende Analysen werden für die gepoolte ITT-Population beider Studien nachfolgend dargestellt: - Anzahl der Patientinnen mit EPDS-Ansprechen - Zeit bis zum ersten EPDS-Ansprechen (inkl. Kaplan-Meier-Plot) Folgende Analysen werden für die gepoolte ITT-Population beider Studien supportiv im Anhang 4-G dargestellt: - Veränderung des EPDS-Gesamtwertes im Vergleich zu Baseline (inkl. grafische Darstellung über die Zeit) - Anzahl der Patientinnen mit EPDS-Verschlechterung Zeit bis zur ersten EPDS-Verschlechterung (inkl. Kaplan-Meier-Plot) Dazugehörige Statistische Methoden: Eine ausführliche Beschreibung der Methodik, die auf dem SKYLARK SAP beruht, findet sich in Abschnitt 4.2.5 bzw. in den Fußnoten. Zudem wurden Heterogenitätstests durchgeführt, die *post-hoc* definiert wurden und deren methodische Grundlage im gleichen Abschnitt zu finden sind.
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.	

Tabelle 4-39: Bewertung des Verzerrungspotenzials für EPDS in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
ROBIN	niedrig	ja	ja	ja	ja	niedrig
SKYLARK	niedrig	ja	ja	ja	ja	niedrig

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das

Studienteam verblindet gegenüber der Behandlungszuweisung. Der selbstberichtete Endpunkt EPDS wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Zusätzlich wurden *post-hoc* Analysen ergänzt, um den Anforderungen an die Nutzenbewertung zu entsprechen. Abgesehen von diesem Zusatz wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Tabelle 4-40: Rücklaufrate für EPDS aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

EPDS-Rücklaufrate	**Zuranolon+PiT N=176**	**PiT N=177**
Tag 3	169 (96 %)	166 (94 %)
Tag 8	168 (95 %)	169 (95 %)
Tag 15	165 (94 %)	162 (92 %)
Tag 21	158 (90 %)	154 (87 %)
Tag 45	157 (89 %)	154 (87 %)
Werte sind aus den MMRM Analysen entnommen (s. Anhang 4-G) Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.		

Die Rücklaufraten für den EPDS lagen zu jedem Zeitpunkt über 87 %.

Tabelle 4-41: Ergebnisse für EPDS-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

EPDS-Ansprechen (Reduktion um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	173
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	56 (32,2)	39 (22,5)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	5 (2,9)	7 (4,0)
RR	1,454	
95 %-KI für das RR	[1,031; 2,049]	
p-Wert für das RR	0,0326	

EPDS-Ansprechen (Reduktion um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
OR	1,550	
95 %-KI für das OR	[0,962; 2,498]	
p-Wert für das OR	0,0716	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	111 (63,8)	101 (58,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (3,4)	4 (2,3)
RR	1,112	
95 %-KI für das RR	[0,946; 1,306]	
p-Wert für das RR	0,1986	
OR	1,224	
95 %-KI für das OR	[0,790; 1,898]	
p-Wert für das OR	0,3663	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	127 (73,0)	108 (62,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	9 (5,2)	11 (6,4)
RR	1,169	
95 %-KI für das RR	[1,018; 1,342]	
p-Wert für das RR	0,0272	
OR	1,597	
95 %-KI für das OR	[1,011; 2,524]	
p-Wert für das OR	0,0449	

EPDS-Ansprechen (Reduktion um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	129 (74,1)	110 (63,6)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	16 (9,2)	19 (11,0)
RR	1,111	
95 %-KI für das RR	[0,963; 1,281]	
p-Wert für das RR	0,1502	
OR	1,603	
95 %-KI für das OR	[0,995; 2,582]	
p-Wert für das OR	0,0525	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	133 (76,4)	111 (64,2)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	17 (9,8)	19 (11,0)
RR	1,162	
95 %-KI für das RR	[1,017; 1,328]	
p-Wert für das RR	0,0269	
OR	1,770	
95 %-KI für das OR	[1,096; 2,857]	
p-Wert für das OR	0,0195	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	

EPDS-Ansprechen (Reduktion um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
Heterogenitätstest (p-Wert)	0,2396	
[1] Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert des EPDS, der Studie, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Studie sowie der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells. [2] Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

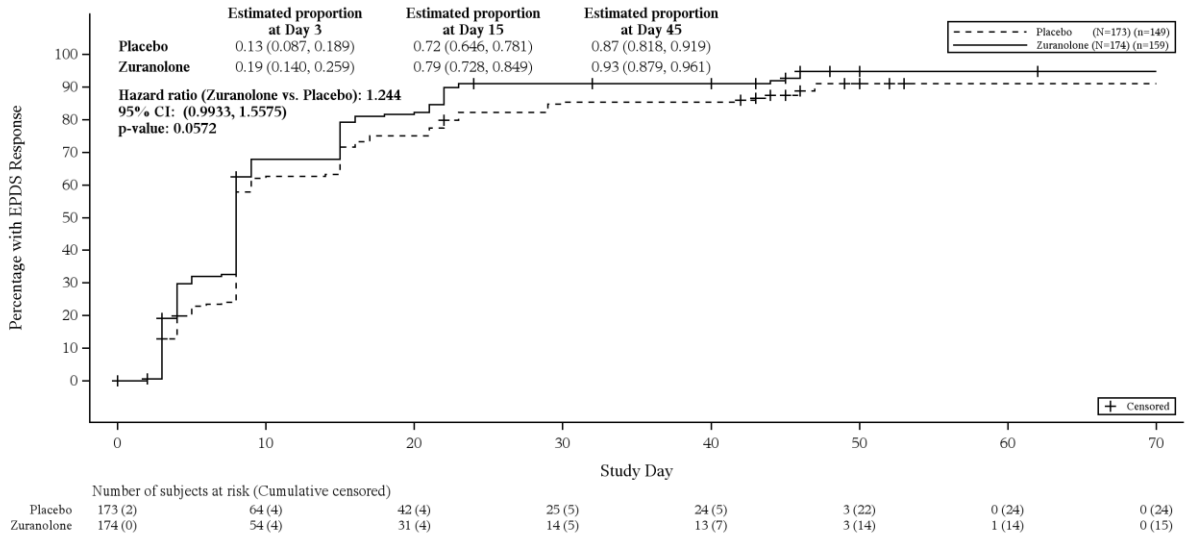
Im Zuranolon+PiT-Arm konnten an den Tagen 3, 15 und 45 statistisch signifikant mehr Patientinnen eine relevante Verbesserung auf der EPDS-Skala erreichen als unter der Vergleichstherapie.

Tabelle 4-42: Ergebnisse für Zeit bis zum ersten EPDS-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten EPDS-Ansprechen	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	173
Anzahl der Patientinnen mit Antwort, n (%) [2]	159 (91,4)	149 (86,1)
Zeit bis zum ersten EPDS-Ansprechen (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	4,0 [3,0; 5,0]	8,0 [4,0; 8,0]
50. Perzentil - Median (95 %-KI)	8,0 [NB; NB]	8,0 [NB; NB]
75. Perzentil - Q3 (95 %-KI)	15,0 [15,0; 18,0]	17,0 [15,0; 23,0]
HR - Hazard Ratio [4]	1,244	
95 %-KI für HR	[0,9933; 1,5575]	
p-Wert für HR	0,0572	
Heterogenitätstest (p-Wert)	0,2211	

Zeit bis zum ersten EPDS-Ansprechen	Zuranolon+PiT	PiT
Das Ansprechen nach EPDS ist definiert als Reduktion um ≥ 4,5 Punkte (15 % der EPDS-Skalenspannweite von 0 – 30) des Gesamtwerts von Baseline. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren EPDS-Bewertung zensiert. [1] Anzahl der Patientinnen mit nicht fehlenden EPDS-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit einer EPDS-Ansprechrate. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. [3] Kaplan-Meier-Schätzer [4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Studie, Behandlung, Baselinewert des EPDS-Gesamtwerts und Antidepressiva zu Baseline (Ja oder Nein) Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Die Zeit bis zum ersten EPDS-Ansprechen zeigte keinen statistisch signifikanten Unterschied zwischen den Behandlungsarmen.



[1] Patientinnen mit nicht fehlenden EPDS-Gesamtwerten zu Baseline sind in die Analyse einbezogen. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Das Ansprechen nach EPDS ist definiert als Reduktion um ≥ 4,5 Punkte (15 % der EPDS-Skalenspannweite von 0 – 30) des Gesamtwerts von Baseline. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren EPDS-Bewertung zensiert.

[3]: Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Behandlung, Baselinewert des EPDS, Antidepressiva zu Baseline (Ja oder Nein) und Studie angepasst ist.

Abbildung 4-7: Kaplan-Meier-Plot für Zeit bis zum ersten EPDS-Ansprechen

4.3.1.3.1.2.4 PHQ-9 – RCT

Tabelle 4-43: Operationalisierung von PHQ-9

Studie	Operationalisierung
ROBIN	Der PHQ-9 ist ein selbst auszufüllender Fragebogen, der den Schweregrad depressiver Symptome bewertet und 9 Items zum emotionalen Zustand des Patienten enthält, darunter: 1. Wenig Interesse oder Freude an Ihren Tätigkeiten. 2. Niedergeschlagenheit, Schwermut oder Hoffnungslosigkeit. 3. Schwierigkeiten, ein- oder durchzuschlafen oder vermehrter Schlaf. 4. Müdigkeit oder Gefühl, keine Energie zu haben. 5. Verminderter Appetit oder übermäßiges Bedürfnis zu essen. 6. Schlechte Meinung von sich selbst; Gefühl, ein Versager zu sein oder die Familie enttäuscht zu haben. 7. Schwierigkeiten, sich auf etwas zu konzentrieren, z. B. beim Zeitunglesen oder Fernsehen. 8. Waren Ihre Bewegungen oder Ihre Sprache so verlangsamt, dass es auch anderen auffallen würde? Oder waren Sie im Gegenteil „zappelig“ oder ruhelos und hatten dadurch einen stärkeren Bewegungsdrang als sonst? 9. Gedanken, dass Sie lieber tot wären oder sich Leid zufügen möchten. Für jede Frage gibt es vier mögliche Antworten (z. B. „Überhaupt nicht“, „An manchen Tagen“, „An mehr als der Hälfte der Tage“ oder „Fast jeden Tag“), um die Häufigkeit des Gefühls anzugeben. Jede Antwort wird mit 0 – 3 Punkten bewertet, und die Punkte für die 9 Fragen werden addiert, um eine Gesamtpunktzahl von 0 – 27 zu erhalten. Je höher die Gesamtpunktzahl, desto schwerwiegender sind die Symptome der Depression. Das Gesamtergebnis aller 9 Items wird wie folgt bewertet: - 1 – 4 = minimale Depression - 5 – 9 = milde Depression - 10 – 14 = mittlere Depression - 15 – 19 = mittelschwere Depression - 20 – 27 = schwere Depression. <u>Erhebende Person und Erhebungszeitpunkte</u> Der Fragebogen zum PHQ-9 wurde von den Patientinnen zu folgenden Zeitpunkten ausgefüllt: - zu den Visiten beim Screening; vor der Verabreichung am Tag 1; 12 ± 1 Stunde nach der abendlichen Verabreichung an den Tagen 3 (nur PA4), 8 und 15; sowie morgens an den Tagen 21 und 45. <u>Schwellenwerte</u> Responder-Analysen wurden *post-hoc* definiert und durchgeführt: - Das Ansprechen/ die Verschlechterung nach PHQ-9 wurde gemäß IQWiG-Methoden als eine Reduktion/ Zunahme um ≥ 15 % der Skalenspannweite, in diesem Fall um ≥ 4,05 Punkte, festgelegt. <u>Dargestellte Analysen:</u> Folgende Analysen werden für die ITT Population der Studie ROBIN in Anhang 4-H dargestellt: - Veränderung des PHQ-9-Gesamtwertes im Vergleich zu Baseline (inkl. grafische Darstellung über die Zeit). - Anzahl der Patientinnen mit PHQ-9-Ansprechen - Zeit bis zum ersten PHQ-9-Ansprechen(inkl. Kaplan-Meier-Plot)

	- Anzahl der Patientinnen mit einer PHQ-9-Verschlechterung - Zeit bis zur ersten PHQ-9-Verschlechterung (inkl. Kaplan-Meier-Plot)
SKY-LARK	Die Definition des Endpunkts PHQ-9 entspricht der Definition in der Studie ROBIN. Erhebende Person und Erhebungszeitpunkte Die PHQ-9 wurde von den Patientinnen selbst ausgefüllt, die Fragebögen wurden zu folgenden Zeitpunkten in der Klinik eingesammelt: - Zum Screening und zu den Visiten an Tag 1, 3, 8, 15, 21, 28 und 45. Schwellenwerte Responder-Analysen wurden, wie in der Beschreibung für ROBIN *post-hoc* definiert und durchgeführt. Dargestellte Analysen: Folgende Analysen werden für die ITT-Population der Studie SKYLARK in Anhang 4-I dargestellt: - Siehe ROBIN. Fehlende Werte: Wenn mehr als ein einzelnes Item fehlt, wird die PHQ-9-Gesamtpunktzahl nicht berechnet und als fehlend belassen. Wenn nur ein einzelner Punkt fehlt, werden die fehlenden Punktwerte durch den Mittelwert aller anderen verfügbaren Punktwerte ersetzt. Dies war für den PHQ-9 bei 3 Patienten der Fall.
Meta-Analyse	Die Definition des Endpunkts PHQ-9 entspricht der Definition in der Studie ROBIN. Aufgrund den leicht abweichenden Erhebungszeitpunkten wurden die die Daten der folgenden Erhebungszeitpunkte gepoolt: Tag 3, 8, 15, 21 und Tag 45. Dargestellte Analysen: Folgende Analysen werden für die gepoolte ITT-Population beider Studien nachfolgend dargestellt: - Anzahl der Patientinnen mit PHQ-9-Ansprechen - Zeit bis zum ersten PHQ-9-Ansprechen(inkl. Kaplan-Meier-Plot) Folgende Analysen werden für die gepoolte ITT-Population beider Studien supportiv im Anhang 4-G dargestellt: - Veränderung des PHQ-9-Gesamtwertes im Vergleich zu Baseline (inkl. grafische Darstellung über die Zeit). - Anzahl der Patientinnen mit einer PHQ-9-Verschlechterung Zeit bis zur ersten PHQ-9-Verschlechterung (inkl. Kaplan-Meier-Plot) Dazugehörige Statistische Methoden: Eine ausführliche Beschreibung der Methodik, die auf dem SKYLARK SAP beruht, findet sich in Abschnitt 4.2.5 bzw. in den Fußnoten. Zudem wurden Heterogenitätstests durchgeführt, die *post-hoc* definiert wurden und deren methodische Grundlage im gleichen Abschnitt zu finden sind.
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.	

Tabelle 4-44: Bewertung des Verzerrungspotenzials für PHQ-9 in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
ROBIN	niedrig	ja	ja	ja	ja	niedrig
SKYLARK	niedrig	ja	ja	ja	ja	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der selbstberichtete Endpunkt PHQ-9 wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Zusätzlich wurden *post-hoc* Analysen ergänzt, um den Anforderungen an die Nutzenbewertung zu entsprechen. Abgesehen von diesem Zusatz wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Tabelle 4-45: Rücklaufrate für PHQ-9 aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

PHQ-9-Rücklaufrate	Zuranolon+PiT N=176	PiT N=177
Tag 3	169 (96 %)	166 (94 %)
Tag 8	168 (95 %)	169 (95 %)
Tag 15	165 (94 %)	163 (92 %)
Tag 21	158 (90 %)	154 (87 %)
Tag 45	157 (89 %)	154 (87 %)
Werte sind aus den MMRM Analysen entnommen (s. Anhang 4-G) Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.		

Die Rücklaufraten für den PHQ-9 lagen zu jedem Zeitpunkt über 87 %.

Tabelle 4-46: Ergebnisse für PHQ-9-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

PHQ-9-Ansprechen (Reduktion um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	173
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	45 (25,9)	42 (24,3)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	5 (2,9)	7 (4,0)
RR	1,080	
95 %-KI für das RR	[0,758; 1,540]	
p-Wert für das RR	0,6705	
OR	1,062	
95 %-KI für das OR	[0,652; 1,728]	
p-Wert für das OR	0,8096	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	112 (64,4)	86 (49,7)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (3,4)	4 (2,3)
RR	1,317	
95 %-KI für das RR	[1,103; 1,571]	
p-Wert für das RR	0,0023	
OR	1,893	
95 %-KI für das OR	[1,229; 2,916]	
p-Wert für das OR	0,0038	
ARR	NB	
95 %-KI für das ARR	NB	

PHQ-9-Ansprechen (Reduktion um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
p-Wert für das ARR	NB	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	130 (74,7)	109 (63,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	9 (5,2)	10 (5,8)
RR	1,177	
95 %-KI für das RR	[1,025; 1,352]	
p-Wert für das RR	0,0209	
OR	1,803	
95 %-KI für das OR	[1,118; 2,905]	
p-Wert für das OR	0,0156	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	124 (71,3)	109 (63,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	16 (9,2)	19 (11,0)
RR	1,099	
95 %-KI für das RR	[0,952; 1,269]	
p-Wert für das RR	0,1984	
OR	1,496	
95 %-KI für das OR	[0,933; 2,398]	
p-Wert für das OR	0,0945	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	131 (75,3)	111 (64,2)

PHQ-9-Ansprechen (Reduktion um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	17 (9,8)	19 (11,0)
RR	1,148	
95 %-KI für das RR	[1,001; 1,315]	
p-Wert für das RR	0,0480	
OR	1,750	
95 %-KI für das OR	[1,076; 2,845]	
p-Wert für das OR	0,0240	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Heterogenitätstest (p-Wert)	0,4204	

[1] Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert des PHQ-9, der Studie, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Studie sowie der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells.

[2] Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Im Zuranolon+PiT-Arm konnten an den Tagen 8, 15 und 45 statistisch signifikant mehr Patientinnen eine relevante Verbesserung auf der PHQ-9-Skala erreichen als unter der Vergleichstherapie.

Tabelle 4-47: Ergebnisse für Zeit bis zum ersten PHQ-9-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten PHQ-9-Ansprechen PHQ-9	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	173
Anzahl der Patientinnen mit Antwort, n (%) [2]	161 (92,5)	149 (86,1)
Zeit bis zur ersten PHQ-9-Ansprechen (Tage) [3]		

Zeit bis zum ersten PHQ-9-Ansprechen PHQ-9	**Zuranolon+PiT**	**PiT**
25. Perzentil - Q1 (95 %-KI)	5,0 [4,0; 8,0]	7,0 [4,0; 8,0]
50. Perzentil - Median (95 %-KI)	8,0 [NB; NB]	8,0 [8,0; 15,0]
75. Perzentil - Q3 (95 %-KI)	15,0 [NB; NB]	22,0 [16,0; 23,0]
HR - Hazard Ratio [4]	1,306	
95 %-KI für HR	[1,0426; 1,6348]	
p-Wert für HR	0,0201	
Heterogenitätstest (p-Wert)	0,1901	

Das Ansprechen nach PHQ-9 ist definiert als Reduktion um ≥ 4,05 Punkte (15 % der PHQ-9-Skalenspannweite von 0 – 27) des Gesamtwerts von Baseline. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren PHQ-9-Bewertung zensiert.

[1] Anzahl der Patientinnen mit nicht fehlenden PHQ-9-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit einer PHQ-9-Ansprechrate. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

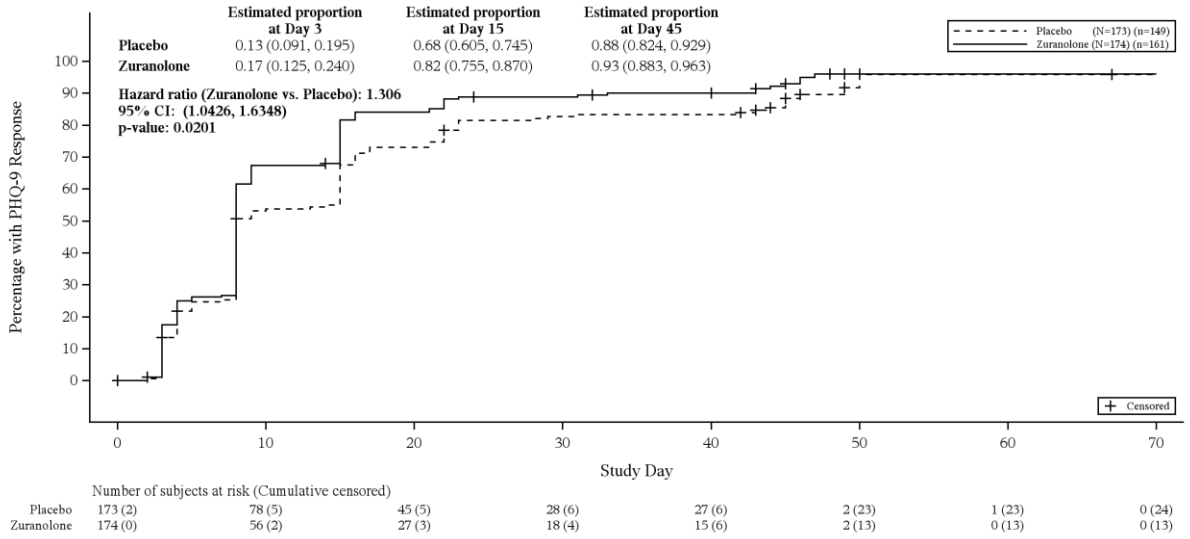
[3] Kaplan-Meier-Schätzer.

[4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Studie, Behandlung, Baselinewert des PHQ-9-Gesamtwerts und Antidepressiva zu Baseline (Ja oder Nein).

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Patientinnen im Zuranolon+PiT-Arm erreichten statistisch signifikant schneller ein Ansprechen auf der PHQ-9-Skala als Patientinnen im Vergleichsarm.



[1] Patientinnen mit nicht fehlenden PHQ-9-Gesamtwerten zu Baseline sind in die Analyse einbezogen. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Das Ansprechen nach PHQ-9 ist definiert als Reduktion um ≥ 4,05 Punkte (15 % der PHQ-9-Skalenspannweite von 0 – 27) des Gesamtwerts von Baseline. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren PHQ-9-Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Behandlung, Baselinewert des PHQ-9, Antidepressiva zu Baseline (Ja oder Nein) und Studie angepasst ist.

Abbildung 4-8: Kaplan-Meier-Plot für Zeit bis zum ersten PHQ-9-Ansprechen

4.3.1.3.1.2.5 HAM-A – RCT

Tabelle 4-48: Operationalisierung von HAM-A

Studie	Operationalisierung
ROBIN	Die HAM-A besteht aus 14 Items und dient zur Beurteilung des Schweregrads von psychischen und somatischen Angstsymptomen, wobei die folgenden 14 Symptome abgedeckt werden: - Ängstliche Stimmung: Sorgen, Erwartung des Schlimmsten, ängstliche Vorahnung, Reizbarkeit. - Anspannung: Gefühle der Anspannung, Ermüdbarkeit, Unfähigkeit sich zu entspannen Schreckhaftigkeit, leichte Tränen, Zittern, Unruhegefühle. - Ängste: Vor Dunkelheit, vor Fremden, vor dem Alleingelassen werden, vor großen Tieren, vor Verkehr, vor Menschenmengen. - Schlaflosigkeit: Einschlafschwierigkeiten, unterbrochener Schlaf, unbefriedigender Schlaf und Müdigkeit beim Aufwachen, Träume, Albträume, Nachtangst. - Intellektuell: Konzentrationsschwierigkeiten, schlechtes Gedächtnis. - Depressive Verstimmung: Interessenverlust, keine Freude mehr an Hobbys, Depressionen, frühes Erwachen, Tagesverlaufsschwankungen. - Somatisch (muskulär): muskuläre Schmerzen und Beschwerden, muskuläre Steifheit, muskuläre Zuckungen, krampfhafte Zuckungen, Zähneknirschen, unsichere Stimme, erhöhter Muskeltonus. - Somatisch (sensorisch): Tinnitus, verschwommenes Sehen, Hitzewallungen und Kälteschauer, Schwächegefühl, Kribbeln. - Herz-Kreislauf-Symptome: Tachykardie, Herzklopfen, Brustschmerzen, Pulsieren der Gefäße, Ohnmachtsgefühl, Herzaussetzer. - Atemwegsbeschwerden: Druck oder Engegefühl in der Brust, Erstickungsgefühl, Seufzen, Atemnot. - Magen-Darm-Symptome: Schwierigkeiten beim Schlucken, Blähungen, Schmerzen vor und nach den Mahlzeiten, Brennendes Gefühl, Völlegefühl, Sodbrennen, Übelkeit, Erbrechen, Gefühl des Absinkens, „Arbeiten“ im Bauch, Borborygmi (Darmgeräusche), Lockerheit des Stuhls, Gewichtsverlust, Verstopfung. - Urogenitale Symptome: Häufiges Wasserlassen, Harndrang, Amenorrhö, Menorrhagie, Entwicklung von Frigidität, vorzeitige Ejakulation, Libidoverlust, Impotenz. - Vegetative Symptome: Mundtrockenheit, Hitzewallungen, Blässe, Neigung zum Schwitzen, Schwindel, Spannungskopfschmerzen, Haarausfall. - Verhalten während des Gesprächs: Zappeln, Unruhe oder Auf- und Abgehen, Zittern der Hände, gerunzelte Stirn, angespanntes Gesicht, Seufzen oder schnelle Atmung, Blässe im Gesicht, Schlucken usw. Jede Antwort wird mit 0 Punkten (nicht vorhanden) bis 4 Punkten (sehr stark) bewertet, und die Punkte für die 14 Fragen werden addiert, um eine Gesamtpunktzahl von 0 – 56 zu erhalten. Je höher die Gesamtpunktzahl, desto stärker sind die Angstsymptome. Die Gesamtpunktzahl für alle 14 Teile wird wie folgt bewertet: • < 17 = leichter Schweregrad • 18 – 24 = leichter bis mittelschwerer Schweregrad • 25 – 30 = mittelschwerer bis schwerer Schweregrad • > 30 = schwerer Schweregrad Erhebende Person und Erhebungszeitpunkte

	Die HAM-A wurde von dem Prüfpersonal zu folgenden Zeitpunkten durchgeführt: - vor der Verabreichung am Tag 1; 12 ± 1 Stunde nach der abendlichen Verabreichung an den Tagen 3, 8 und 15; und morgens an den Tagen 21 und 45. Schwellenwerte Responder-Analysen wurden *post-hoc* definiert und durchgeführt: - Das Ansprechen nach HAM-A wurde gemäß IQWiG-Methoden als eine Reduktion um ≥ 15 % der Skalenspannweite, in diesem Fall um ≥ 8,4 Punkte, festgelegt. Dargestellte Analysen: Folgende Analysen werden für die ITT Population der Studie ROBIN in Anhang 4-H dargestellt: - Anzahl der Patientinnen mit HAM-A-Ansprechen - Zeit bis zum ersten HAM-A-Ansprechen (inkl. Kaplan-Meier-Plot)
SKY-LARK	Die Definition des Endpunkts HAM-A entspricht der Definition in der ROBIN-Studie. Erhebende Person und Erhebungszeitpunkte Die HAM-A wurde von dem Prüfpersonal zu folgenden Zeitpunkten durchgeführt: - an den Tagen 1, 3, 8, 15, 21, 28 und 45. Schwellenwerte Responder-Analysen wurden *post-hoc* definiert und durchgeführt: - Das Ansprechen nach HAM-A wurde gemäß IQWiG-Methoden als eine Reduktion um ≥ 15 % der Skalenspannweite, in diesem Fall um ≥ 8,4 Punkte, festgelegt. Dargestellte Analysen: Folgende Analysen werden für die ITT-Population der Studie SKYLARK in Anhang 4-I dargestellt: - Siehe ROBIN
Meta-Analyse	Die Definition des Endpunkts HAM-A entspricht der Definition in der Studie ROBIN. Aufgrund den leicht abweichenden Erhebungszeitpunkten wurden die die Daten der folgenden Erhebungszeitpunkte gepoolt: Tag 3, 8, 15, 21 und Tag 45. Dargestellte Analysen: Folgende Analysen werden für die gepoolte ITT-Population beider Studien nachfolgend dargestellt: - Anzahl der Patientinnen mit HAM-A-Ansprechen Zeit bis zum ersten HAM-A-Ansprechen (inkl. Kaplan-Meier-Plot) Dazugehörige Statistische Methoden: Eine ausführliche Beschreibung der Methodik, die auf dem SKYLARK SAP beruht, findet sich in Abschnitt 4.2.5 bzw. in den Fußnoten. Zudem wurden Heterogenitätstests durchgeführt, die *post-hoc* definiert wurden und deren methodische Grundlage im gleichen Abschnitt zu finden sind.
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.	

Tabelle 4-49: Bewertung des Verzerrungspotenzials für HAM-A in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
ROBIN	niedrig	ja	ja	ja	ja	niedrig
SKYLARK	niedrig	ja	ja	ja	ja	niedrig

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der Endpunkt HAM-A wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben, um die Angstzustände der Patientinnen zu dokumentieren. Zusätzlich wurden *post-hoc* Analysen ergänzt, um den Anforderungen an die Nutzenbewertung zu entsprechen. Abgesehen von diesem Zusatz wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet. Das ITT-Prinzip wurde adäquat umgesetzt. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Tabelle 4-50: Rücklaufrate für HAM-A aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

HAM-A-Rücklaufrate	Zuranolon+PiT N=176	PiT N=177
Tag 3	172 (98 %)	170 (96 %)
Tag 8	168 (95 %)	169 (95 %)
Tag 15	166 (94 %)	163 (92 %)
Tag 21	158 (90 %)	156 (88 %)
Tag 45	157 (89 %)	154 (87 %)
Werte sind aus den MMRM Analysen entnommen (s. Anhang 4-G) Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.		

Die Rücklaufraten für die HAM-A lagen zu jedem Zeitpunkt über 87 %.

Tabelle 4-51: Ergebnisse für HAM-A-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

HAM-A-Ansprechen (Reduktion um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	173
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	75 (43,1)	58 (33,5)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (1,1)	3 (1,7)
RR	1,304	
95 %-KI für das RR	[1,009; 1,686]	
p-Wert für das RR	0,0423	
OR	1,575	
95 %-KI für das OR	[1,014; 2,446]	
p-Wert für das OR	0,0432	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	112 (64,4)	86 (49,7)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (3,4)	4 (2,3)
RR	1,312	
95 %-KI für das RR	[1,101; 1,564]	
p-Wert für das RR	0,0024	
OR	1,975	
95 %-KI für das OR	[1,270; 3,073]	
p-Wert für das OR	0,0025	
ARR	NB	
95 %-KI für das ARR	NB	

HAM-A-Ansprechen (Reduktion um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
p-Wert für das ARR	NB	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	125 (71,8)	95 (54,9)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	8 (4,6)	10 (5,8)
RR	1,323	
95 %-KI für das RR	[1,129; 1,552]	
p-Wert für das RR	0,0006	
OR	2,236	
95 %-KI für das OR	[1,423; 3,513]	
p-Wert für das OR	0,0005	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	116 (66,7)	102 (59,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	16 (9,2)	17 (9,8)
RR	1,144	
95 %-KI für das RR	[0,975; 1,343]	
p-Wert für das RR	0,1000	
OR	1,453	
95 %-KI für das OR	[0,923; 2,287]	
p-Wert für das OR	0,1061	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	124 (71,3)	106 (61,3)

HAM-A-Ansprechen (Reduktion um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	17 (9,8)	19 (11,0)
RR	1,178	
95 %-KI für das RR	[1,014; 1,369]	
p-Wert für das RR	0,0322	
OR	1,664	
95 %-KI für das OR	[1,041; 2,660]	
p-Wert für das OR	0,0333	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Heterogenitätstest (p-Wert)	0,6030	
[1] Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert des HAM-A, der Studie, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Studie sowie der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells. [2] Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Im Zuranolon+PiT-Arm konnten an den Tagen 3, 8, 15 und 45 statistisch signifikant mehr Patientinnen eine relevante Verbesserung auf der HAM-A-Skala erreichen als unter der Vergleichstherapie.

Tabelle 4-52: Ergebnisse für Zeit bis zum ersten HAM-A-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten HAM-A- Ansprechen	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 177**
Anzahl der auswertbaren Patientinnen [1]	174	173
Anzahl der Patientinnen mit Antwort, n (%) [2]	152 (87,4)	140 (80,9)

Zeit bis zum ersten HAM-A- Ansprechen	Zuranolon+PiT	PiT
Zeit bis zum ersten HAM-A-Ansprechen (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	3,0 [3,0; 4,0]	4,0 [3,0; 4,0]
50. Perzentil - Median (95 %-KI)	8,0 [5,0; 8,0]	8,0 [8,0; 14,0]
75. Perzentil - Q3 (95 %-KI)	15,0 [15,0; 21,0]	22,0 [20,0; 44,0]
HR - Hazard Ratio [4]	1,356	
95 %-KI für HR	[1,0738; 1,7128]	
p-Wert für HR	0,0105	
Heterogenitätstest (p-Wert)	0,7530	

Das Ansprechen nach HAM-A ist definiert als Reduktion um ≥ 8,4 Punkte (15 % der HAM-A-Skalenspannweite von 0 – 56) des Gesamtwerts von Baseline. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren HAM-A-Bewertung zensiert.

[1] Anzahl der Patientinnen mit nicht fehlenden HAM-A-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit einer HAM-A-Ansprechrate. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

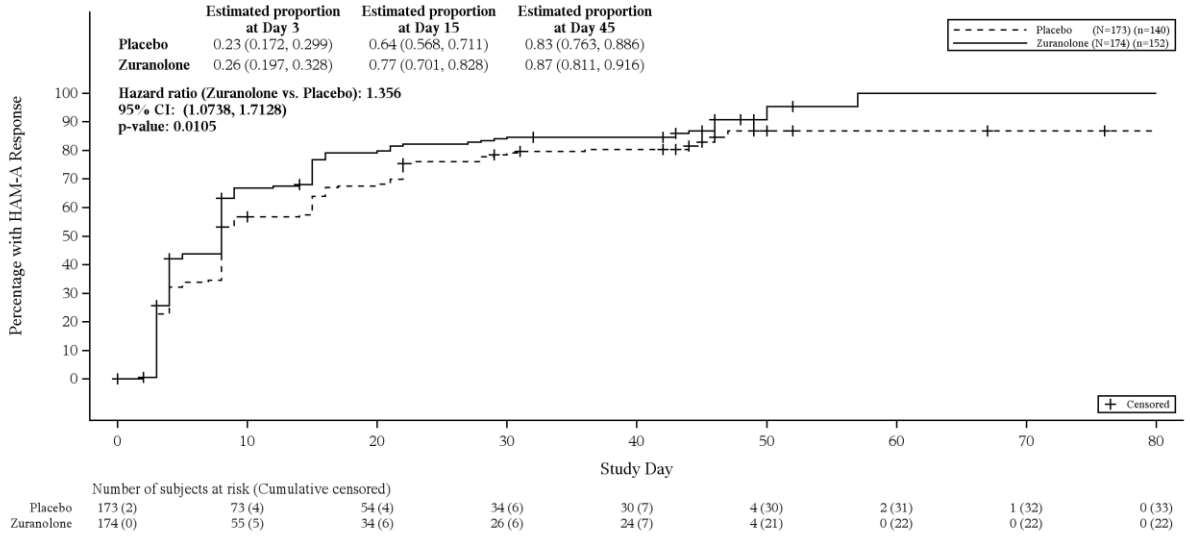
[3] Kaplan-Meier-Schätzer

[4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Studie, Behandlung, Baselinewert des HAM-A-Gesamtwerts und Antidepressiva zu Baseline (Ja oder Nein)

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Patientinnen im Zuranolon+PiT-Arm sprachen statistisch signifikant schneller auf der HAM-A-Skala an als Patientinnen im Vergleichsarm.



[1] Patientinnen mit nicht fehlenden HAM-A-Gesamtwerten zu Baseline sind in die Analyse einbezogen. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Das Ansprechen nach HAM-A ist definiert als Reduktion um ≥ 8,4 Punkte (15 % der HAM-A-Skalenspannweite von 0 – 56) des Gesamtwerts von Baseline. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren HAM-A-Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Behandlung, Baselinewert des HAM-A, Antidepressiva zu Baseline (Ja oder Nein) und Studie angepasst ist.

Abbildung 4-9: Kaplan-Meier-Plot für Zeit bis zum ersten HAM-A-Ansprechen

4.3.1.3.1.3 Lebensqualität – RCT

4.3.1.3.1.3.1 SF-36 – RCT

Tabelle 4-53: Operationalisierung von SF-36

Studie	Operationalisierung
ROBIN	Der SF-36 ist ein aus 36 Fragen bestehender, multidimensionaler, validierter und generischer Lebensqualitätsfragebogen, der von Patient*innen selbst ausgefüllt wird. Der SF-36 besteht aus 8 Domänen die wiederum in unterschiedlichen Summenscores, dem PCS und MCS zusammengefasst werden: - PCS: 1. körperliche Funktionsfähigkeit, 2. körperliche Rollenfunktion, 3. körperliche Schmerzen, 4. allgemeine Gesundheitswahrnehmung, - MCS 5. Vitalität, 6. soziale Funktionsfähigkeit, 7. emotionale Rollenfunktion, 8. und psychisches Wohlbefinden. Je höher der SF-36-Wert, desto besser ist die Lebensqualität. <u>Erhebende Person und Erhebungszeitpunkte</u> Der SF-36 wurde von den Patientinnen zu folgenden Zeitpunkten selbst ausgefüllt: - Während der Screening-Besuche; vor der Verabreichung am Tag 1; am Tag 3 (nur PA4); an Tage 15 – 12 ± 1 Stunden nach der Verabreichung am Vorabend; und am Morgen des Tages 45. <u>Schwellenwerte</u> Responder-Analysen wurden *post-hoc* definiert und durchgeführt: - Das Ansprechen im PCS bzw. MCS nach SF-36 wurde gemäß IQWiG-Methoden als eine Reduktion um ≥ 15 % der Skalenspannweite, entsprechend ≥ 9,4 Punkte für PCS und ≥ 9,6 Punkte für MCS, festgelegt. - Das Ansprechen in den 8 einzelnen Subskalen nach SF-36 wurde gemäß IQWiG-Methoden als eine Reduktion um ≥ 15 % der Skalenspannweite, in diesem Fall jeweils um ≥ 15 Punkte, festgelegt. <u>Dargestellte Analysen:</u> Folgende Analysen werden für die ITT-Population der Studie ROBIN nachfolgend dargestellt: - Anzahl der Patientinnen mit PCS-Ansprechen auf dem SF-36 - Zeit bis zum ersten PCS-Ansprechen auf dem SF-36 (inkl. Kaplan-Meier-Plot) - Anzahl der Patientinnen mit MCS-Ansprechen auf dem SF-36 - Zeit bis zum ersten MCS-Ansprechen auf dem SF-36 (inkl. Kaplan-Meier-Plot) - Anzahl der Patientinnen mit Ansprechen in folgenden Subskalen des SF-36: o körperliche Schmerzen o soziale Funktionsfähigkeit o psychisches Wohlbefinden - Zeit bis zum ersten Ansprechen in folgenden Subskalen des SF-36 (inkl. Kaplan-Meier-

	Plot) o körperliche Schmerzen o soziale Funktionsfähigkeit o psychisches Wohlbefinden Folgende Analysen werden für die ITT Population der Studie ROBIN in Anhang 4-H dargestellt: - Veränderung des PCS-Gesamtwertes des SF-36 im Vergleich zu Baseline (inkl. grafische Darstellung über die Zeit). - Veränderung des MCS-Gesamtwertes des SF-36 im Vergleich zu Baseline (inkl. grafische Darstellung über die Zeit). - Veränderung jeder einzelnen Subskala des SF-36 im Vergleich zu Baseline (inkl. grafische Darstellung über die Zeit). - Anzahl der Patientinnen mit PCS-Verschlechterung auf dem SF-36 - Zeit bis zur ersten PCS-Verschlechterung auf dem SF-36 (inkl. Kaplan-Meier-Plot) - Anzahl der Patientinnen mit MCS-Verschlechterung auf dem SF-36 - Zeit bis zur ersten MCS-Verschlechterung auf dem SF-36 (inkl. Kaplan-Meier-Plot) - Anzahl der Patientinnen mit Verschlechterung in folgenden Subskalen des SF-36: o körperliche Schmerzen o soziale Funktionsfähigkeit o psychisches Wohlbefinden - Zeit bis zur ersten Verschlechterung in folgenden Subskalen des SF-36 (inkl. Kaplan-Meier-Plot) o körperliche Schmerzen o soziale Funktionsfähigkeit o psychisches Wohlbefinden <u>Statistische Methoden:</u> Eine ausführliche Beschreibung der Methodik, die auf ROBIN SAP beruht, findet sich in Abschnitt 4.2.5 bzw. in den Fußnoten.
SKY-LARK	Der Endpunkt SF-36 wurde in der SKYLARK-Studie nicht erhoben.
Meta-Analyse	Da der SF-36 nur in der ROBIN-Studie erhoben wurde kann keine Meta-Analyse durchgeführt werden.
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.	

Tabelle 4-54: Bewertung des Verzerrungspotenzials für SF-36 in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
ROBIN	niedrig	ja	ja	ja	ja	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

Bei der Studie ROBIN handelt es sich um eine doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der generische Endpunkt SF-36 wird wie im Studienprotokoll und statistischen Analyseplan präspezifiziert erhoben, um die Lebensqualität der Patientinnen zu erheben. Zusätzlich wurden *post-hoc* Analysen ergänzt, um den Anforderungen an die Nutzenbewertung zu entsprechen. Abgesehen von diesem Zusatz wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet. Das ITT-Prinzip wurde adäquat umgesetzt. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Tabelle 4-55: Rücklaufraten des SF-36 aus RCT mit dem zu bewertenden Arzneimittel

SF-36-Rücklaufrate	**Zuranolon+PiT N=77**	**PiT N=76**
Tag 15	74 (96 %)	73 (96 %)
Tag 45	73 (95 %)	69 (91 %)
Werte sind aus den MMRM Analysen entnommen (s. Anhang 4-G) Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.		

Die Rücklaufraten für den SF-36 lagen zu jedem Zeitpunkt über 90 %.

4.3.1.3.1.3.1.1 PCS des SF-36

Tabelle 4-56: Ergebnisse für das PCS-Ansprechen auf dem SF-36 aus RCT mit dem zu bewertenden Arzneimittel

PCS-Ansprechen auf dem SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	15 (19,7)	9 (12,0)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	2 (2,6)	2 (2,7)
RR	1,500	
95 %-KI für das RR	[0,718; 3,133]	
p-Wert für das RR	0,2802	

PCS-Ansprechen auf dem SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
OR	1,670	
95 %-KI für das OR	[0,646; 4,315]	
p-Wert für das OR	0,2899	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	12 (15,8)	7 (9,3)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	3 (3,9)	6 (8,0)
RR	1,543	
95 %-KI für das RR	[0,659; 3,615]	
p-Wert für das RR	0,3177	
OR	1,694	
95 %-KI für das OR	[0,599; 4,791]	
p-Wert für das OR	0,3200	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	

[1] Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert der PCS des SF-36, der Behandlung, dem Bewertungszeitpunkt und der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. Das RR basiert auf einer Poisson-Verteilung. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells.

[2] Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien.
Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Für das Ansprechen auf der PCS des SF-36 lag zu keinem Zeitpunkt ein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor.

Tabelle 4-57: Ergebnisse für Zeit bis zum ersten PCS-Ansprechen des SF-36 aus RCT mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten PCS-Ansprechen des SF-36	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Anzahl der Patientinnen mit Antwort, n (%) [2]	22 (28,9)	12 (16,0)
Zeit bis zum ersten PCS-Ansprechen des SF-36 (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	37,0 [15,0; NB]	NB [44,0; NB]
50. Perzentil - Median (95 %-KI)	87,0 [NB; NB]	NB [NB; NB]
75. Perzentil - Q3 (95 %-KI)	87,0 [NB; NB]	NB [NB; NB]
HR - Hazard Ratio [4]	1,619	
95 %-KI für HR	[0,7953; 3,2953]	
p-Wert für HR	0,1841	

Das Ansprechen nach SF-36 PCS ist definiert als Zunahme um ≥ 9,4 Punkte (15 % der Skalenspannweite) des Gesamtwerts von Baseline. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[1] Anzahl der Patientinnen mit nicht fehlenden PCS-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit einer PCS-Ansprechrate. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

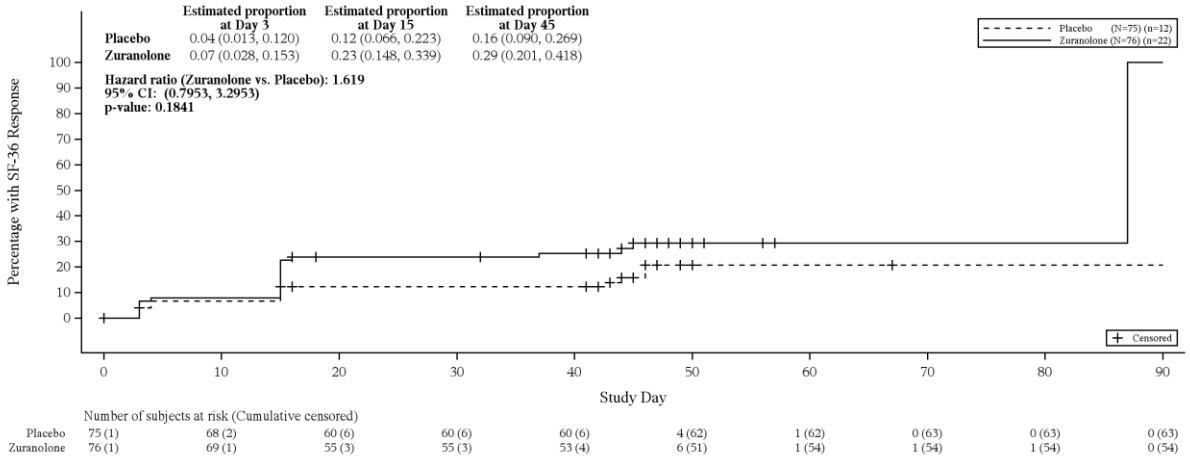
[3] Kaplan-Meier-Schätzer

[4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Antidepressiva zu Baseline, Baselinewert des PCS-Gesamtwerts und Behandlung.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Für die Zeit bis zum Ansprechen auf der PCS des SF-36 wurde kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen beobachtet.



[1] Patientinnen mit nicht fehlenden PCS-Gesamtwerten zu Baseline sind in der Analyse berücksichtigt. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Ansprechen auf der PCS ist definiert als Zunahme um ≥ 9,4 Punkte entsprechend 15 % der Skalenspannweite. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Antidepressiva zu Baseline, Baselinewert des PCS-Gesamtwerts und Behandlung angepasst ist.

Abbildung 4-10: Kaplan-Meier-Plot für Zeit bis zum ersten PCS-Ansprechen des SF-36

4.3.1.3.1.3.1.2 MCS des SF-36

Tabelle 4-58: Ergebnisse für MCS-Ansprechen auf dem SF-36 aus RCT mit dem zu bewertenden Arzneimittel

MCS-Ansprechen auf dem SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	50 (65,8)	47 (62,7)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	2 (2,6)	2 (2,7)
RR	1,058	
95 %-KI für das RR	[0,835; 1,341]	
p-Wert für das RR	0,6393	
OR	1,176	
95 %-KI für das OR	[0,599; 2,307]	
p-Wert für das OR	0,6376	

MCS-Ansprechen auf dem SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
ARR	0,042	
95 %-KI für das ARR	[-0,110; 0,195]	
p-Wert für das ARR	0,5843	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	59 (77,6)	48 (64,0)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	3 (3,9)	6 (8,0)
RR	1,223	
95 %-KI für das RR	[0,995; 1,502]	
p-Wert für das RR	0,0555	
OR	2,028	
95 %-KI für das OR	[0,993; 4,144]	
p-Wert für das OR	0,0525	
ARR	0,158	
95 %-KI für das ARR	[0,019; 0,297]	
p-Wert für das ARR	0,0260	

[1] Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert der MCS des SF-36, der Behandlung, dem Bewertungszeitpunkt und der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. Das RR basiert auf einer Poisson-Verteilung. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells.

[2] Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien.
Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Für das Ansprechen auf der MCS des SF-36 lag zu keinem Zeitpunkt ein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor.

Tabelle 4-59: Ergebnisse für Zeit bis zum ersten MCS-Ansprechen auf dem SF-36 aus RCT mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten MCS-Ansprechen des SF-36	**Zuranolon+PiT**	**PiT**
Day 15	**N = 77**	**N = 76**

Zeit bis zum ersten MCS-Ansprechen des SF-36	**Zuranolon+PiT**	**PiT**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Anzahl der Patientinnen mit Antwort, n (%) [2]	66 (86,8)	59 (78,7)
Zeit bis zum ersten MCS-Ansprechen des SF-36 (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	15,0 [4,0; 15,0]	14,0 [3,0; 15,0]
50. Perzentil - Median (95 %-KI)	15,0 [NB; NB]	15,0 [NB; NB]
75. Perzentil - Q3 (95 %-KI)	43,0 [16,0; 46,0]	45,0 [16,0; 46,0]
HR - Hazard Ratio [4]	1,051	
95 %-KI für HR	[0,7375; 1,4967]	
p-Wert für HR	0,7845	

Das Ansprechen nach SF-36 MCS ist definiert als Zunahme um ≥ 9.6 Punkte (15 % der Skalenspannweite) des Gesamtwerts von Baseline. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[1] Anzahl der Patientinnen mit nicht fehlenden MCS-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit einer MCS-Ansprechrate. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

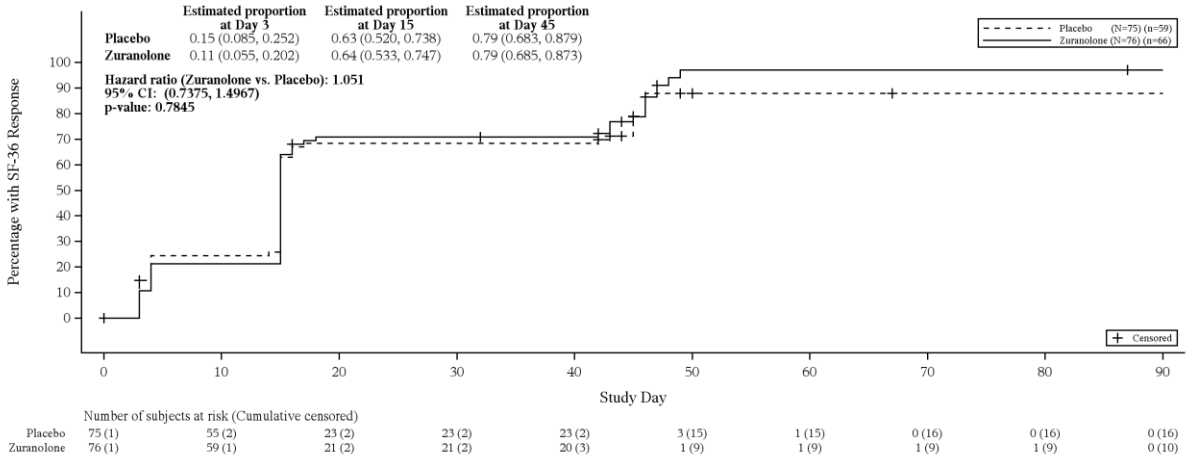
[3] Kaplan-Meier-Schätzer

[4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Antidepressiva zu Baseline, Baselinewert des MCS-Gesamtwerts und Behandlung.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Für die Zeit bis zum Ansprechen auf der MCS des SF-36 wurde kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen beobachtet.



[1] Patientinnen mit nicht fehlenden MCS-Gesamtwerten zu Baseline sind in der Analyse berücksichtigt. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Ansprechen auf der MCS ist definiert als Zunahme um ≥ 9,6 Punkte entsprechend 15 % der Skalenspannweite. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Antidepressiva zu Baseline, Baselinewert des MCS-Gesamtwerts und Behandlung angepasst ist.

Abbildung 4-11: Kaplan-Meier-Plot für Zeit bis zum ersten MCS-Ansprechen auf dem SF-36

4.3.1.3.1.3.1.3 Subskalen des SF-36

a. Subskala „Körperliche Schmerzen“

Tabelle 4-60: Ergebnisse für das Ansprechen auf der Subskala „Körperliche Schmerzen“ des SF-36 aus der RCT mit dem zu bewertenden Arzneimittel

Ansprechen Subskala „Körperliche Schmerzen“ des SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	40 (52,6)	32 (42,7)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	2 (2,6)	2 (2,7)
RR	1,204	
95 %-KI für das RR	[0,877; 1,652]	
p-Wert für das RR	0,2513	
OR	1,478	

Ansprechen Subskala „Körperliche Schmerzen“ des SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
95 %-KI für das OR	[0,751; 2,905]	
p-Wert für das OR	0,2578	
ARR	0,089	
95 %-KI für das ARR	[-0,045; 0,224]	
p-Wert für das ARR	0,1934	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	44 (57,9)	29 (38,7)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	3 (3,9)	6 (8,0)
RR	1,461	
95 %-KI für das RR	[1,051; 2,032]	
p-Wert für das RR	0,0243	
OR	2,266	
95 %-KI für das OR	[1,123; 4,572]	
p-Wert für das OR	0,0224	
ARR	0,173	
95 %-KI für das ARR	[0,036; 0,309]	
p-Wert für das ARR	0,0130	

Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert für Subskala Körperliche Schmerzen des SF-36, der Behandlung, dem Bewertungszeitpunkt und der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells.

[1] Anzahl der Patientinnen mit nicht fehlenden Werten für die Subskala Körperliche Schmerzen Gesamtwerte zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit Ansprechen auf der Subskala Körperliche Schmerzen des SF-36 (Zunahme um ≥ 15 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 100 des Subskalenwerts im Vergleich zu Baseline). Fehlende Subskalenwerte bei der Untersuchung wurden als "kein Ansprechen" imputiert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Für das Ansprechen auf der Subskala „Körperliche Schmerzen“ des SF-36 lag zu Tag 45 ein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor. Es beschrieben mehr Patientinnen, die Zuranolon erhielten, eine relevante Verbesserung ihrer Körperlichen Schmerzen als unter der Vergleichstherapie.

Tabelle 4-61: Ergebnisse für Zeit bis zum ersten Ansprechen auf der Subskala „Körperliche Schmerzen“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten Ansprechen Subskala „Körperliche Schmerzen“	**Zuranolon+PiT**	**PiT**
Day 15	**N = 77**	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Anzahl der Patientinnen mit Antwort, n (%) [2]	54 (71,1)	43 (57,3)
Zeit bis zum ersten Ansprechen Subskala „Körperliche Schmerzen“ (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	4,0 [3,0; 15,0]	15,0 [3,0; 15,0]
50. Perzentil - Median (95 %-KI)	15,0 [15,0; 43,0]	43,0 [15,0; 46,0]
75. Perzentil - Q3 (95 %-KI)	47,0 [45,0; NB]	NB [46,0; NB]
HR - Hazard Ratio [4]	1,229	
95 %-KI für HR	[0,8214; 1,8390]	
p-Wert für HR	0,3159	

Ansprechen auf der Subskala ist definiert als Zunahme um ≥ 15 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 100. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[1] Anzahl der Patientinnen mit nicht fehlenden Subskalen-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit Ansprechen auf einer Subskalen. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

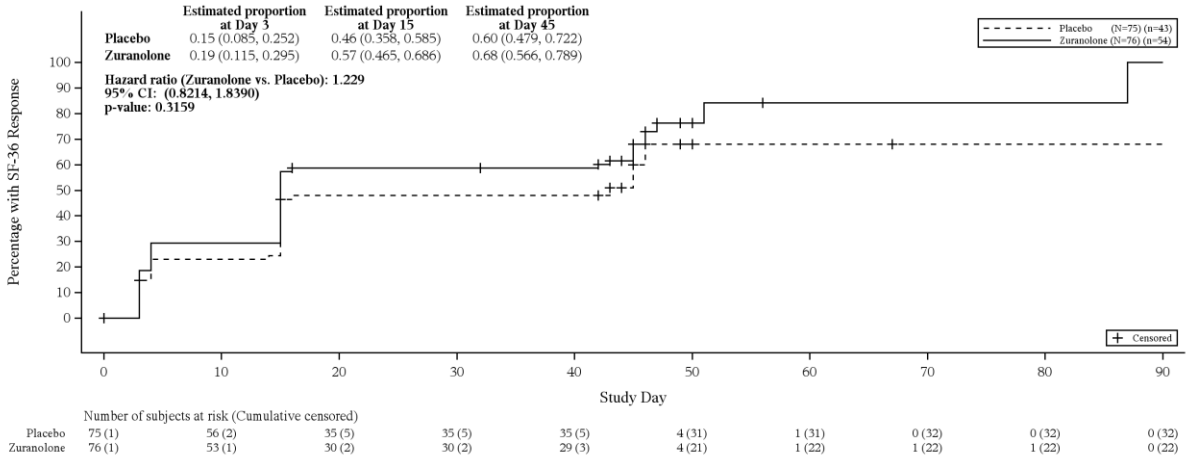
[3] Kaplan-Meier-Schätzer

[4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Antidepressiva zu Baseline, Baselinewert des Subskalen-Gesamtwerts und Behandlung.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Für die Zeit bis zum ersten Ansprechen auf der Subskala „Körperliche Schmerzen“ des SF-36 wird kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen beobachtet.



[1] Patientinnen mit nicht fehlenden Subskalen-Gesamtwerten zu Baseline sind in der Analyse berücksichtigt. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Ansprechen auf der Subskala ist definiert als Zunahme um ≥ 15 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 100. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Antidepressiva zu Baseline, Baselinewert des Subskalen-Gesamtwerts und Behandlung angepasst ist.

Abbildung 4-12: Kaplan-Meier-Plot für Zeit bis zum ersten Ansprechen auf der Subskala „Körperliche Schmerzen“ des SF-36

b. Subskala „Soziale Funktionsfähigkeit“

Tabelle 4-62: Ergebnisse für das Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel

Ansprechen Subskala „Soziale Funktionsfähigkeit“ des SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	50 (65,8)	43 (57,3)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	2 (2,6)	2 (2,7)
RR	1,149	
95 %-KI für das RR	[0,899; 1,470]	
p-Wert für das RR	0,2671	
OR	1,398	

Ansprechen Subskala „Soziale Funktionsfähigkeit“ des SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
95 %-KI für das OR	[0,716; 2,728]	
p-Wert für das OR	0,3266	
ARR	0,077	
95 %-KI für das ARR	[-0,075; 0,228]	
p-Wert für das ARR	0,3202	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	62 (81,6)	49 (65,3)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	3 (3,9)	6 (8,0)
RR	1,218	
95 %-KI für das RR	[1,011; 1,467]	
p-Wert für das RR	0,0378	
OR	2,360	
95 %-KI für das OR	[1,109; 5,023]	
p-Wert für das OR	0,0259	
ARR	0,150	
95 %-KI für das ARR	[0,018; 0,283]	
p-Wert für das ARR	0,0261	

Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert für Subskala Soziale Funktionsfähigkeit des SF-36, der Behandlung, dem Bewertungszeitpunkt und der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells.

[1] Anzahl der Patientinnen mit nicht fehlenden Werten für die Subskala Soziale Funktionsfähigkeit Gesamtwerte zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit Ansprechen auf der Subskala Soziale Funktionsfähigkeit des SF-36 (Zunahme um ≥ 15 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 100 des Subskalenwerts im Vergleich zu Baseline). Fehlende Subskalenwerte bei der Untersuchung wurden als "kein Ansprechen" imputiert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

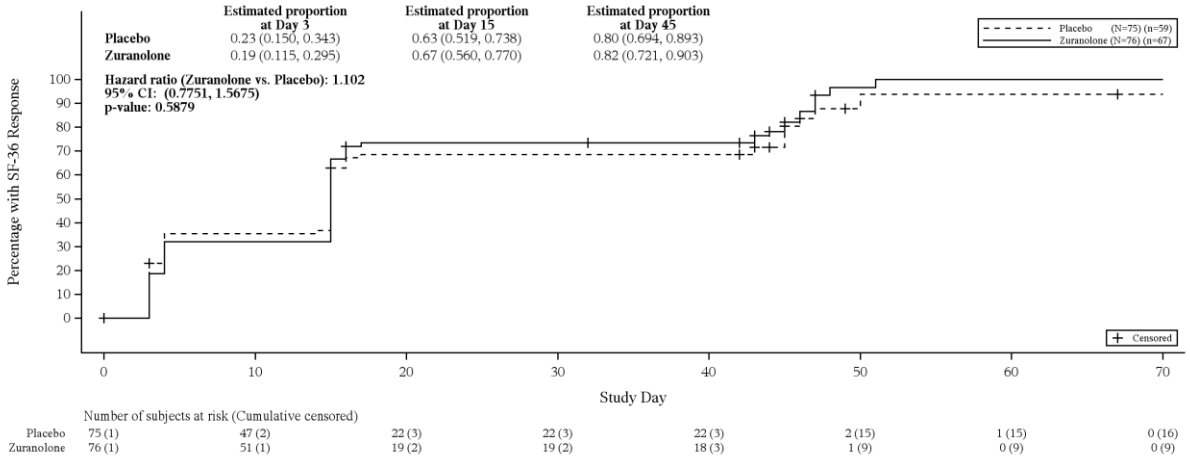
Quelle: [72]

Für das Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“ des SF-36 lag zu Tag 45 ein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor. Zum Beobachtungsende beschrieben deutlich mehr Patientinnen, die zuvor Zuranolon erhielten, eine relevante Verbesserung ihrer soziale Funktionsfähigkeit als unter der Vergleichstherapie.

Tabelle 4-63: Ergebnisse für Zeit bis zum ersten Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten Ansprechen Subskala „Soziale Funktionsfähigkeit“	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Anzahl der Patientinnen mit Antwort, n (%) [2]	67 (88,2)	59 (78,7)
Zeit bis zum ersten Ansprechen Subskala „Soziale Funktionsfähigkeit“ (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	4,0 [3,0; 15,0]	4,0 [3,0; 4,0]
50. Perzentil - Median (95 %-KI)	15,0 [NB; NB]	15,0 [NB; NB]
75. Perzentil - Q3 (95 %-KI)	43,0 [15,0; 46,0]	45,0 [16,0; 47,0]
HR - Hazard Ratio [4]	1,102	
95 %-KI für HR	[0,7751; 1,5675]	
p-Wert für HR	0,5879	
Ansprechen auf der Subskala ist definiert als Zunahme um ≥ 15 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 100. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert. [1] Anzahl der Patientinnen mit nicht fehlenden Subskalen-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit Ansprechen auf einer Subskalen. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. [3] Kaplan-Meier-Schätzer [4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Antidepressiva zu Baseline, Baselinewert des Subskalen-Gesamtwerts und Behandlung. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Für die Zeit bis zum ersten Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“ des SF-36 wird kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen beobachtet.



[1] Patientinnen mit nicht fehlenden Subskalen-Gesamtwerten zu Baseline sind in der Analyse berücksichtigt. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Ansprechen auf der Subskala ist definiert als Zunahme um ≥ 15 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 100. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Antidepressiva zu Baseline, Baselinewert des Subskalen-Gesamtwerts und Behandlung angepasst ist.

Abbildung 4-13: Kaplan-Meier-Plot für Zeit bis zum ersten Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“ des SF-36

c. Subskala „Psychisches Wohlbefinden“

Tabelle 4-64: Ergebnisse für das Ansprechen auf der Subskala „Psychisches Wohlbefinden“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel

Ansprechen Subskala „Psychisches Wohlbefinden“ des SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	54 (71,1)	47 (62,7)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	2 (2,6)	2 (2,7)
RR	1,149	
95 %-KI für das RR	[0,918; 1,439]	
p-Wert für das RR	0,2253	
OR	1,549	

Ansprechen Subskala „Psychisches Wohlbefinden“ des SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
95 %-KI für das OR	[0,775; 3,098]	
p-Wert für das OR	0,2157	
ARR	0,110	
95 %-KI für das ARR	[-0,038; 0,258]	
p-Wert für das ARR	0,1464	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	62 (81,6)	50 (66,7)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	3 (3,9)	6 (8,0)
RR	1,240	
95 %-KI für das RR	[1,028; 1,497]	
p-Wert für das RR	0,0249	
OR	2,380	
95 %-KI für das OR	[1,127; 5,028]	
p-Wert für das OR	0,0230	
ARR	0,182	
95 %-KI für das ARR	[0,054; 0,309]	
p-Wert für das ARR	0,0053	

Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert für Subskala Psychisches Wohlbefinden des SF-36, der Behandlung, dem Bewertungszeitpunkt und der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells.

[1] Anzahl der Patientinnen mit nicht fehlenden Werten für die Subskala Psychisches Wohlbefinden Gesamtwerte zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit Ansprechen auf der Subskala Psychisches Wohlbefinden des SF-36 (Zunahme um ≥ 15 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 100 des Subskalenwerts im Vergleich zu Baseline). Fehlende Subskalenwerte bei der Untersuchung wurden als "kein Ansprechen" imputiert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Für das Ansprechen auf der Subskala „Psychisches Wohlbefinden“ des SF-36 lag zu Tag 45 ein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor. Zum Beobachtungsende beschrieben deutlich mehr Patientinnen, die zuvor Zuranolon erhielten, eine relevante Verbesserung ihres psychischen Wohlbefindens als unter der Vergleichstherapie

Tabelle 4-65: Ergebnisse für Zeit bis zum ersten Ansprechen auf der Subskala „Psychisches Wohlbefinden“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten Ansprechen Subskala „Psychisches Wohlbefinden“	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Anzahl der Patientinnen mit Antwort, n (%) [2]	69 (90,8)	62 (82,7)
Zeit bis zum ersten Ansprechen Subskala „Psychisches Wohlbefinden“ (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	4,0 [3,0; 15,0]	3,0 [3,0; 4,0]
50. Perzentil - Median (95 %-KI)	15,0 [NB; NB]	15,0 [14,0; 15,0]
75. Perzentil - Q3 (95 %-KI)	43,0 [15,0; 46,0]	42,0 [15,0; 45,0]
HR - Hazard Ratio [4]	1,022	
95 %-KI für HR	[0,7245; 1,4427]	
p-Wert für HR	0,8999	

Ansprechen auf der Subskala ist definiert als Zunahme um ≥ 15 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 100. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[1] Anzahl der Patientinnen mit nicht fehlenden Subskalen-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit Ansprechen auf einer Subskalen. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

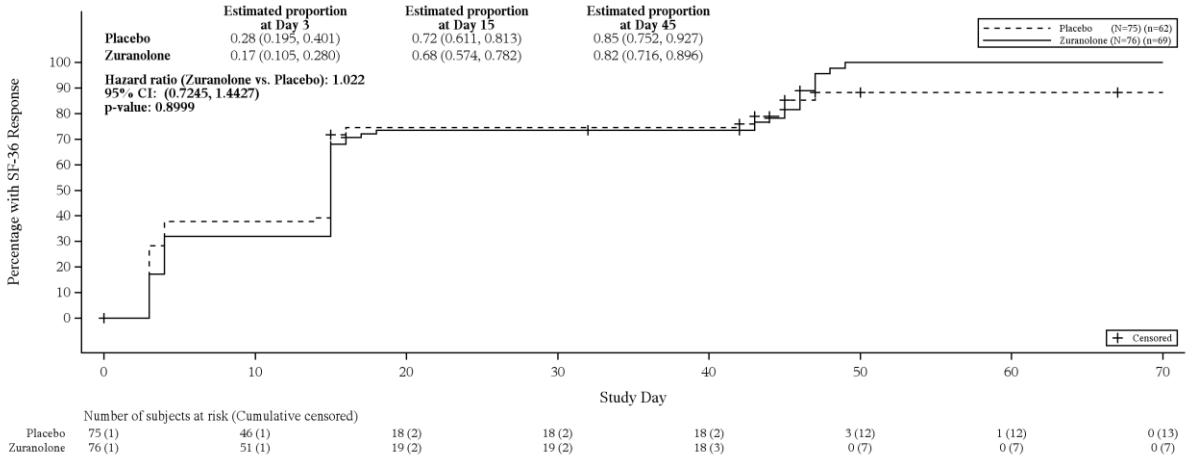
[3] Kaplan-Meier-Schätzer

[4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Antidepressiva zu Baseline, Baselinewert des Subskalen-Gesamtwerts und Behandlung.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Für die Zeit bis zum ersten Ansprechen auf der Subskala „Psychisches Wohlbefinden“ des SF-36 wird kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen beobachtet.



[1] Patientinnen mit nicht fehlenden Subskalen-Gesamtwerten zu Baseline sind in der Analyse berücksichtigt. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Ansprechen auf der Subskala ist definiert als Zunahme um ≥ 15 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 100. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Antidepressiva zu Baseline, Baselinewert des Subskalen-Gesamtwerts und Behandlung angepasst ist.

Abbildung 4-14: Kaplan-Meier-Plot für Zeit bis zum ersten Ansprechen auf der Subskala „Psychisches Wohlbefinden“ des SF-36

4.3.1.3.1.3.2 BIMF – RCT

Tabelle 4-66: Operationalisierung von BIMF

Studie	Operationalisierung
ROBIN	Der BIMF ist ein 20 Item umfassender Fragebogen, der zur Selbsteinschätzung der Mutterrolle/ mütterlichen Funktionsfähigkeit für den Zeitraum der ersten 12 Monate nach der Geburt entworfen wurde. Er deckt eine breite Palette funktionaler Bereiche ab, darunter Selbstfürsorge, Säuglingspflege, Mutter-Kind-Interaktion, psychisches Wohlbefinden der Mutter, soziale Unterstützung, Management und Anpassung. Jedes Item wird auf einer Skala von 0 (stimme überhaupt nicht zu) bis 6 (stimme voll und ganz zu) bewertet. Vor der Berechnung der Gesamtpunktzahl müssen die Fragen 16 und 18 umgekehrt bewertet werden – dabei wird die Antwort „0" zu „6", „1" zu „5" usw. Die BIMF-Gesamtpunktzahl ist die Summe der 20 Einzelbewertungen und liegt zwischen 0 und 120. Niedrigere Gesamtwerte weisen auf eine eingeschränkte Funktion in der Mutterrolle hin. Die 20 Items des BIMF umfassen: 1. Ich bin eine gute Mutter. 2. Ich fühle mich ausgeruht. 3. Ich fühle mich wohl mit der Art und Weise, wie ich mich entschieden habe, mein Baby zu ernähren (Stillen oder Flasche). 4. Mein Baby und ich verstehen uns. 5. Ich bin in der Lage, mich zu entspannen und Zeit mit meinem Baby zu genießen. 6. Es gibt Personen in meinem Leben, denen ich mein Baby anvertrauen kann, wenn ich eine Pause brauche. 7. Ich fühle mich wohl, wenn ich einer befreundeten Person oder einem Verwandten die Pflege meines Babys überlasse (z. B. Vater des Säuglings oder Partner*in). 8. Ich stehe in mir ausreichendem Kontakt mit anderen erwachsenen Personen. 9. Ich bekomme genug Ermutigung von anderen Personen. 10. Ich vertraue meinen Gefühlen (Instinkten), wenn es darum geht, Entscheidungen für mein Baby zu treffen. 11. Ich nehme mir jede Woche etwas Zeit, um Dinge für mich selbst zu tun. 12. Ich kümmere mich gut um die physischen Bedürfnisse meines Babys (Füttern, Windeln wechseln, Medizinische Termine). 13. Ich kümmere mich gut um meine eigenen physischen Bedürfnisse meines Babys (Essen, Duschen, usw.) 14. Ich treffe gute Entscheidungen über die Gesundheit und das Wohlbefinden meines Babys. 15. Mein Baby und ich finden uns in eine Routine ein. 16. Ich mache mir Sorgen darüber, wie andere mich (als Mutter) beurteilen. 17. Ich bin dazu fähig mich um mein Baby und meine anderen Verantwortlichkeiten zu kümmern. 18. Angst oder Sorgen beeinträchtigen oft meine Funktionsfähigkeit als Mutter. 19. Mit der Zeit werde ich besser darin, mich um mein Baby zu kümmern. 20. Ich bin zufrieden mit der Arbeit, die ich als neue Mutter leiste. Neben dem Gesamtwert des BIMF werden zwei weitere Subskalen „Kompetenz der Mutter" und „Bedürfnisse der Mutter" basierend auf den Antworten der Patientinnen berechnet. Die Punktzahl der Subskala „Kompetenz der Mutter" wird als Summe der Punkte der Items 1, 3, 4, 5, 10, 12, 14, 15, 17, 19 und 20 berechnet. Die Punktzahl der Subskala „Bedürfnisse der Mutter" wird als Summe der Punkte der Items 2, 6, 7, 8, 9, 11 und 13 berechnet.

Erhebende Person und Erhebungszeitpunkte

Der BIMF wurde von den Patienten zu folgenden Zeitpunkten selbst ausgefüllt:

- Während der Screening-Besuche; vor der Verabreichung am Tag 1; 12 ± 1 Stunde nach der Verabreichung am Abend vor den Tagen 3 (nur PA4), 8 und 15; und am Morgen der Tage 21 und 45.

Schwellenwerte

Responder-Analysen wurden *post-hoc* definiert und durchgeführt:

- Das Ansprechen nach BIMF-Gesamtwert wurde gemäß IQWiG-Methoden als eine Reduktion um ≥ 15 % der Skalenspannweite, in diesem Fall um ≥ 18 Punkte, festgelegt.
- Das Ansprechen auf der Subskala „Kompetenz der Mutter“ des BIMF wurde gemäß IQWiG-Methoden als eine Reduktion um ≥ 15 % der Skalenspannweite, in diesem Fall um ≥ 9,9 Punkte, festgelegt.
- Das Ansprechen auf der Subskala „Bedürfnisse der Mutter“ des BIMF wurde gemäß IQWiG-Methoden als eine Reduktion um ≥ 15 % der Skalenspannweite, in diesem Fall um ≥ 6,3 Punkte, festgelegt.

Dargestellte Analysen:

Folgende Analysen werden für die ITT-Population der Studie ROBIN nachfolgend dargestellt:

- Anzahl der Patientinnen mit BIMF-Gesamtwert-Ansprechen
- Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen (inkl. Kaplan-Meier-Plot)
- Anzahl der Patientinnen mit Ansprechen auf den folgenden Subskalen des BIMF:
    - Kompetenz der Mutter
    - Bedürfnisse der Mutter
- Zeit bis zum ersten Ansprechen auf den folgenden Subskalen des BIMF (inkl. Kaplan-Meier-Plot):
    - Kompetenz der Mutter
    - Bedürfnisse der Mutter

Folgende Analysen werden für die ITT Population der Studie ROBIN in Anhang 4-H dargestellt:

- Veränderung des BIMF-Gesamtwertes im Vergleich zu Baseline (inkl. grafische Darstellung über die Zeit).
- Veränderung auf den Subskalen „Kompetenz der Mutter“ und „Bedürfnisse der Mutter“ im Vergleich zu Baseline (inkl. grafische Darstellung über die Zeit)
- Anzahl der Patientinnen mit BIMF-Gesamtwert-Verschlechterung
- Zeit bis zur ersten BIMF-Gesamtwert-Verschlechterung (inkl. Kaplan-Meier-Plot)
- Anzahl der Patientinnen mit Verschlechterung auf den folgenden Subskalen des BIMF:
    - Kompetenz der Mutter
    - Bedürfnisse der Mutter
- Zeit bis zur ersten Verschlechterung auf den folgenden Subskalen des BIMF (inkl. Kaplan-Meier-Plot):
    - Kompetenz der Mutter
    - Bedürfnisse der Mutter

Statistische Methoden:

Eine ausführliche Beschreibung der Methodik, die auf dem ROBIN SAP beruht, findet sich in Abschnitt 4.2.5 bzw. in den Fußnoten.

Wenn mehr als 4 Einzelwert-Scores fehlen, wird der BIMF-Gesamtwert nicht berechnet und als fehlend belassen. Wenn weniger als oder gleich 4 einzelne Item-Werte fehlen, werden die fehlenden Werte durch den Mittelwert aller anderen verfügbaren Werte ersetzt (unter Berücksichtigung der umgekehrten Bewertung der Fragen 16 und 18).

	Wenn mehr als 2 einzelne Item-Werte aus der Subskalen-Wertung „Kompetenz der Mutter" fehlen, wird sie als fehlend belassen. Wenn weniger als oder gleich 2 einzelne Item-Werte fehlen, werden die fehlenden Werte durch den Mittelwert aller anderen verfügbaren Werte ersetzt. Wenn mehr als 1 einzelner Item-Wert aus der Subskalen-Wertung „Bedürfnisse der Mutter" fehlt, wird er als fehlend gesetzt. Andernfalls wird der fehlende Wert durch den Mittelwert aller anderen verfügbaren Werte ersetzt.
SKY-LARK	Der Endpunkt BIMF wurde in der SKYLARK-Studie nicht erhoben.
Meta-Analyse	Da der BIMF nur in der ROBIN-Studie erhoben wurde kann keine Meta-Analyse durchgeführt werden.
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.	

Tabelle 4-67: Bewertung des Verzerrungspotenzials für BIMF in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
ROBIN	niedrig	ja	ja	ja	ja	niedrig
SKYLARK	niedrig	ja	ja	ja	ja	niedrig

Bei der Studie ROBIN handelt es sich um eine doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der BIMF ist ein Fragebogen, den die Patientinnen eigenständig ausfüllen. Der Endpunkt wurde, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert, erhoben. Es wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. *Post-hoc* definierte Analysen wurden ergänzt, um den Anforderungen der Nutzenbewertung zu entsprechen. Somit ist eine ergebnisunabhängige Berichterstattung für den Endpunkt gewährleistet. Das ITT-Prinzip wurde adäquat umgesetzt. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Tabelle 4-68: Rücklaufrate für BIMF aus RCT mit dem zu bewertenden Arzneimittel

BIMF-Rücklaufrate	**Zuranolon+PiT N=77**	**PiT N=76**
Tag 3	72 (94 %)	72 (95 %)
Tag 8	75 (97 %)	74 (97 %)
Tag 15	74 (96 %)	73 (96 %)
Tag 21	74 (96 %)	73 (96 %)
Tag 45	73(95 %)	69 (91 %)
Werte sind aus den MMRM Analysen entnommen (s. Anhang 4-G) Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.		

Die Rücklaufraten für den BIMF lagen zu jedem Zeitpunkt über 90 %.

4.3.1.3.1.3.2.1 Gesamtwert

Tabelle 4-69: Ergebnisse für BIMF-Gesamtwert-Ansprechen aus RCT mit dem zu bewertenden Arzneimittel

BIMF-Gesamtwert-Ansprechen (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	11 (14,5)	7 (9,3)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	4 (5,3)	3 (4,0)
RR	1,451	
95 %-KI für das RR	[0,600; 3,508]	
p-Wert für das RR	0,4081	
OR	1,355	
95 %-KI für das OR	[0,496; 3,700]	
p-Wert für das OR	0,5529	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	

BIMF-Gesamtwert-Ansprechen (Zunahme um ≥ 15 % der Skalenspannweite)	Zuranolon+PiT	PiT
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	23 (30,3)	19 (25,3)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	1 (1,3)	1 (1,3)
RR	1,118	
95 %-KI für das RR	[0,679; 1,840]	
p-Wert für das RR	0,6609	
OR	1,074	
95 %-KI für das OR	[0,526; 2,194]	
p-Wert für das OR	0,8440	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	35 (46,1)	20 (26,7)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	2 (2,6)	2 (2,7)
RR	1,616	
95 %-KI für das RR	[1,040; 2,513]	
p-Wert für das RR	0,0330	
OR	2,020	
95 %-KI für das OR	[1,001; 4,077]	
p-Wert für das OR	0,0496	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	35 (46,1)	35 (46,7)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	2 (2,6)	2 (2,7)

BIMF-Gesamtwert-Ansprechen (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
RR	0,924	
95 %-KI für das RR	[0,661; 1,290]	
p-Wert für das RR	0,6410	
OR	0,822	
95 %-KI für das OR	[0,419; 1,613]	
p-Wert für das OR	0,5684	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	49 (64,5)	37 (49,3)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	3 (3,9)	6 (8,0)
RR	1,223	
95 %-KI für das RR	[0,914; 1,636]	
p-Wert für das RR	0,1753	
OR	1,687	
95 %-KI für das OR	[0,843; 3,374]	
p-Wert für das OR	0,1393	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	

BIMF-Gesamtwert-Ansprechen (Zunahme um ≥ 15 % der Skalenspannweite)	Zuranolon+PiT	PiT
Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert für BIMF-Gesamtwert, der Behandlung, dem Bewertungszeitpunkt und der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells. [1] Anzahl der Patientinnen mit nicht fehlenden Werten für BIMF-Gesamtwert zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit Ansprechen auf den BIMF-Gesamtwert (Zunahme um ≥ 18 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 120 des Gesamtwerts im Vergleich zu Baseline). Fehlende Gesamtwerte bei der Untersuchung wurden als "kein Ansprechen" imputiert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

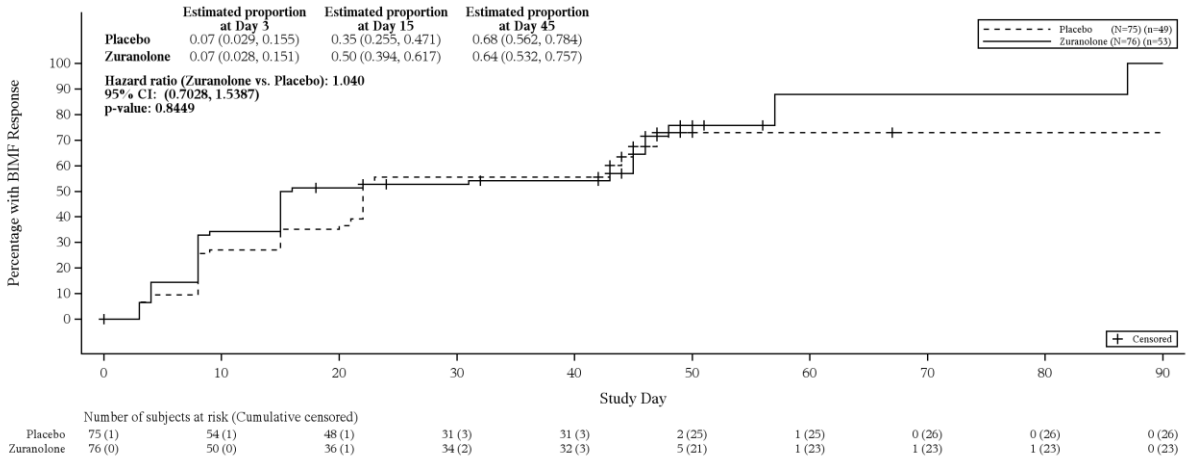
Für das BIMF-Gesamtwert-Ansprechen lag zu Tag 15 ein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor. So beschrieben deutlich mehr Patientinnen, die Zuranolon erhielten, eine relevante Verbesserung ihrer mütterlichen Funktionsfähigkeit als unter der Vergleichstherapie.

Tabelle 4-70: Ergebnisse für Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen Gesamtwert aus RCT mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen	Zuranolon+PiT	PiT
	N = 77	N = 76
Anzahl der evaluierbaren Patientinnen [1]	76	75
Anzahl der Patientinnen mit Antwort, n (%) [2]	53 (69,7)	49 (65,3)
Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	8,0 [8,0; 15,0]	8,0 [8,0; 15,0]
50. Perzentil - Median (95 %-KI)	15,5 [15,0; 45,0]	22,0 [21,0; 44,0]
75. Perzentil - Q3 (95 %-KI)	48,0 [45,0; NB]	NB [45,0; NB]
HR - Hazard Ratio [4]	1,040	
95 %-KI für HR	[0,7028; 1,5387]	
p-Wert für HR	0,8449	

Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen	Zuranolon+PiT	PiT
Ansprechen für BIMF Gesamtwert ist definiert als Zunahme um ≥ 18 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 120. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert. [1] Anzahl der Patientinnen mit nicht fehlenden BIMF-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit Ansprechen nach BIMF-Gesamtwert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. [3] Kaplan-Meier-Schätzer [4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Antidepressiva zu Baseline, Baselinewert des BIMF-Gesamtwerts und Behandlung. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Für die Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen wurde kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen beobachtet.



[1] Patientinnen mit nicht fehlenden BIMF-Gesamtwerten zu Baseline sind in der Analyse berücksichtigt. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Ansprechen nach BIMF-Gesamtwert ist definiert als Zunahme um ≥ 18 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 120. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Antidepressiva zu Baseline, Baselinewert des BIMF-Gesamtwerts und Behandlung angepasst ist.

Abbildung 4-15: Kaplan-Meier-Plot für Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen

4.3.1.3.1.3.2.2 Subskalen des BIMF

a. Subskala „Kompetenz der Mutter“

Tabelle 4-71: Ergebnisse für Ansprechen auf der Subskala „Kompetenz der Mutter“ des BIMF aus RCT mit dem zu bewertenden Arzneimittel

Ansprechen Subskala „Kompetenz der Mutter“ des BIMF (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	7 (9,2)	8 (10,7)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	4 (5,3)	3 (4,0)
RR	0,703	
95 %-KI für das RR	[0,280; 1,766]	
p-Wert für das RR	0,4536	
OR	0,525	
95 %-KI für das OR	[0,162; 1,696]	
p-Wert für das OR	0,2814	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	18 (23,7)	14 (18,7)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	1 (1,3)	1 (1,3)
RR	1,033	
95 %-KI für das RR	[0,593; 1,800]	
p-Wert für das RR	0,9078	
OR	0,872	
95 %-KI für das OR	[0,377; 2,019]	
p-Wert für das OR	0,7493	
ARR	NB	
95 %-KI für das ARR	NB	

Ansprechen Subskala „Kompetenz der Mutter“ des BIMF (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
p-Wert für das ARR	NB	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	27 (35,5)	19 (25,3)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	2 (2,6)	2 (2,7)
RR	1,142	
95 %-KI für das RR	[0,725; 1,800]	
p-Wert für das RR	0,5668	
OR	1,084	
95 %-KI für das OR	[0,487; 2,410]	
p-Wert für das OR	0,8438	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	31 (40,8)	21 (28,0)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	2 (2,6)	2 (2,7)
RR	1,186	
95 %-KI für das RR	[0,783; 1,798]	
p-Wert für das RR	0,4206	
OR	1,221	
95 %-KI für das OR	[0,568; 2,624]	
p-Wert für das OR	0,6092	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	41 (53,9)	24 (32,0)

Ansprechen Subskala „Kompetenz der Mutter“ des BIMF (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	3 (3,9)	6 (8,0)
RR	1,373	
95 %-KI für das RR	[0,921; 2,046]	
p-Wert für das RR	0,1194	
OR	1,948	
95 %-KI für das OR	[0,889; 4,269]	
p-Wert für das OR	0,0959	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert für Subskala, der Behandlung, dem Bewertungszeitpunkt und der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells. [1] Anzahl der Patientinnen mit nicht fehlenden Werten für Subskala zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit Ansprechen auf der Subskala (Zunahme um ≥ 9,9 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 66 des Subskalenwertes im Vergleich zu Baseline). Fehlende Subskalenwerte bei der Untersuchung wurden als "kein Ansprechen" imputiert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Für das Ansprechen auf der Subskala „Kompetenz der Mutter“ lag zu keinem Zeitpunkt ein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor.

Tabelle 4-72: Ergebnisse für Zeit bis zum ersten Ansprechen auf der Subskala „Kompetenz der Mutter“ des BIMF aus RCT mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten Ansprechen Subskala „Kompetenz der Mutter“ des BIMF	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75

Zeit bis zum ersten Ansprechen Subskala „Kompetenz der Mutter“ des BIMF	**Zuranolon+PiT**	**PiT**
Anzahl der Patientinnen mit Antwort, n (%) [2]	45 (59,2)	35 (46,7)
Zeit bis zum ersten Ansprechen Subskala „Kompetenz der Mutter“ (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	9,0 [8,0; 15,0]	15,0 [8,0; 22,0]
50. Perzentil - Median (95 %-KI)	43,0 [15,0; NB]	46,0 [23,0; NB]
75. Perzentil - Q3 (95 %-KI)	57,0 [57,0; NB]	NB [NB; NB]
HR - Hazard Ratio [4]	0,986	
95 %-KI für HR	[0,6289; 1,5444]	
p-Wert für HR	0,9492	

Ansprechen für die Subskala ist definiert als Zunahme um ≥ 9,9 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 66. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[1] Anzahl der Patientinnen mit nicht fehlenden Subskalenwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit Ansprechen nach Subskalenwerten. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

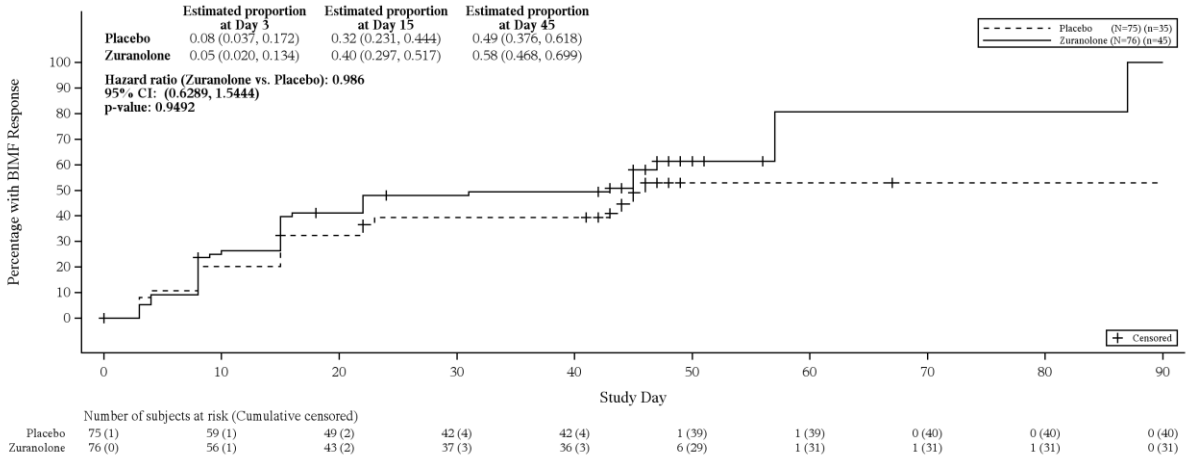
[3] Kaplan-Meier-Schätzer

[4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Antidepressiva zu Baseline, Baselinewert der Subskala und Behandlung.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Für die Zeit bis zum ersten Ansprechen auf der Subskala „Kompetenz der Mutter“ des BIMF wurde kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen beobachtet.



[1] Patientinnen mit nicht fehlenden Subskalenwerten zu Baseline sind in der Analyse berücksichtigt. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Ansprechen nach Subskala ist definiert als Zunahme um ≥ 9,9 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 66. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Antidepressiva zu Baseline, Baselinewert der Subskala und Behandlung angepasst ist.

Abbildung 4-16: Kaplan-Meier-Plot für Zeit bis zum ersten Ansprechen auf der Subskala „Kompetenz der Mutter“ des BIMF

b. Subskala „Bedürfnisse der Mutter“

Tabelle 4-73: Ergebnisse für Ansprechen auf der Subskala „Bedürfnisse der Mutter“ aus RCT mit dem zu bewertenden Arzneimittel

Ansprechen Subskala „Bedürfnisse der Mutter“ des BIMF (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	17 (22,4)	12 (16,0)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	4 (5,3)	3 (4,0)
RR	1,372	
95 %-KI für das RR	[0,706; 2,665]	
p-Wert für das RR	0,3505	
OR	1,409	
95 %-KI für das OR	[0,625; 3,176]	

Ansprechen Subskala „Bedürfnisse der Mutter“ des BIMF (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
p-Wert für das OR	0,4080	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	31 (40,8)	33 (44,0)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	1 (1,3)	1 (1,3)
RR	0,948	
95 %-KI für das RR	[0,665; 1,352]	
p-Wert für das RR	0,7692	
OR	0,836	
95 %-KI für das OR	[0,442; 1,583]	
p-Wert für das OR	0,5831	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	35 (46,1)	33 (44,0)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	2 (2,6)	2 (2,7)
RR	1,054	
95 %-KI für das RR	[0,747; 1,487]	
p-Wert für das RR	0,7654	
OR	1,050	
95 %-KI für das OR	[0,539; 2,047]	
p-Wert für das OR	0,8850	
ARR	NB	
95 %-KI für das ARR	NB	

Ansprechen Subskala „Bedürfnisse der Mutter“ des BIMF (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
p-Wert für das ARR	NB	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	32 (42,1)	40 (53,3)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	2 (2,6)	2 (2,7)
RR	0,801	
95 %-KI für das RR	[0,580; 1,107]	
p-Wert für das RR	0,1788	
OR	0,603	
95 %-KI für das OR	[0,315; 1,153]	
p-Wert für das OR	0,1259	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	50 (65,8)	41 (54,7)
Anzahl der Patientinnen mit fehlender Bewertung, n (%)	3 (3,9)	6 (8,0)
RR	1,165	
95 %-KI für das RR	[0,905; 1,501]	
p-Wert für das RR	0,2364	
OR	1,604	
95 %-KI für das OR	[0,805; 3,194]	
p-Wert für das OR	0,1792	
ARR	NB	
95 %-KI für das ARR	NB	
p-Wert für das ARR	NB	

Ansprechen Subskala „Bedürfnisse der Mutter“ des BIMF (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon+PiT**	**PiT**
Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert für Subskala, der Behandlung, dem Bewertungszeitpunkt und der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur. RR, OR und ARR werden aus demselben GEE-Modell mit Log-, Logit- und Identitätslinkfunktion geschätzt. NB zeigt an, dass nicht berechenbar aufgrund der Nichtkonvergenz des Modells. [1] Anzahl der Patientinnen mit nicht fehlenden Werten für Subskala zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit Ansprechen auf der Subskala (Zunahme um ≥ 6,3 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 42 des Subskalenwertes im Vergleich zu Baseline). Fehlende Subskalenwerte bei der Untersuchung wurden als "kein Ansprechen" imputiert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

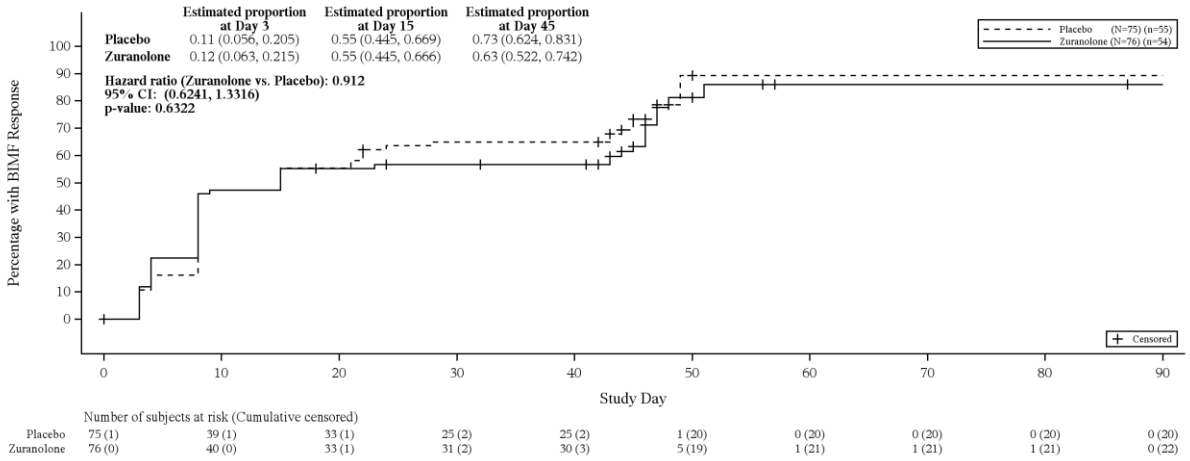
Für das Ansprechen auf der Subskala „Bedürfnisse der Mutter“ des BIMF lag zu keinem Zeitpunkt ein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor.

Tabelle 4-74: Ergebnisse für Zeit bis zum ersten Ansprechen auf der Subskala „Bedürfnisse der Mutter“ des BIMF aus RCT mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten Ansprechen Subskala „Bedürfnisse der Mutter“ des BIMF	**Zuranolon+PiT**	**PiT**
	N = 77	**N = 76**
Anzahl der evaluierbaren Patientinnen [1]	76	75
Anzahl der Patientinnen mit Antwort, n (%) [2]	54 (71,1)	55 (73,3)
Zeit bis zum ersten Ansprechen Subskala „Bedürfnisse der Mutter“ (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	8,0 [4,0; 8,0]	8,0 [4,0; 8,0]
50. Perzentil - Median (95 %-KI)	15,0 [8,0; 44,0]	15,0 [8,0; 22,0]
75. Perzentil - Q3 (95 %-KI)	47,0 [46,0; NB]	47,0 [28,0; NB]
HR - Hazard Ratio [4]	0,912	
95 %-KI für HR	[0,6241; 1,3316]	
p-Wert für HR	0,6322	

Zeit bis zum ersten Ansprechen Subskala „Bedürfnisse der Mutter“ des BIMF	Zuranolon+PiT	PiT
Ansprechen für die Subskala ist definiert als Zunahme um ≥ 6,3 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 42. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert. [1] Anzahl der Patientinnen mit nicht fehlenden Subskalenwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit Ansprechen nach Subskalenwerten. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. [3] Kaplan-Meier-Schätzer [4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Antidepressiva zu Baseline, Baselinewert der Subskala und Behandlung. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Für die Zeit bis zum ersten Ansprechen auf der Subskala „Bedürfnisse der Mutter“ des BIMF wurde kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen beobachtet.



[1] Patientinnen mit nicht fehlenden Subskalenwerten zu Baseline sind in der Analyse berücksichtigt. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Ansprechen nach Subskala ist definiert als Zunahme um ≥ 6,3 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 42. Tage werden ab dem Datum der ersten Dosis berechnet. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[3] Der geschätzte Anteil der Patientinnen mit Ereignissen basiert auf der Kaplan-Meier-Produktgrenzenmethode. Das HR, das 95 %-KI und der p-Wert basieren auf einem Cox-Proportional-Hazards-Modell, das für Antidepressiva zu Baseline, Baselinewert der Subskala und Behandlung angepasst ist.

Abbildung 4-17: Kaplan-Meier-Plot für Zeit bis zum ersten Ansprechen auf der Subskala „Bedürfnisse der Mutter“ des BIMF

4.3.1.3.1.4 Sicherheit – RCT

4.3.1.3.1.4.1 C-SSRS – RCT

Tabelle 4-75: Operationalisierung von C-SSRS

<table>
<tr><th>Studie</th><th>Operationalisierung</th></tr>
<tr><td>ROBIN</td><td>Die C-SSRS (v14.0 „Baseline“ und „Since last Visit“) wurde durch die Patient*innen berichtet und ist ein etabliertes Fragebogeninstrument zur standardisierten Untersuchung von suizidalen Gedanken und Verhalten. Der Fragebogen besteht aus zwei Domänen, Suizidgedanken und suizidales Verhalten mit jeweils 5 Fragen und jeweils zwei Antwortmöglichkeiten (ja/nein).
Eine zusätzliche "Ja/Nein"-Frage wird verwendet, um zu erfassen, ob die Person selbstverletzendes Verhalten ohne suizidale Absicht gezeigt hat.
<table>
<tr><td colspan="2">Suizidale Gedanken</td></tr>
<tr><td>Frage 1</td><td>Wunsch zu sterben</td></tr>
<tr><td>Frage 2</td><td>Nicht-spezifische aktive Suizidgedanken</td></tr>
<tr><td>Frage 3</td><td>Aktive Selbstmordgedanken mit beliebigen Methoden (nicht geplant) ohne Absicht zu handeln</td></tr>
<tr><td>Frage 4</td><td>Aktive Selbstmordgedanken mit einer gewissen Handlungsabsicht, ohne spezifischen Plan</td></tr>
<tr><td>Frage 5</td><td>Aktive Selbstmordgedanken mit konkretem Plan und Absicht</td></tr>
<tr><td colspan="2">Suizidales Verhalten</td></tr>
<tr><td>Frage 6</td><td>Vorbereitende Handlungen oder Verhaltensweisen</td></tr>
<tr><td>Frage 7</td><td>Abgebrochener Versuch</td></tr>
<tr><td>Frage 8</td><td>Unterbrochener Versuch</td></tr>
<tr><td>Frage 9</td><td>Tatsächlicher Versuch</td></tr>
<tr><td>Frage 10</td><td>Suizid</td></tr>
</table>
<u>Erhebende Person und Erhebungszeitpunkte</u>
Erhebungszeitpunkte sind zu Baseline, Tag 1, 3, 8, 15, 21, sowie 45.
<u>Dargestellte Analysen in Anhang 4-H</u>:
- Ergebnisse für C-SSRS zu Baseline aus der Studie ROBIN mit dem zu bewertenden Arzneimittel
- Ergebnisse für C-SSRS an Tag 3 aus der Studie ROBIN mit dem zu bewertenden Arzneimittel
- Ergebnisse für C-SSRS an Tag 15 aus der Studie ROBIN mit dem zu bewertenden Arzneimittel
- Ergebnisse für C-SSRS bei der letzten Bewertung nach Baseline aus der der Studie ROBIN mit dem zu bewertenden Arzneimittel</td></tr>
<tr><td>SKY-LARK</td><td>Die Definition des Endpunkts C-SSRS entspricht der Definition in der Studie ROBIN.
<u>Erhebende Person und Erhebungszeitpunkte</u>
Erhebungszeitpunkte sind zu Baseline, Tag 1, 3, 8, 15, 21, 28 sowie 45.
<u>Dargestellte Analysen in Anhang 4-I:</u>
- Ergebnisse für C-SSRS zu Baseline aus der Studie SKYLARK mit dem zu bewertenden</td></tr>
</table>

Studie	Operationalisierung
	Arzneimittel - Ergebnisse für C-SSRS an Tag 3 aus der Studie SKYLARK mit dem zu bewertenden Arzneimittel - Ergebnisse für C-SSRS an Tag 15 aus der Studie SKYLARK mit dem zu bewertenden Arzneimittel - Ergebnisse für C-SSRS bei der letzten Bewertung nach Baseline aus der der Studie SKYLARK mit dem zu bewertenden Arzneimittel
Meta-Analyse	Die Definition des Endpunkts C-SSRS entspricht der Definition in der Studie ROBIN. Dargestellte Analysen: Folgende Analysen werden für die gepoolte Sicherheits-Population beider Studien nachfolgend dargestellt: - Ergebnisse für C-SSRS zu Baseline aus der Meta-Analyse mit dem zu bewertenden Arzneimittel - Ergebnisse für C-SSRS an Tag 3 aus der Meta-Analyse mit dem zu bewertenden Arzneimittel - Ergebnisse für C-SSRS an Tag 15 aus der Meta-Analyse mit dem zu bewertenden Arzneimittel - Ergebnisse für C-SSRS bei der letzten Bewertung nach Baseline aus der Meta-Analyse mit dem zu bewertenden Arzneimittel
Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.	

Tabelle 4-76: Bewertung des Verzerrungspotenzials für C-SSRS in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
ROBIN	niedrig	ja	ja	ja	ja	niedrig
SKYLARK	niedrig	ja	ja	ja	ja	niedrig

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der Endpunkt C-SSRS wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Eine ereignisunabhängige Berichterstattung ist gewährleistet. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Tabelle 4-77: Ergebnisse für C-SSRS zu Baseline aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Anzahl der Patientinnen mit mindestens einer „Ja“ Antwort zu Baseline	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 171**
Suizidale Gedanken n, (%)	58 (33,0)	48 (28,1)
Wunsch zu sterben	47 (26,7)	42 (24,6)
Nicht-spezifische aktive Suizidgedanken	20 (11,4)	13 (7,6)
Aktive Selbstmordgedanken mit beliebigen Methoden (nicht geplant) ohne Absicht zu handeln	7 (4,0)	4 (2,3)
Aktive Selbstmordgedanken mit einer gewissen Handlungsabsicht, ohne spezifischen Plan	3 (1,7)	1 (0,6)
Aktive Selbstmordgedanken mit konkretem Plan und Absicht	1 (0,6)	1 (0,6)
Suizidales Verhalten	2 (1,1)	1 (0,6)
Vorbereitende Handlungen oder Verhaltensweisen	1 (0,6)	0
Abgebrochener Versuch	1 (0,6)	0
Unterbrochener Versuch	0	0
Tatsächlicher Versuch	0	1 (0,6)
Suizid	0	0
Selbstmordgedanken oder suizidales Verhalten	58 (33,0)	48 (28,1)
Selbstverletzendes Verhalten ohne suizidale Absicht	2 (1,1)	3 (1,8)

Hinweis 1: Für Selbstmordgedanken (1 – 5) oder suizidales Verhalten (6 – 10) wird eine Antwort mit „Ja" gewertet, wenn bei der Visite eine Frage zu Selbstmordgedanken oder suizidalem Verhalten mit „Ja" beantwortet wird. Die Kategorie Selbstmordgedanken oder suizidales Verhalten umfasst Patientinnen, die entweder bei Selbstmordgedanken (1 – 5) oder bei suizidalem Verhalten (6 – 10) gezählt werden.

Hinweis 2: Als Baseline gelten alle Bewertungen, die vor der ersten Dosis des Prüfpräparats durchgeführt wurden, wobei Lebenszeitbewertungen ausgeschlossen sind und der schlechteste Wert für jede Frage gezählt wird.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Tabelle 4-78: Ergebnisse für C-SSRS an Tag 3 aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

	Zuranolon+PiT	**PiT**	**Zuranolon+PiT vs. PiT**		
N	176	171	**RR Effektmaß [95 %-KI] p-Wert**	**OR Effektmaß [95 %-KI]**	**ARR Effektmaß [95 %-KI]**
C-SSRS Selbstmordgedanken (Frage 1 – 5)					
Anzahl der Patientinnen	7 (4,0)	11 (6,4)	0,61 [0,244; 1,547] 0,3013	0,60 [0,226; 1,582]	-0,03 [-0,073; 0,022]
C-SSRS Suizidales Verhalten (Frage 6 – 10)					
Anzahl der Patientinnen	0	0	NA	NA	NA
C-SSRS Selbstmordgedanken oder suizidales Verhalten (Frage 1 – 10)					
Anzahl der Patientinnen	7 (4,0)	11 (6,4)	0,61 [0,244; 1,547] 0,3013	0,60 [0,226; 1,582]	-0,03 [-0,073; 0,022]
C-SSRS Selbstverletzendes Verhalten ohne suizidale Absicht					
Anzahl der Patientinnen	0	1 (0,6)	0,32 [0,013; 7,847] 0,4867	0,32 [0,013; 7,910]	-0,01 [-0,022; 0,010]

Hinweis 1: Für Selbstmordgedanken (1 – 5) oder suizidales Verhalten (6 – 10) wird eine Antwort mit „Ja" gewertet, wenn bei der Visite eine Frage zu Selbstmordgedanken oder suizidalem Verhalten mit „Ja" beantwortet wird. Die Kategorie Selbstmordgedanken oder suizidales Verhalten umfasst Patientinnen, die entweder bei Selbstmordgedanken (1 – 5) oder bei suizidalem Verhalten (6 – 10) gezählt werden.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

An Tag 3 lag kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor.

Tabelle 4-79: Ergebnisse für C-SSRS an Tag 15 aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

	Zuranolon+PiT	**PiT**	**Zuranolon+PiT vs. PiT**		
N	176	171	**RR Effektmaß [95 %-KI] p-Wert**	**OR Effektmaß [95 %-KI]**	**ARR Effektmaß [95 %-KI]**
C-SSRS Selbstmordgedanken (Frage 1 – 5)					

	Zuranolon+PiT	PiT	Zuranolon+PiT vs. PiT		
N	176	171	**RR Effektmaß [95 %-KI] p-Wert**	**OR Effektmaß [95 %-KI]**	**ARR Effektmaß [95 %-KI]**
Anzahl der Patientinnen	9 (5,1)	9 (5,3)	0,95 [0,388; 2,339] 0,9157	0,95 [0,367; 2,457]	0,00 [-0,052; 0,046]
C-SSRS Suizidales Verhalten (Frage 6 – 10)					
Anzahl der Patientinnen	0	0	NA	NA	NA
C-SSRS Selbstmordgedanken oder suizidales Verhalten (Frage 1 – 10)					
Anzahl der Patientinnen	9 (5,1)	9 (5,3)	0,95 [0,388; 2,339] 0,9157	0,95 [0,367; 2,457]	0,00 [-0,052; 0,046]
C-SSRS Selbstverletzendes Verhalten ohne suizidale Absicht					
Anzahl der Patientinnen	0	0	NA	NA	NA

Hinweis 1: Für Selbstmordgedanken (1 – 5) oder suizidales Verhalten (6 – 10) wird eine Antwort mit „Ja" gewertet, wenn bei der Visite eine Frage zu Selbstmordgedanken oder suizidalem Verhalten mit „Ja" beantwortet wird. Die Kategorie Selbstmordgedanken oder suizidales Verhalten umfasst Patientinnen, die entweder bei Selbstmordgedanken (1 – 5) oder bei suizidalem Verhalten (6 – 10) gezählt werden.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

An Tag 15 lag kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor.

Tabelle 4-80: Ergebnisse für C-SSRS bei der letzten Bewertung nach Baseline aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

	Zuranolon+PiT	PiT	Zuranolon+PiT vs. PiT		
N	176	171	**RR Effektmaß [95 %-KI] p-Wert**	**OR Effektmaß [95 %-KI]**	**ARR Effektmaß [95 %-KI]**
C-SSRS Selbstmordgedanken (Frage 1 – 5)					
Anzahl der Patientinnen	4 (2,3)	9 (5,3)	0,43 [0,135; 1,367] 0,1525	0,42 [0,126; 1,377]	-0,03 [-0,071; 0,010]

	Zuranolon+PiT	PiT	Zuranolon+PiT vs. PiT		
N	176	171	**RR Effektmaß [95 %-KI] p-Wert**	**OR Effektmaß [95 %-KI]**	**ARR Effektmaß [95 %-KI]**
C-SSRS Suizidales Verhalten (Frage 6 – 10)					
Anzahl der Patientinnen	1 (0,6)	0	2,90 [0,119; 70,639] 0,5138	2,91 [0,118; 72,033]	0,01 [-0,010; 0,021]
C-SSRS Selbstmordgedanken oder suizidales Verhalten (Frage 1 – 10)					
Anzahl der Patientinnen	4 (2,3)	9 (5,3)	0,43 [0,135; 1,367] 0,1525	0,42 [0,126; 1,377]	-0,03 [-0,071; 0,010]
C-SSRS Selbstverletzendes Verhalten ohne suizidale Absicht					
Anzahl der Patientinnen	0	1 (0,6)	0,32 [0,013; 7,849] 0,4867	0,32 [0,013; 7,911]	-0,01 [-0,022; 0,010]

Hinweis 1: Für Selbstmordgedanken (1 – 5) oder suizidales Verhalten (6 – 10) wird eine Antwort mit „Ja" gewertet, wenn bei der Visite eine Frage zu Selbstmordgedanken oder suizidalem Verhalten mit „Ja" beantwortet wird. Die Kategorie Selbstmordgedanken oder suizidales Verhalten umfasst Patientinnen, die entweder bei Selbstmordgedanken (1 – 5) oder bei suizidalem Verhalten (6 – 10) gezählt werden.

Hinweis 2: Die letzte Bewertung während der Studie ist definiert als die letzte Bewertung nach Baseline, die nach der ersten Dosis der Studienmedikation durchgeführt wurde.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Zur letzten Bewertung lag kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor.

4.3.1.3.1.4.2 UE – RCT

Tabelle 4-81: Operationalisierung von UE

Studie	Operationalisierung
ROBIN	Ein unerwünschtes Ereignis (UE) wurde definiert als das Auftreten jeglicher unerwünschter medizinischen Erscheinung oder die Verschlechterung einer bereits bestehenden Erkrankung nach oder während der Verabreichung eines pharmazeutischen Produktes, unabhängig davon, ob ein kausaler Zusammenhang mit dem Produkt besteht. In klinischen Studien kann ein UE jede unvorteilhafte medizinische Erscheinung umfassen, die zu jedem Zeitpunkt auftritt, einschließlich der Baseline- oder Washout-Phasen, selbst wenn keine Studienmedikation verabreicht wurde. Unerwünschte Ereignisse, die spontan von den Patientinnen berichtet wurden und/oder als Antwort auf eine offene Frage des Studienpersonals oder durch Beobachtung festgestellt wurden, wurden während der Studie am Prüfzentrum dokumentiert. Klinisch signifikante Veränderungen von Laborwerten, Blutdruck und Puls mussten nicht als unerwünschte Ereignisse gemeldet werden, es sei denn, sie erforderten eine korrigierende medizinische Maßnahme durch das Prüfpersonal, stellten ein schwerwiegendes unerwünschtes Ereignis dar oder führten zum Abbruch der Verabreichung der Studienmedikation. Informationen zu unerwünschten Ereignissen wurden ab der Unterzeichnung der ICF bis zum letzten Besuch der Studie für die jeweilige Patientin erhoben. Unerwünschte Ereignisse, die nach der ersten Verabreichung der Studienmedikation auftreten, werden als Treatment Emergent Adverse Event (TEAE) bezeichnet. Die Kodierung der UE erfolgte für den CSR gemäß SAP nach MedDRA-Terminologie der Version 19 oder höher mit entsprechenden Bezeichnungen für SOC und PT. Im Rahmen der Analysen für das Dossier wurden eine einheitliche Hochstufung auf die Version 24.0 basierend auf den Master MedDRA-Datensätzen für alle AE Lower-Level-Terms (LLT) vorgenommen. Für jedes unerwünschte Ereignis bewertete und dokumentierte das Prüfpersonal den Beginn (Datum und Uhrzeit), die Behebung oder das klinische Plateau (Datum und Uhrzeit), den Schweregrad, die Kausalität, die ergriffenen Maßnahmen, das Ergebnis und ob das Ereignis zum Abbruch der Studie durch die Patientin geführt hat. Der Schweregrad wurde nach folgender Skala bewertet: - Mild: Wahrnehmung eines Zeichens oder Symptoms, aber leicht tolerierbar. - Moderat: Beschwerden, die ausreichen, um normale Aktivitäten zu beeinträchtigen. - Schwer: Symptome, die die Fähigkeit zur Durchführung normaler Aktivitäten unmöglich machen. Ein SUE ist ein UE, das während der Studie und bei jeglicher Dosis der Studienmedikation auftrat und eines oder mehrere der folgenden Kriterien erfüllte: - Es führte zum Tod. - Es war unmittelbar lebensbedrohlich. - Es erforderte einen stationären Krankenhausaufenthalt oder eine Verlängerung eines bestehenden Krankenhausaufenthalts. - Es führte zu einer anhaltenden oder signifikanten Behinderung oder Arbeitsunfähigkeit. - Es führte zu einer angeborenen Anomalie oder einem Geburtsfehler. - Es handelte sich um ein wichtiges medizinisches Ereignis, das die Patientin gefährden oder eine medizinische Intervention erfordern könnte, um eines der oben genannten Ergebnisse zu verhindern.

	Erhebende Person und Erhebungszeitpunkte Unerwünschte Ereignisse, die spontan von den Patientinnen berichtet wurden und/oder als Antwort auf eine offene Frage des Studienpersonals oder durch Beobachtung festgestellt wurden, wurden während der Studie am Prüfzentrum dokumentiert. Dargestellte Analysen: Folgende Analysen werden für die Sicherheits-Population der Studie ROBIN in Anhang 4-H dargestellt: - Gesamtraten (inkl. Kaplan-Meier-Plots) o Anzahl der Patientinnen mit mindestens einem UE o Anzahl der Patientinnen mit mindestens einem schweren UE o Anzahl der Patientinnen mit mindestens einem SUE o Anzahl der Patientinnen mit Therapieabbruch aufgrund von UE - UE nach SOC und PT (gemäß MedDRA) o Jegliche UE, analysiert nach SOC und PT (alle SOC/PT, die bei ≥ 10 % der Patientinnen in mindestens einem Arm oder bei ≥ 10 Patientinnen und bei ≥ 1 % der Patientinnen in mindestens einem Arm auftraten) o Auftreten von schweren UE nach SOC und PT (alle SOC/PT, die bei ≥ 5 % der Patientinnen in mindestens einem Arm auftraten oder bei ≥ 10 Patientinnen und bei ≥ 1 % der Patientinnen in mindestens einem Arm auftraten) o Auftreten von SUE nach SOC und PT (alle SOC/PT, die bei ≥ 5 % der Patientinnen in mindestens einem Arm auftraten oder bei ≥ 10 Patientinnen und bei ≥ 1 % der Patientinnen in mindestens einem Arm auftraten)
SKYLARK	Die Definition und Erhebung der UE entspricht der Definition in der ROBIN-Studie und weicht nur insofern ab, dass eine höhere MedDRA Version ≥ 23 genutzt wurde. Im Rahmen der Analysen für das Dossier wurden eine einheitliche Hochstufung auf die Version 24.0 basierend auf den Master MedDRA-Datensätzen für alle AE Lower-Level-Terms (LLT) vorgenommen Dargestellte Analysen: Folgende Analysen werden für die Sicherheits-Population der Studie SKYLARK in Anhang 4-I dargestellt: - Siehe ROBIN.
Meta-Analyse	Die Definition der UE und die Zeitpunkte der Erhebung ihre Schweregrade entspricht der Beschreibung für ROBIN. Dargestellte Analysen: Folgende Analysen werden für die gepoolte Sicherheits-Population beider Studien nachfolgend dargestellt: Gesamtraten Anzahl der Patientinnen mit mindestens einem UE Anzahl der Patientinnen mit mindestens einem schweren UE Anzahl der Patientinnen mit mindestens einem SUE Anzahl der Patientinnen mit Therapieabbruch aufgrund von UE UE nach SOC und PT (gemäß MedDRA) Jegliche UE, analysiert nach SOC und PT (alle SOC/PT, die bei ≥ 10 % der Patientinnen in mindestens einem Arm oder bei ≥ 10 Patientinnen und bei ≥ 1 % der Patientinnen in mindestens einem Arm auftraten) Auftreten von schweren UE nach SOC und PT (alle SOC/PT, die bei ≥ 5 % der Patientinnen in mindestens einem Arm auftraten oder bei ≥ 10 Patient*innen und bei ≥ 1 % der Patientinnen in mindestens einem Arm auftraten)

	Auftreten von SUE nach SOC und PT (alle SOC/PT, die bei ≥ 5 % der Patientinnen in mindestens einem Arm auftraten oder bei ≥ 10 Patient*innen und bei ≥ 1 % der Patientinnen in mindestens einem Arm auftraten) Statistische Methoden: Eine ausführliche Beschreibung der Methodik, die auf dem SKYLARK SAP beruht, findet sich in Abschnitt 4.2.5 bzw. in den Fußnoten. Zudem wurden Heterogenitätstests durchgeführt, die *post-hoc* definiert wurden und deren methodische Grundlage im gleichen Abschnitt zu finden sind.
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.	

Tabelle 4-82: Bewertung des Verzerrungspotenzials für Sicherheit in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
ROBIN	niedrig	ja	ja	ja	ja	niedrig
SKYLARK	niedrig	ja	ja	ja	ja	niedrig

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. In den Studien ROBIN und SKYLARK erfolgte die Erhebung jeglicher Unerwünschten Ereignisse der Patientinnen objektiv, unabhängig davon, ob das Ereignis in kausalem Zusammenhang mit der Verabreichung der Studienmedikation stand.

Die Erhebung und Klassifikation der UE erfolgten auf Basis allgemein anerkannter internationaler Standards. Somit ist eine ergebnisunabhängige Erhebung gewährleistet.

Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Tabelle 4-83: Ergebnisse für Anzahl der Patientinnen mit mindestens einem UE aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Anzahl der Patientinnen mit mindestens einem UE	Zuranolon+PiT	PiT
	N = 176	N = 171
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	112 (63,6)	90 (52,6)

Anzahl der Patientinnen mit mindestens einem UE	Zuranolon+PiT	PiT
	N = 176	N = 171
RR	1,21	
95 %-KI für RR	[1,009; 1,449]	
p-Wert für RR	0,0396	
OR	1,58	
95 %-KI für OR	[1,025; 2,420]	
ARR	0,11	
95 %-KI für ARR	[0,007; 0,213]	
Heterogenitätstest (p-Wert)	0,7330	
Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Es traten bei 63,6 % der Patientinnen im Zuranolon+PiT-Arm UE auf. Im Vergleichsarm trat mindestens 1 UE bei 52,6 % der Patientinnen auf.

Tabelle 4-84: Ergebnisse für Anzahl der Patientinnen mit mindestens einem UE nach SOC aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Anzahl der Patientinnen mit mindestens einem UE nach SOC	Zuranolon+PiT	PiT
Erkrankungen des Nervensystems	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	77 (43,8)	45 (26,3)
RR	1,66	
95 %-KI für RR	[1,230; 2,248]	
p-Wert für RR	0,0010	
OR	2,18	
95 %-KI für OR	[1,385; 3,423]	
ARR	0,17	
95 %-KI für ARR	[0,076; 0,273]	
Heterogenitätstest (p-Wert)	0,2479	
Erkrankungen des Gastrointestinaltrakts	**N = 176**	**N = 171**

Anzahl der Patientinnen mit mindestens einem UE nach SOC	**Zuranolon+PiT**	**PiT**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	27 (15,3)	28 (16,4)
RR	0,94	
95 %-KI für RR	[0,577; 1,522]	
p-Wert für RR	0,7922	
OR	0,93	
95 %-KI für OR	[0,520; 1,647]	
ARR	-0,01	
95 %-KI für ARR	[-0,087; 0,067]	
Heterogenitätstest (p-Wert)	0,9027	
Infektionen und parasitäre Erkrankungen	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	25 (14,2)	14 (8,2)
RR	1,73	
95 %-KI für RR	[0,934; 3,224]	
p-Wert für RR	0,0813	
OR	1,86	
95 %-KI für OR	[0,930; 3,707]	
ARR	0,06	
95 %-KI für ARR	[-0,006; 0,126]	
Heterogenitätstest (p-Wert)	0,8260	
Allgemeine Erkrankungen und Beschwerden am Verabreichungsort	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	16 (9,1)	8 (4,7)
RR	1,94	
95 %-KI für RR	[0,854; 4,422]	
p-Wert für RR	0,1133	
OR	2,04	
95 %-KI für OR	[0,848; 4,894]	

Anzahl der Patientinnen mit mindestens einem UE nach SOC	**Zuranolon+PiT**	**PiT**
ARR	0,04	
95 %-KI für ARR	[-0,009; 0,097]	
Heterogenitätstest (p-Wert)	0,1127	
Skelettmuskulatur-, Bindegewebs- und Knochenerkrankungen	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	14 (8,0)	9 (5,3)
RR	1,51	
95 %-KI für RR	[0,672; 3,399]	
p-Wert für RR	0,3179	
OR	1,56	
95 %-KI für OR	[0,655; 3,695]	
ARR	0,03	
95 %-KI für ARR	[-0,025; 0,079]	
Heterogenitätstest (p-Wert)	NB	
Psychiatrische Erkrankungen	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	12 (6,8)	14 (8,2)
RR	0,83	
95 %-KI für RR	[0,397; 1,749]	
p-Wert für RR	0,6288	
OR	0,82	
95 %-KI für OR	[0,368; 1,829]	
ARR	-0,01	
95 %-KI für ARR	[-0,069; 0,042]	
Heterogenitätstest (p-Wert)	0,0021	
Untersuchungen	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	11 (6,3)	13 (7,6)
RR	0,82	

Anzahl der Patientinnen mit mindestens einem UE nach SOC	**Zuranolon+PiT**	**PiT**
95 %-KI für RR	[0,379; 1,784]	
p-Wert für RR	0,6203	
OR	0,81	
95 %-KI für OR	[0,353; 1,862]	
ARR	-0,01	
95 %-KI für ARR	[-0,067; 0,040]	
Heterogenitätstest (p-Wert)	0,0093	
Erkrankungen der Atemwege, des Brustraums und Mediastinums	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	10 (5,7)	5 (2,9)
RR	1,94	
95 %-KI für RR	[0,678; 5,568]	
p-Wert für RR	0,2161	
OR	2,00	
95 %-KI für OR	[0,669; 5,978]	
ARR	0,03	
95 %-KI für ARR	[-0,015; 0,070]	
Heterogenitätstest (p-Wert)	NB	

UEs wurden mit MedDRA Version 24.0 kodiert.

Beinhaltet Patientinnen mit UE nach SOC ≥ 10 % in irgendeinem Arm ODER Anzahl der Patientinnen mit UE nach SOC mit mindestens 10 Patientinnen mit dem jeweiligen UE und ≥ 1 % in irgendeinem Arm. SOC wird in abnehmender Häufigkeit der Zuranolon+PiT-Spalte dargestellt.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

In der SOC „Erkrankungen des Nervensystems“ traten statistisch signifikant häufiger UE im Zuranolon+PiT-Arm auf als im Vergleichsarm. Häufige Nebenwirkungen aus dieser SOC sind für Zuranolon bekannt und werden auch in der Fachinformation genannt [1].

Für die weiteren dargestellten SOC mit UE bei mindestens 10 % in irgendeinem Arm bzw. der Anzahl der Patientinnen mit UE nach SOC mit mindestens 10 Patientinnen mit dem jeweiligen

UE und mindestens 1 % in einem Studienarm lag kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen vor.

Tabelle 4-85: Ergebnisse für Anzahl der Patientinnen mit mindestens einem UE nach PT aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Anzahl der Patientinnen mit mindestens einem UE nach PT	**Zuranolon+PiT**	**PiT**
Erkrankungen des Nervensystems/ Somnolenz	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	38 (21,6)	13 (7,6)
RR	2,84	
95 %-KI für RR	[1,569; 5,142]	
p-Wert für RR	0,0006	
OR	3,35	
95 %-KI für OR	[1,713; 6,539]	
ARR	0,14	
95 %-KI für ARR	[0,067; 0,213]	
Heterogenitätstest (p-Wert)	0,0369	
Erkrankungen des Nervensystems/ Schwindel	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	19 (10,8)	14 (8,2)
RR	1,32	
95 %-KI für RR	[0,683; 2,545]	
p-Wert für RR	0,4097	
OR	1,36	
95 %-KI für OR	[0,657; 2,802]	
ARR	0,03	
95 %-KI für ARR	[-0,035; 0,088]	
Heterogenitätstest (p-Wert)	0,9185	
Erkrankungen des Nervensystems/ Kopfschmerz	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	16 (9,1)	22 (12,9)
RR	0,71	

Anzahl der Patientinnen mit mindestens einem UE nach PT	**Zuranolon+PiT**	**PiT**
95 %-KI für RR	[0,384; 1,299]	
p-Wert für RR	0,2634	
OR	0,68	
95 %-KI für OR	[0,343; 1,339]	
ARR	-0,04	
95 %-KI für ARR	[-0,103; 0,028]	
Heterogenitätstest (p-Wert)	0,9189	
Erkrankungen des Nervensystems/ Sedierung	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	15 (8,5)	1 (0,6)
RR	14,57	
95 %-KI für RR	[1,946; 109,126]	
p-Wert für RR	0,0091	
OR	15,84	
95 %-KI für OR	[2,068; 121,283]	
ARR	0,08	
95 %-KI für ARR	[0,037; 0,122]	
Heterogenitätstest (p-Wert)	0,8830	
Erkrankungen des Gastrointestinaltrakts/ Durchfall	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	11 (6,3)	4 (2,3)
RR	2,67	
95 %-KI für RR	[0,868; 8,229]	
p-Wert für RR	0,0868	
OR	2,78	
95 %-KI für OR	[0,869; 8,918]	
ARR	0,04	
95 %-KI für ARR	[-0,003; 0,081]	
Heterogenitätstest (p-Wert)	NB	

Anzahl der Patientinnen mit mindestens einem UE nach PT	**Zuranolon+PiT**	**PiT**
Erkrankungen des Gastrointestinaltrakts/ Übelkeit	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 UE, n (%)	8 (4,5)	12 (7,0)
RR	0,65	
95 %-KI für RR	[0,272; 1,545]	
p-Wert für RR	0,3276	
OR	0,63	
95 %-KI für OR	[0,251; 1,584]	
ARR	-0,02	
95 %-KI für ARR	[-0,074; 0,024]	
Heterogenitätstest (p-Wert)	NB	

UEs wurden mit MedDRA Version 24.0 kodiert.

Beinhaltet Patientinnen mit UE nach PT ≥ 10 % in irgendeinem Arm ODER die Anzahl der Patientinnen mit UE nach PT mit mindestens 10 Patientinnen mit dem jeweiligen UE und ≥ 1% in irgendeinem Arm. PT wird in absteigender Häufigkeit der Zuranolon+PiT-Spalte dargestellt.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

Die PTs „Somnolenz“ und „Sedierung “ traten statistisch signifikant häufiger im Zuranolon+PiT-Arm auf als im Vergleichsarm.

Für den PT „Somnolenz“ zeigte der Heterogenitätstest eine statistische Signifikanz an, weshalb die Ergebnisse in den Einzelstudien überprüft wurden. Die Effekte sind gleichgerichtet, jedoch in der Studie SKYLARK mit stärker ausgeprägtem Unterschied zwischen den Behandlungsgruppen. Sowohl Somnolenz als auch Sedierung gehören zu den bekannten, in der Fachinformation genannten, sehr häufigen Nebenwirkungen, die jedoch zumeist transient zu Beginn der Behandlung auftreten und im Zusammenhang mit dem Wirkmechanismus von Zuranolon stehen [1].

4.3.1.3.1.4.3 Schwere UE – RCT

Tabelle 4-86: Ergebnisse für Anzahl der Patientinnen mit mindestens einem schweren UE aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Anzahl der Patientinnen mit mindestens einem schweren UE	**Zuranolon+PiT**	**PiT**
Grad 3/ 4/ 5	**N = 176**	**N = 171**
Anzahl der Patientinnen mit ≥ 1 schweren UE, n (%)	6 (3,4)	4 (2,3)
RR	1,46	
95 %-KI für RR	[0,419; 5,074]	
p-Wert für RR	0,5540	
OR	1,47	
95 %-KI für OR	[0,408; 5,316]	
ARR	0,01	
95 %-KI für ARR	[-0,024; 0,046]	
Heterogenitätstest (p-Wert)	NB	
UEs wurden mit MedDRA Version 24.0 kodiert. Der Schweregrad der UEs basiert auf der erfassten Schwere der UEs: Mild wird als Grad 1, Moderat als Grad 2 und Schwer als Grad 3 oder höher klassifiziert. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Bezüglich des Auftretens von schweren UE zeigte sich kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen.

Es gab keine Daten, die den Kriterien entsprachen: Patientinnen mit schweren UE nach SOC oder PT ≥ 5 % in irgendeinem Arm ODER die Anzahl der Patientinnen mit UE nach SOE oder PT mit mindestens 10 Patientinnen mit der jeweiligen UE und ≥ 1 % in irgendeinem Arm.

4.3.1.3.1.4.4 SUE – RCT

Tabelle 4-87: Ergebnisse für Anzahl der Patientinnen mit mindestens einem SUE aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Anzahl der Patientinnen mit mindestens einem SUE	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 171**
Anzahl der Patientinnen mit ≥ 1 SUE, n (%)	3 (1,7)	1 (0,6)
RR	2,91	
95 %-KI für RR	[0,306; 27,748]	
p-Wert für RR	0,3521	
OR	2,95	
95 %-KI für OR	[0,304; 28,622]	
ARR	0,01	
95 %-KI für ARR	[-0,011; 0,033]	
Heterogenitätstest (p-Wert)	NB	
UEs wurden mit MedDRA Version 24.0 kodiert. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Bezüglich des Auftretens von SUE zeigte sich kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen.

Es gab keine Daten, die den Kriterien entsprachen: Patientinnen mit SUE nach SOC oder PT ≥ 5 % in irgendeinem Arm ODER die Anzahl der Patientinnen mit SUE nach SOC oder PT mit mindestens 10 Patientinnen mit der jeweiligen UE und ≥ 1 % in irgendeinem Arm.

4.3.1.3.1.4.5 Therapieabbruch aufgrund von UE – RCT

Tabelle 4-88: Ergebnisse für Therapieabbruch aufgrund von UE aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Anzahl der Patientinnen mit Therapieabbruch aufgrund von UE	**Zuranolon+PiT**	**PiT**
	N = 176	**N = 171**
Anzahl der Patientinnen mit ≥ 1 Therapieabbruch aufgrund von UE, n (%)	5 (2,8)	2 (1,2)
RR	2,43	
95 %-KI für RR	[0,478; 12,351]	
p-Wert für RR	0,2848	
OR	2,47	
95 %-KI für OR	[0,473; 12,911]	
ARR	0,02	
95 %-KI für ARR	[-0,013; 0,046]	
Heterogenitätstest (p-Wert)	NB	
UEs wurden mit MedDRA Version 24.0 kodiert. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Im Zuranolon+PiT-Arm brachen 2,8 % der Patientinnen die Therapie aufgrund eines UE ab. Im Vergleichsarm brachen 1,2 % der Patientinnen die Therapie wegen eines UE ab.

4.3.1.3.2 Subgruppenanalysen – RCT

Für die Darstellung der Ergebnisse aus Subgruppenanalysen gelten die gleichen Anforderungen wie für die Darstellung von Ergebnissen aus Gesamtpopulationen in Abschnitt 4.3.1.3.1.1. [19]

Darüber hinaus sind folgende Kriterien zu berücksichtigen:

- Subgruppenanalysen sind nur für die Merkmale (zum Beispiel Alter) durchzuführen, bei denen die resultierenden Subgruppen jeweils mindestens zehn Patienten umfassen.
- Subgruppenanalysen sind für binäre Ereignisse je Merkmal nur dann durchzuführen, wenn in einer der Subgruppen mindestens zehn Ereignisse aufgetreten sind.
- Für Überlebenszeitanalysen müssen Kaplan-Meier-Kurven zu den einzelnen Subgruppen nur für Subgruppenanalysen mit statistisch signifikantem Interaktionsterm ($p < 0,05$) dargestellt werden.
- Ergebnisse zu UE nach SOC und PT müssen nur dargestellt werden, wenn das jeweilige Ergebnis für die Gesamtpopulation statistisch signifikant ist. Zu den UE-Gesamtraten (UE, schwere UE, SUE und Abbrüche wegen UE) müssen Subgruppenanalysen unabhängig vom Vorliegen statistischer Signifikanz in der Gesamtpopulation dargestellt werden.
- Bei Vorliegen mehrerer Studien und Durchführung von Meta-Analysen zu diesen Studien gelten die zuvor genannten Kriterien für die jeweilige Meta-Analyse, nicht für die Einzelstudien.
- Für Studien des pharmazeutischen Unternehmers sind entsprechende Analysen für alle benannten Effektmodifikatoren zu allen relevanten Endpunkten nach den zuvor genannten Kriterien vorzulegen und daher gegebenenfalls posthoc durchzuführen.
- Wird für die Nutzenbewertung nur die Teilpopulation einer Studie herangezogen (zum Beispiel wegen Zulassungsbeschränkungen, aufgrund von durch den Gemeinsamen Bundesausschuss bestimmte Teilpopulationen), so gelten die genannten Kriterien für diese Teilpopulation, und die Subgruppenanalysen sind für die Teilpopulation und nicht für die Gesamtpopulation der Studie durchzuführen.
- Subgruppenanalysen, bei denen der Interaktionsterm nicht statistisch signifikant ist, können auch in einem separaten Anhang des vorliegenden Modul 4 dargestellt werden. Dabei kann die Ausgabe der Statistik-Software unverändert verwendet werden, sofern diese alle notwendigen Angaben enthält. Eine ausschließliche Darstellung in Modul 5 ist aber nicht ausreichend.

Beschreiben Sie die Ergebnisse von Subgruppenanalysen. Stellen Sie dabei zunächst tabellarisch dar, zu welchen der in Abschnitt 4.2.5.5 genannten Effektmodifikatoren

[16] unbesetzt

Subgruppenanalysen zu den relevanten Endpunkten vorliegen, und ob diese a priori geplant und im Studienprotokoll festgelegt waren oder posthoc durchgeführt wurden.

Orientieren Sie sich an der beispielhaften Angabe in der ersten Tabellenzeile.

Tabelle 4-89 Matrix der durchgeführten Subgruppenanalysen

Endpunkt Studie	**Alter**	**Schwere der Erkrankung**	**Abstammung**	**Antidepressiva-zu Baseline**	**BMI**	**Beginn der PPD**	**PPD in der Familie**	**PPD in der Anamnese**	**Depression mit erhöhter Angst**
Gesamtmortalität									
Todesfälle	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Morbidität									
HAMD-17*	●	●	●	●	●	●	●	●	●
MADRS, EPDS, PHQ-9, HAM-A	○	○	○	○	○	○	○	○	○
Lebensqualität (Nur RCT ROBIN)	○	○	○	○	○	○	○	○	○
Sicherheit	○	○	○	○	○	○	○	○	○

* In beiden RCT-Studien waren Subgruppenanalysen für den primären Endpunkt HAMD-17 vorgesehen. Daher kann dieser auch für die Meta-Analyse als prädefiniert angesehen werden. Für alle anderen Endpunkte werden die Subgruppenanalysen als *post hoc* gekennzeichnet.

●: A priori geplante Subgruppenanalyse. ○: Posthoc durchgeführte Subgruppenanalyse. n.d.: Subgruppenanalyse nicht durchgeführt.

Stellen Sie anschließend in Tabelle 4-90 die Ergebnisse der Interaktionsterme für alle Subgruppenanalysen je Endpunkt in tabellarischer Form dar, und zwar für jede einzelne Studie separat. Kennzeichnen Sie dabei statistisch signifikante ($p < 0,05$) Interaktionsterme.

Tabelle 4-90: Ergebnis des Interaktionsterms der Subgruppenanalysen je Endpunkt für die Meta-Analyse

Endpunkt (Analyse)	**Alter**	**Schwere der Erkrank-ung**	**Abstam-mung**	**Antidepres-siva zu Baseline**	**BMI (a) (b) (c)**	**Beginn der PPD**	**PPD in der Familie**	**PPD in der Anamnese**	**Depression mit erhöhter Angst**
Gesamtmortalität									
Todesfälle	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.	k.A.
Morbidität									
HAMD-17-Ansprechen (Responder)	0,7893	0,1159	0,7112	0,3953	0,0544 0,7454 **0,0050** Tabelle 4-91	0,9854	0,7105	0,4546	0,7712
HAMD-17-Ansprechen (TTE)	0,7470	0,3274	0,6573	0,3456	0,8058 0,4150 0,2364	0,9643	0,4536	0,5674	0,9349
HAMD-17-Remission (Responder)	0,8968	0,4510	0,3231	0,9309	0,0536 0,8689 **0,0096** Tabelle 4-92	0,8366	0,5247	0,9457	0,6685
HAMD-17-Remission (TTE)	0,6820	0,9559	0,6759	0,7675	0,3927 0,6047 0,1153	0,8772	0,3600	0,8427	0,9763
MADRS-Ansprechen (Responder)	0,6907	0,1829	0,2759	0,4559	0,4399 0,6433 0,1305	0,6484	0,3165	0,5021	0,9245
MADRS-Ansprechen (TTE)	0,5305	0,1521	0,5299	0,4611	0,5074 0,9726 0,4076	0,5188	0,2215	0,1683	0,6537
MADRS-Remission (Responder)	0,3213	0,6235	0,1735	0,2610	0,1435 0,7677 **0,0221**	0,8252	0,6960	0,3236	0,9719

Endpunkt (Analyse)	**Alter**	**Schwere der Erkrank-ung**	**Abstam-mung**	**Antidepres-siva zu Baseline**	**BMI (a) (b) (c)**	**Beginn der PPD**	**PPD in der Familie**	**PPD in der Anamnese**	**Depression mit erhöhter Angst**
					Tabelle 4-93				
MADRS-Remission (TTE)	0,1245	0,5638	0,9481	0,1299	0,5658 0,5800 0,1809	0,8802	0,4304	0,1721	0,6016
EPDS-Ansprechen (Responder)	0,1005	0,5721	0,7875	0,5831	0,3792 0,2070 0,7341	0,7453	0,3889	0,9601	0,5690
EPDS-Ansprechen (TTE)	0,2629	0,8330	0,9447	0,6621	0,7132 0,7920 0,8781	0,6695	0,3367	0,8346	0,2925
PHQ-9-Ansprechen (Responder)	0,2405	0,3380	0,9161	0,5749	0,4674 0,3775 0,9096	0,1683	0,3614	0,6607	0,4688
PHQ-9-Ansprechen (TTE)	0,2568	0,5702	0,7285	0,8506	0,8923 0,8225 0,6772	0,3514	0,3614	0,8014	0,7255
HAM-A-Ansprechen (Responder)	0,1466	0,1978	0,5076	0,1049	0,1682 0,7543 0,0500○	0,8866○	0,9863	0,2693	0,1680
HAM-A-Ansprechen (TTE)	0,2158	0,2589	0,6115	0,1843	0,2326 0,5576 **0,0396** Tabelle 4-94	0,8159	0,6308	0,4879	0,9213
Lebensqualität									
SF-36 PCS-Ansprechen (Responder)	0,8519	0,1583	k.A.	0,8388	k.A.	k.A.	k.A.	k.A.	k.A.
SF-36 PCS-Ansprechen (TTE)	0,9839	0,5115	0,2240	0,7428	0,5053 0,7792	0,7856	0,6010	0,9905	k.A.

Endpunkt (Analyse)	**Alter**	**Schwere der Erkrank-ung**	**Abstam-mung**	**Antidepres-siva zu Baseline**	**BMI (a) (b) (c)**	**Beginn der PPD**	**PPD in der Familie**	**PPD in der Anamnese**	**Depression mit erhöhter Angst**
					0,2644				
SF-36 MCS-Ansprechen (Responder)	0,6564	0,0827	0,4123	0,1680	0,9671 0,5710 0,5664	0,1213	0,7735	0,5831	k.A.
SF-36 MCS-Ansprechen (TTE)	0,1271	0,6552	0,8129	0,0636	0,8242 0,8935 0,9031	0,2624	0,7180	0,7477	k.A.
SF-36 Ansprechen „Körperliche Schmerzen“ (Responder)	0,4810	0,0435	0,3673	0,9229	0,2272 0,4188 0,4956	0,9132	0,0871	0,7891	k.A.
SF-36 Ansprechen „Körperliche Schmerzen“ (TTE)	0,8354	0,4413	0,5332	0,8762	0,0746 0,2058 0,4139	0,8995	0,1032	0,9011	k.A.
SF-36 Ansprechen „Soziale Funktionsfähigkeit“ (Responder)	0,1894	**0,0182** Tabelle 4-95	0,4365	0,8798	0,0993 0,7657 0,0643	0,6775	0,5700	0,5607	k.A.
SF-36 Ansprechen „Soziale Funktionsfähigkeit“ (TTE)	0,2357	0,1611	0,4115	0,5789	0,1564 0,1381 0,9306	0,8168	0,3876	0,6635	k.A.
SF-36 Ansprechen „Psychisches Wohlbefinden“ (Responder)	0,7202	0,0842	0,6531	0,4569	0,4533 0,6886 0,1624	0,2206	0,4189	0,4385	k.A.
SF-36 Ansprechen „Psychisches Wohlbefinden“ (TTE)	0,3366	0,5922	0,7326	0,3062	0,7956 0,9448 0,8100	0,4544	0,1849	0,1486	k.A.
BIMF-Gesamtwert-	05672	**0,0460**	0,2594	0,3130	0,7804	0,1820	0,2852	0,3302	k.A.

Endpunkt (Analyse)	**Alter**	**Schwere der Erkrank-ung**	**Abstam-mung**	**Antidepres-siva zu Baseline**	**BMI (a) (b) (c)**	**Beginn der PPD**	**PPD in der Familie**	**PPD in der Anamnese**	**Depression mit erhöhter Angst**
Ansprechen (Responder)		Tabelle 4-96			0,8584 0,8736				
BIMF-Gesamtwert - Ansprechen (TTE)	0,5662	0,3352	**0,0407** Tabelle 4-97	0,4783	0,5878 0,7743 0,7183	0,0711	0,2804	0,7903	k.A.
BIMF Ansprechen „Kompetenz der Mutter“ (Responder)	0,2338	0,2223	0,0892	0,4340	0,2679 0,5156 0,4894	0,7906	0,9940	0,7300	k.A.
BIMF Ansprechen „Kompetenz der Mutter“ (TTE)	0,1662	0,2407	**0,0265** Tabelle 4-98	0,8668	0,5715 0,3071 0,6373	0,8913	0,9293	0,4008	k.A.
BIMF Ansprechen „Bedürfnisse der Mutter“ (Responder)	0,4535	0,2066	0,2544	0,8692	0,4882 0,9559 0,3988	0,8924	0,7200	0,1251	k.A.
BIMF Ansprechen „Bedürfnisse der Mutter“ (TTE)	0,5604	0,6169	0,0573	0,7003	0,9306 0,6535 0,5159	**0,0176** Tabelle 4-99	0,9344	0,5099	k.A.
Sicherheit									
UE Gesamtrate	0,3402	0,5897○	0,1246	0,1563	0,9097 0,4109 0,5114	0,5737	0,1908	0,7370	0,3535
SOC „Erkrankungen des Nervensystems“	0,2751	0,6566	0,5249	0,7755	0,8897 0,6102 0,7479	0,4126	0,0907	0,9681	0,4147
PT „Somnolenz“	0,2287	0,3742	0,7226	0,2233	0,8101 0,8538 0,6181	0,0703	0,5056	0,7191	0,3976
PT „Sedierung“	0,4475	k.A.	0,7763	k.A	0,8539 0,4976	0,4011	0,9667	0,9261	0,5474

Endpunkt (Analyse)	**Alter**	**Schwere der Erkrank-ung**	**Abstam-mung**	**Antidepres-siva zu Baseline**	**BMI (a) (b) (c)**	**Beginn der PPD**	**PPD in der Familie**	**PPD in der Anamnese**	**Depression mit erhöhter Angst**
					0,4642				
Schwere UE	k.A	k.A	k.A	k.A	k.A	k.A	k.A	k.A	k.A
SUE	k.A	k.A	k.A	k.A	k.A	k.A	k.A	k.A	k.A

k.A.: keine Angabe.

(a), (b) und (c) Repräsentieren p-Werte der Interaktionstests mit dem Vergleich verschiedener BMI-Gruppen: ≤ 24,9 vs. 25 – 29,9; ≤ 24,9 vs. ≥ 30 und 25 – 29,9 vs. ≥ 30.

Geforderte Subgruppenanalysen mit statistisch nicht signifikantem Interaktionswert finden sich in Anhang 4-G.

Quelle: [72]

Stellen Sie schließlich alle Subgruppenergebnisse dar.

Sofern eine Effektmodifikation für mehr als ein Subgruppenmerkmal vorliegt, kann eine Untersuchung auf eine Wechselwirkung höherer Ordnung sinnvoll sein. Dies gilt insbesondere dann, wenn diese Effektmodifikation konsistent über mehrere Endpunkte besteht. Zur Interpretation der Ergebnisse sollte dann für diese Endpunkte zusätzlich eine Subgruppenanalyse durchgeführt werden, die die Merkmale mit Effektmodifikation kombiniert. Beispiel: Für die Endpunkte Mortalität, gesundheitsbezogene Lebensqualität und schwere unerwünschte Ereignisse liegt sowohl für das Merkmal Geschlecht (mit den Ausprägungen „weiblich" und „männlich") als auch für das Merkmal Schweregrad (mit den Ausprägungen „niedrig" und „hoch") eine Effektmodifikation vor. Die zusätzliche Subgruppenanalyse erfolgt dann für die drei genannten Endpunkte für das kombinierte Merkmal Geschlecht/Schweregrad mit den vier Ausprägungen weiblich/niedrig, weiblich/hoch, männlich/niedrig und männlich/hoch.

Sofern die vorliegenden Studien beziehungsweise Daten für eine Meta-Analyse medizinisch und methodisch geeignet sind, fassen Sie die Ergebnisse mithilfe einer Meta-Analyse quantitativ zusammen und stellen Sie die Ergebnisse der Meta-Analyse (als Forest-Plot) dar.

Beschreiben Sie die Ergebnisse zusammenfassend. Begründen Sie Ihr Vorgehen, wenn Sie keine Meta-Analyse durchführen beziehungsweise wenn Sie nicht alle Studien in die Meta-Analyse einschließen.

Sofern Informationen beziehungsweise Ergebnisse zu durchgeführten Subgruppenanalysen für den jeweiligen Endpunkt im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU Dossiers zu verweisen.

Tabelle 4-91: Ergebnisse für Subgruppe BMI im Endpunkt HAMD-17-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

HAMD-17-Ansprechen (50 %-Reduktion von Baseline)	**Zuranolon + PiT**	**PiT**
BMI (kg/ m²) Gruppe 2: < 24,9		
Anzahl der auswertbaren Patientinnen [1]	33	43
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	9 (27,3)	10 (23,3)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	0	2 (4,7)
RR	1,274	
95 %-KI für das RR	[0,610; 2,661]	

HAMD-17-Ansprechen (50 %-Reduktion von Baseline)	**Zuranolon + PiT**	**PiT**
p-Wert für das RR	0,5195	
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	13 (39,4)	14 (32,6)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (6,1)	1 (2,3)
RR	1,253	
95 %-KI für das RR	[0,684; 2,297]	
p-Wert für das RR	0,4649	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	17 (51,5)	14 (32,6)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (6,1)	2 (4,7)
RR	1,621	
95 %-KI für das RR	[0,948; 2,772]	
p-Wert für das RR	0,0777	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	19 (57,6)	15 (34,9)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	5 (15,2)	6 (14,0)
RR	1,724	
95 %-KI für das RR	[1,066; 2,786]	
p-Wert für das RR	0,0263	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	18 (54,5)	22 (51,2)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (18,2)	3 (7,0)
RR	1,085	
95 %-KI für das RR	[0,716; 1,644]	
p-Wert für das RR	0,7003	
BMI (kg/ m²) Gruppe 2: 25-29,9		
Anzahl der auswertbaren Patientinnen [1]	45	54

HAMD-17-Ansprechen (50 %-Reduktion von Baseline)	Zuranolon + PiT	PiT
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	18 (40,0)	3 (5,6)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	1 (2,2)	0
RR	6,684	
95 %-KI für das RR	[2,110; 21,175]	
p-Wert für das RR	0,0012	
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	30 (66,7)	12 (22,2)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (4,4)	2 (3,7)
RR	2,834	
95 %-KI für das RR	[1,650; 4,866]	
p-Wert für das RR	0,0002	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	30 (66,7)	20 (37,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (4,4)	4 (7,4)
RR	1,708	
95 %-KI für das RR	[1,141; 2,557]	
p-Wert für das RR	0,0093	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	29 (64,4)	25 (46,3)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	3 (6,7)	6 (11,1)
RR	1,342	
95 %-KI für das RR	[0,936; 1,925]	
p-Wert für das RR	0,1100	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	33 (73,3)	27 (50,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	5 (11,1)	6 (11,1)

HAMD-17-Ansprechen (50 %-Reduktion von Baseline)	**Zuranolon + PiT**	**PiT**
RR	1,399	
95 %-KI für das RR	[1,020; 1,920]	
p-Wert für das RR	0,0373	
BMI (kg/ m²) Gruppe 2: ≥ 30		
Anzahl der auswertbaren Patientinnen [1]	95	76
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	29 (30,5)	19 (25,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	1 (1,1)	1 (1,3)
RR	1,214	
95 %-KI für das RR	[0,743; 1,983]	
p-Wert für das RR	0,4389	
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	52 (54,7)	31 (40,8)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (2,1)	1 (1,3)
RR	1,339	
95 %-KI für das RR	[0,974; 1,841]	
p-Wert für das RR	0,0722	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	59 (62,1)	36 (47,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	3 (3,2)	4 (5,3)
RR	1,299	
95 %-KI für das RR	[0,981; 1,722]	
p-Wert für das RR	0,0683	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	54 (56,8)	36 (47,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	8 (8,4)	5 (6,6)
RR	1,200	

HAMD-17-Ansprechen (50 %-Reduktion von Baseline)	Zuranolon + PiT	PiT
95 %-KI für das RR	[0,901; 1,598]	
p-Wert für das RR	0,2129	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	55 (57,9)	36 (47,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (6,3)	10 (13,2)
RR	1,211	
95 %-KI für das RR	[0,908; 1,616]	
p-Wert für das RR	0,1933	
Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert des HAMD-17, der Studie, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Studie sowie der Interaktion zwischen Zeitpunkten und Behandlung mit einer austauschbaren Kovarianzstruktur Log-Link-Funktion und Poisson-Verteilung Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien. [1] Anzahl der Patientinnen mit nicht fehlenden HAMD-17-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit HAMD-17-Ansprechen (≥ 50 % Reduktion gegenüber Baseline). Fehlende Werte wurden als „kein Ansprechen“ imputiert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

In den Gruppen der Patientinnen mit einem BMI < 24,9 bzw. von 25 – 29,9 sprachen an den Tagen 21 bzw. 3, 8, 15 und 45 statistisch signifikant mehr Patientinnen nach HAMD-17 im Zuranolon+PiT Arm als im Vergleichsarm an. In der Gruppe der Patientinnen mit einem BMI von ≥ 30 gab es keinen statistisch signifikanten Behandlungsunterschied.

Tabelle 4-92: Ergebnisse für Subgruppe BMI im Endpunkt HAMD-17-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

HAMD-17-Remission (Gesamtwert ≤ 7)	Zuranolon + PiT	PiT
BMI (kg/ m²) Gruppe 2: < 24,9		
Anzahl der auswertbaren Patientinnen [1]	33	43
Tag 3		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	4 (12,1)	3 (7,0)

HAMD-17-Remission (Gesamtwert ≤ 7)	**Zuranolon + PiT**	**PiT**
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	0	2 (4,7)
RR	1,773	
95 %-KI für das RR	[0,394; 7,973]	
p-Wert für das RR	0,4552	
Tag 8		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	6 (18,2)	7 (16,3)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (6,1)	1 (2,3)
RR	1,144	
95 %-KI für das RR	[0,418; 3,130]	
p-Wert für das RR	0,7930	
Tag 15		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	9 (27,3)	9 (20,9)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (6,1)	2 (4,7)
RR	1,344	
95 %-KI für das RR	[0,614; 2,942]	
p-Wert für das RR	0,4599	
Tag 21		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	8 (24,2)	7 (16,3)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	5 (15,2)	6 (14,0)
RR	1,532	
95 %-KI für das RR	[0,629; 3,732]	
p-Wert für das RR	0,3478	
Tag 45		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	13 (39,4)	13 (30,2)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (18,2)	3 (7,0)
RR	1,348	
95 %-KI für das RR	[0,743; 2,444]	

HAMD-17-Remission (Gesamtwert ≤ 7)	**Zuranolon + PiT**	**PiT**
p-Wert für das RR	0,3255	
BMI (kg/ m²) Gruppe 2: 25-29,9		
Anzahl der auswertbaren Patientinnen [1]	45	54
Tag 3		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	9 (20,0)	1 (1,9)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	1 (2,2)	0
RR	9,341	
95 %-KI für das RR	[1,364; 63,987]	
p-Wert für das RR	0,0228	
Tag 8		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	16 (35,6)	4 (7,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (4,4)	2 (3,7)
RR	4,172	
95 %-KI für das RR	[1,526; 11,403]	
p-Wert für das RR	0,0054	
Tag 15		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	17 (37,8)	4 (7,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (4,4)	4 (7,4)
RR	4,903	
95 %-KI für das RR	[1,784; 13,475]	
p-Wert für das RR	0,0021	
Tag 21		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	22 (48,9)	12 (22,2)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	3 (6,7)	6 (11,1)
RR	2,081	
95 %-KI für das RR	[1,163; 3,725]	
p-Wert für das RR	0,0136	

HAMD-17-Remission (Gesamtwert ≤ 7)	**Zuranolon + PiT**	**PiT**
Tag 45		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	24 (53,3)	16 (29,6)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	5 (11,1)	6 (11,1)
RR	1,702	
95 %-KI für das RR	[1,047; 2,766]	
p-Wert für das RR	0,0320	
BMI (kg/ m²) Gruppe 2: ≥ 30		
Anzahl der auswertbaren Patientinnen [1]	95	76
Tag 3		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	9 (9,5)	5 (6,6)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	1 (1,1)	1 (1,3)
RR	1,430	
95 %-KI für das RR	[0,502; 4,073]	
p-Wert für das RR	0,5033	
Tag 8		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	19 (20,0)	14 (18,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (2,1)	1 (1,3)
RR	1,076	
95 %-KI für das RR	[0,585; 1,979]	
p-Wert für das RR	0,8142	
Tag 15		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	32 (33,7)	19 (25,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	3 (3,2)	4 (5,3)
RR	1,344	
95 %-KI für das RR	[0,833; 2,168]	
p-Wert für das RR	0,2262	
Tag 21		

HAMD-17-Remission (Gesamtwert ≤ 7)	**Zuranolon + PiT**	**PiT**
Anzahl der Patientinnen mit einer Remission, n (%) [2]	25 (26,3)	19 (25,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	8 (8,4)	5 (6,6)
RR	1,052	
95 %-KI für das RR	[0,633; 1,748]	
p-Wert für das RR	0,8459	
Tag 45		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	38 (40,0)	17 (22,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (6,3)	10 (13,2)
RR	1,771	
95 %-KI für das RR	[1,095; 2,864]	
p-Wert für das RR	0,0198	
Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert des HAMD-17, der Studie, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Studie sowie der Interaktion zwischen Zeitpunkten und Behandlung mit einer austauschbaren Kovarianzstruktur Log-Link-Funktion und Poisson-Verteilung. Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien. [1] Anzahl der Patientinnen mit nicht fehlenden HAMD-17-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit HAMD-17-Remission (Gesamtwert ≤ 7%). Fehlende Werte wurden als „kein Ansprechen“ imputiert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

In den Gruppen der Patientinnen mit einem BMI < 24,9 lag zu keinem Zeitpunkt ein statistisch signifikanter Behandlungsunterschied vor. In der Gruppe der Patientinnen mit einem BMI von 25 – 29,9 erreichten zu allen Zeitpunkten statistisch signifikant mehr Patientinnen eine HAMD-17-Remission im Zuranolon+PiT Arm als im Vergleichsarm. In der Gruppe der Patientinnen mit einem BMI von ≥ 30 lag ein statistisch signifikanter Vorteil zugunsten der Zuranolon+PiT-Behandlung nur zu Tag 45 vor.

Tabelle 4-93: Ergebnisse für Subgruppe BMI im Endpunkt Zeit bis zur ersten MADRS-Remission aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Zeit bis zur ersten MADRS-Remission (Gesamtwert ≤ 10)	**Zuranolon + PiT**	**PiT**
BMI (kg/ m²) Gruppe 2: ≤ 24,9		
Anzahl der auswertbaren Patientinnen [1]	33	43

Zeit bis zur ersten MADRS-Remission (Gesamtwert ≤ 10)	**Zuranolon + PiT**	**PiT**
Tag 8		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	7 (21,2)	9 (20,9)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (6,1)	1 (2,3)
RR	1,061	
95 %-KI für das RR	[0,428; 2,628]	
p-Wert für das RR	0,8982	
Tag 15		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	13 (39,4)	12 (27,9)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (6,1)	2 (4,7)
RR	1,480	
95 %-KI für das RR	[0,770; 2,845]	
p-Wert für das RR	0,2394	
Tag 45		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	15 (45,5)	14 (32,6)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (18,2)	3 (7,0)
RR	1,453	
95 %-KI für das RR	[0,836; 2,524]	
p-Wert für das RR	0,1851	
BMI (kg/ m²) Gruppe 2: 25-29,9		
Anzahl der auswertbaren Patientinnen [1]	45	54
Tag 8		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	19 (42,2)	8 (14,8)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (4,4)	2 (3,7)
RR	2,521	
95 %-KI für das RR	[1,232; 5,156]	
p-Wert für das RR	0,0113	
Tag 15		

Zeit bis zur ersten MADRS-Remission (Gesamtwert ≤ 10)	**Zuranolon + PiT**	**PiT**
Anzahl der Patientinnen mit einer Remission, n (%) [2]	22 (48,9)	9 (16,7)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (4,4)	4 (7,4)
RR	2,699	
95 %-KI für das RR	[1,399; 5,208]	
p-Wert für das RR	0,0031	
Tag 45		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	27 (60,0)	20 (37,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	5 (11,1)	6 (11,1)
RR	1,501	
95 %-KI für das RR	[0,991; 2,273]	
p-Wert für das RR	0,0553	
BMI (kg/ m²) Gruppe 2: ≥ 30		
Anzahl der auswertbaren Patientinnen [1]	95	75
Tag 8		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	24 (25,3)	21 (28,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (2,1)	1 (1,3)
RR	0,906	
95 %-KI für das RR	[0,551; 1,488]	
p-Wert für das RR	0,6958	
Tag 15		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	43 (45,3)	24 (32,0)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	4 (4,2)	4 (5,3)
RR	1,412	
95 %-KI für das RR	[0,953; 2,090]	
p-Wert für das RR	0,0851	
Tag 45		
Anzahl der Patientinnen mit einer Remission, n (%) [2]	40 (42,1)	24 (32,0)

Zeit bis zur ersten MADRS-Remission (Gesamtwert ≤ 10)	Zuranolon + PiT	PiT
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	6 (6,3)	10 (13,3)
RR	1,314	
95 %-KI für das RR	[0,879; 1,964]	
p-Wert für das RR	0,1836	
Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert des MADRS, der Studie, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Studie sowie der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur Kovarianzstruktur Log-Link-Funktion und Poisson-Verteilung. Die GEE-Analyse verwendete gemeinsame geplante Bewertungszeitpunkte in beiden Studien. [1] Anzahl der Patientinnen mit nicht fehlenden MADRS-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit einer MADRS-Remission (definiert als MADRS-Gesamtwert ≤ 10). Fehlende MADRS-Gesamtwerte bei der Untersuchung wurden als keine Remission imputiert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

In der Gruppe der Patientinnen mit einem BMI von 25 – 29,9 erreichten an Tag 8 und 15 statistisch signifikant mehr Patientinnen eine Remission gemessen nach MADRS im Zuranolon+PiT Arm als im Vergleichsarm. In den Gruppen der Patientinnen mit einem BMI < 24,9 bzw. ≥ 30 lag zu keinem Zeitpunkt ein statistisch signifikanter Behandlungsunterschied vor.

Tabelle 4-94: Ergebnisse für Subgruppe BMI im Endpunkt Zeit bis zum ersten HAM-A-Ansprechen aus der Meta-Analyse mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten HAM-A-Ansprechen	Zuranolon + PiT	PiT
BMI (kg/ m²) Gruppe 2: ≤ 24,9		
Anzahl der auswertbaren Patientinnen [1]	33	43
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	29 (87,9)	34 (79,1)
Zeit bis zum ersten HAM-A-Ansprechen (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	4,0 [3,0; 8,0]	4,0 [3,0; 8,0]
50. Perzentil - Median (95 %-KI)	8,0 [4,0; 15,0]	9,0 [5,0; 21,0]
75. Perzentil - Q3 (95 %-KI)	16,0 [8,0; 44,0]	22,0 [20,0; NB]
HR - Hazard Ratio [4]	1,489	

Zeit bis zum ersten HAM-A-Ansprechen	**Zuranolon + PiT**	**PiT**
95 %-KI für den HR	[0,8976; 2,4700]	
p-Wert für den HR	0,1232	
BMI (kg/ m²) Gruppe 2: 25-29,9		
Anzahl der auswertbaren Patientinnen [1]	45	54
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	38 (84,4)	44 (81,5)
Zeit bis zum ersten HAM-A-Ansprechen (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	3,0 [3,0; 4,0]	8,0 [3,0; 8,0]
50. Perzentil - Median (95 %-KI)	5,0 [4,0; 8,0]	8,5 [8,0; 16,0]
75. Perzentil - Q3 (95 %-KI)	15,0 [8,0; 21,0]	28,0 [16,0; 46,0]
HR - Hazard Ratio [4]	1,791	
95 %-KI für den HR	[1,0970; 2,9229]	
p-Wert für den HR	0,0198	
BMI (kg/ m²) Gruppe 2: ≥ 30		
Anzahl der auswertbaren Patientinnen [1]	95	76
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	84 (88,4)	62 (81,6)
Zeit bis zum ersten HAM-A-Ansprechen (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	3,0 [3,0; 4,0]	3,0 [3,0; 4,0]
50. Perzentil - Median (95 %-KI)	8,0 [4,0; 8,0]	8,0 [4,0; 9,0]
75. Perzentil - Q3 (95 %-KI)	15,0 [9,0; 43,0]	22,0 [15,0; NB]
HR - Hazard Ratio [4]	1,099	
95 %-KI für den HR	[0,7886; 1,5303]	
p-Wert für den HR	0,5784	

Zeit bis zum ersten HAM-A-Ansprechen	Zuranolon + PiT	PiT
Das Ansprechen nach HAM-A ist definiert als Reduktion um ≥ 8,4 Punkte (15 % der HAM-A-Skalenspannweite von 0 – 56) des Gesamtwerts von Baseline. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren HAM-A-Bewertung zensiert. [1] Anzahl der Patientinnen mit nicht fehlenden HAM-A-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit einer HAM-A-Ansprechrate. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. [3] Kaplan-Meier-Schätzer [4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Studie, Behandlung, Baselinewert des HAM-A-Gesamtwerts und Antidepressiva zu Baseline (Ja oder Nein) Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

In den Gruppen der Patientinnen mit einem BMI von 25 – 29,9 sprachen die Patientinnen im HAM-A statistisch signifikant schneller im Zuranolon+PiT Arm an als im Vergleichsarm. In den Gruppen der Patientinnen mit einem BMI < 24,9 und ≥ 30 lag kein statistisch signifikanter Behandlungsunterschied vor. Die entsprechenden Kaplan-Meier-Plots finden sich in Anhang 4-G.

Tabelle 4-95: Ergebnisse für Subgruppe „Schwere der Erkrankung“ im Endpunkt Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“ des SF-36 aus RCT mit dem zu bewertenden Arzneimittel

Ansprechen Subskala „Soziale Funktionsfähigkeit“ des SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	Zuranolon + PiT	PiT
HAMD-17 < 28		
Anzahl der auswertbaren Patientinnen [1]	30	22
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	18 (60,0)	16 (72,7)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (6,7)	0
RR	0,817	
95 %-KI für das RR	[0,555; 1,203]	
p-Wert für das RR	0,3049	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	25 (83,3)	17 (77,3)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (6,7)	0

Ansprechen Subskala „Soziale Funktionsfähigkeit“ des SF-36 (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon + PiT**	**PiT**
RR	1,067	
95 %-KI für das RR	[0,813; 1,402]	
p-Wert für das RR	0,6391	
HAMD-17 ≥ 28		
Anzahl der auswertbaren Patientinnen [1]	46	53
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	32 (69,6)	27 (50,9)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	0	2 (3,8)
RR	1,382	
95 %-KI für das RR	[1,009; 1,893]	
p-Wert für das RR	0,0440	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	37 (80,4)	32 (60,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	1 (2,2)	6 (11,3)
RR	1,348	
95 %-KI für das RR	[1,045; 1,739]	
p-Wert für das RR	0,0216	

Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, Baselinewert der Subskala, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur Log-Link-Funktion und Poisson-Verteilung.

[1] Anzahl der Patientinnen mit nicht fehlenden Subskala-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“ des SF-36 (Zunahme um ≥ 15 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 100 im Vergleich zu Baseline). Fehlende Werte wurden als „kein Ansprechen“ imputiert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

In der Gruppe der Patientinnen mit einem HAMD-17 Wert von < 28 gab es keinen statistisch signifikanten Behandlungsunterschied für das Ansprechen auf der Subskala „Soziale Funktionsfähigkeit“. In den Gruppen der Patientinnen mit einem HAMD-17 ≥ 28 lag ein

statistisch signifikanter Vorteil zugunsten des Zuranolon+PiT-Arms für das Ansprechen bezüglich „Sozialer Funktionsfähigkeit“ vor.

Tabelle 4-96: Ergebnisse für Subgruppe „Schwere der Erkrankung“ im Endpunkt BIMF-Gesamtwert-Ansprechen aus RCT mit dem zu bewertenden Arzneimittel

BIMF-Gesamtwert-Ansprechen (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon + PiT**	**PiT**
HAMD-17 < 28		
Anzahl der auswertbaren Patientinnen [1]	30	22
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	4 (13,3)	2 (9,1)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	4 (13,3)	2 (9,1)
RR	1,069	
95 %-KI für das RR	[0,215; 5,309]	
p-Wert für das RR	0,9347	
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	11 (36,7)	5 (22,7)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	1 (3,3)	0
RR	1,176	
95 %-KI für das RR	[0,486; 2,845]	
p-Wert für das RR	0,7187	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	14 (46,7)	6 (27,3)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (6,7)	0
RR	1,247	
95 %-KI für das RR	[0,549; 2,834]	
p-Wert für das RR	0,5974	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	12 (40,0)	14 (63,6)

BIMF-Gesamtwert-Ansprechen (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon + PiT**	**PiT**
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (6,7)	0
RR	0,458	
95 %-KI für das RR	[0,263; 0,798]	
p-Wert für das RR	0,0058	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	20 (66,7)	14 (63,6)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	2 (6,7)	0
RR	0,764	
95 %-KI für das RR	[0,476; 1,225]	
p-Wert für das RR	0,2635	
HAMD-17 ≥ 28		
Anzahl der auswertbaren Patientinnen [1]	46	53
Tag 3		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	7 (15,2)	5 (9,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	0	1 (1,9)
RR	1,623	
95 %-KI für das RR	[0,561; 4,697]	
p-Wert für das RR	0,3715	
Tag 8		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	12 (26,1)	14 (26,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	0	1 (1,9)
RR	0,994	
95 %-KI für das RR	[0,522; 1,891]	
p-Wert für das RR	0,9851	
Tag 15		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	21 (45,7)	14 (26,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	0	2 (3,8)

BIMF-Gesamtwert-Ansprechen (Zunahme um ≥ 15 % der Skalenspannweite)	**Zuranolon + PiT**	**PiT**
RR	1,739	
95 %-KI für das RR	[1,020; 2,966]	
p-Wert für das RR	0,0421	
Tag 21		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	23 (50,0)	21 (39,6)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	0	2 (3,8)
RR	1,270	
95 %-KI für das RR	[0,828; 1,948	
p-Wert für das RR	0,2737	
Tag 45		
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	29 (63,0)	23 (43,4)
Anzahl der Patientinnen mit fehlender Auswertung, n (%)	1 (2,2)	6 (11,3)
RR	1,462	
95 %-KI für das RR	[0,001; 2,136]	
p-Wert für das RR	0,0495	

Das verwendete Modell ist ein GEE-Modell für binäre Antworten, mit Faktoren für Antidepressiva zu Baseline, BIMF-Gesamtwert zu Baseline, der Behandlung, dem Bewertungszeitpunkt, der Interaktion zwischen Zeitpunkten und Behandlung mit einer unstrukturierten Kovarianzstruktur Log-Link-Funktion und Poisson-Verteilung.

[1] Anzahl der Patientinnen mit nicht fehlenden BIMF-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit BIMF-Ansprechen (Zunahme um ≥ 18 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 120 im Vergleich zu Baseline). Fehlende Werte wurden als „kein Ansprechen“ imputiert. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

In der Gruppe der Patientinnen mit einem HAMD-17 Wert von < 28 erreichten an Tag 21 statistisch mehr Patientinnen im Vergleichsarm ein Ansprechen im BIMF-Gesamtscore. In den Gruppen der Patientinnen mit einem HAMD-17 ≥ 28 lag ein statistisch signifikanter Vorteil zugunsten des Zuranolon+PiT-Arms für das Ansprechen im BIMF-Gesamtscore vor. Zu allen anderen Zeitpunkten gab es keinen statistisch signifikanten Behandlungsunterschied.

Tabelle 4-97: Ergebnisse für Subgruppe Abstammung im Endpunkt Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen aus RCT mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen	**Zuranolon + PiT**	**PiT**
Schwarz oder Afroamerikanisch		
Anzahl der auswertbaren Patientinnen [1]	31	32
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	20 (64,5)	22 (71,9)
Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	15,0 [8,0; 45,0]	8,0 [3,0; 22,0]
50. Perzentil - Median (95 %-KI)	45,0 [15,0; 57,0]	22,0 [15,0; 44,0]
75. Perzentil - Q3 (95 %-KI)	57,0 [46,0; NB]	45,0 [43,0; NB]
HR - Hazard Ratio [4]	0,584	
95 %-KI für den HR	[0,3106; 1,0985]	
p-Wert für den HR	0,0952	
Weiß		
Anzahl der auswertbaren Patientinnen [1]	44	40
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	32 (72,7)	24 (60,0)
Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	8,0 [4,0; 8,0]	8,0 [8,0; 22,0)
50. Perzentil - Median (95 %-KI)	15,0 [8,0; 43,0]	22,5 [15,0; NB]
75. Perzentil - Q3 (95 %-KI)	48,0 [16,0; NB]	NB [45,0; NB]
HR - Hazard Ratio [4]	1,418	
95 %-KI für den HR	[0,8315; 2,4184]	
p-Wert für den HR	0,1997	

Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen	Zuranolon + PiT	PiT
Das Ansprechen nach BIMF-Gesamtwert ist definiert als Reduktion um ≥ 18 Punkte (15 % der Skalenspannweite von 0 – 120) des Gesamtwerts von Baseline. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren HAM-A-Bewertung zensiert. [1] Anzahl der Patientinnen mit nicht fehlenden BIMF-Gesamtwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit einer BIMF-Ansprechrate. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. [3] Kaplan-Meier-Schätzer [4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Antidepressiva zu Baseline, Baselinewert des BIMF-Gesamtwerts und Behandlung. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Für das Merkmal Abstammung konnte für die Zeit bis zum ersten BIMF-Gesamtwert-Ansprechen in keiner Subgruppe ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen gezeigt werden. Die entsprechenden Kaplan-Meier-Plots finden sich in Anhang 4-H.

Tabelle 4-98: Ergebnisse für Subgruppe Abstammung im Endpunkt Zeit bis zum ersten Ansprechen nach -Subskala „Kompetenz der Mutter“ des BIMF aus RCT mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten Ansprechen Subskala „Kompetenz der Mutter“ des BIMF	Zuranolon + PiT	PiT
Schwarz oder Afroamerikanisch		
Anzahl der auswertbaren Patientinnen [1]	31	32
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	14 (45,2)	14 (43,8)
Zeit bis zum ersten Ansprechen Subskala „Kompetenz der Mutter“ des BIMF (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	43,0 [8,0; 47,0]	15,0 [4,0; 44,0]
50. Perzentil - Median (95 %-KI)	57,0 [45,0; NB]	45,0 [22,0; NB]
75. Perzentil - Q3 (95 %-KI)	87,0 [57,0; NB]	NB [NB; NB]
HR - Hazard Ratio [4]	0,490	
95 %-KI für den HR	[0,2148; 1,1155]	
p-Wert für den HR	0,0891	
Weiß		

Zeit bis zum ersten Ansprechen Subskala „Kompetenz der Mutter“ des BIMF	**Zuranolon + PiT**	**PiT**
Anzahl der auswertbaren Patientinnen [1]	44	40
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	30 (68,2)	19 (47,5)
Zeit bis zum ersten Ansprechen Subskala „Kompetenz der Mutter“ des BIMF (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	8,0 [4,0; 15,0)	15,0 [4,0; 23,0]
50. Perzentil - Median (95 %-KI)	15,0 [15,0; 22,0]	46,0 [22,0; NB]
75. Perzentil - Q3 (95 %-KI)	NB [22,0; NB]	NB [NB; NB]
HR - Hazard Ratio [4]	1,496	
95 %-KI für den HR	[0,8351; 2,6782]	
p-Wert für den HR	0,1757	
Ansprechen für die Subskala ist definiert als Zunahme um ≥ 9,9 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 66. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert. [1] Anzahl der Patientinnen mit nicht fehlenden Subskalenwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert. [2] Anzahl der Patientinnen mit Ansprechen nach Subskalenwerten. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen. [3] Kaplan-Meier-Schätzer [4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Antidepressiva zu Baseline, Baselinewert der Subskala und Behandlung. Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis. Quelle: [72]		

Für das Merkmal Abstammung konnte für die Zeit bis zum ersten BIMF-Ansprechen auf der Subskala „Kompetenz der Mutter“ in keiner Subgruppe ein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen gezeigt werden. Die entsprechenden Kaplan-Meier-Plots finden sich in Anhang 4-H.

Tabelle 4-99: Ergebnisse für Subgruppe Beginn der PPD im Endpunkt Zeit bis zum ersten Ansprechen auf der Subskala „Bedürfnisse der Mutter“ des BIMF aus RCT mit dem zu bewertenden Arzneimittel

Zeit bis zum ersten Ansprechen Subskala „Bedürfnisse der Mutter“ des BIMF	**Zuranolon + PiT**	**PiT**
3. Trimester		
Anzahl der auswertbaren Patientinnen [1]	32	32
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	19 (59,4)	27 (84,4)

Zeit bis zum ersten Ansprechen Subskala „Bedürfnisse der Mutter“ des BIMF	**Zuranolon + PiT**	**PiT**
Zeit bis zum ersten Ansprechen Subskala „Bedürfnisse der Mutter“ des BIMF (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	6,0 [3,0; 8,0]	8,0 [3,0; 8,0]
50. Perzentil - Median (95 %-KI)	46,0 [8,0; 51,0]	9,0 [8,0; 22,0]
75. Perzentil - Q3 (95 %-KI)	NB [46,0; NB]	43,0 [15,0; NB]
HR - Hazard Ratio [4]	0,523	
95 %-KI für den HR	[0,2803; 0,9743]	
p-Wert für den HR	0,0411	
Postpartum		
Anzahl der auswertbaren Patientinnen [1]	44	43
Anzahl der Patientinnen mit einem Ansprechen, n (%) [2]	35 (79,5)	28 (65,1)
Zeit bis zum ersten Ansprechen Subskala „Bedürfnisse der Mutter“ des BIMF (Tage) [3]		
25. Perzentil - Q1 (95 %-KI)	8,0 [4,0; 8,0]	8,0 [4,0; 15,0)
50. Perzentil - Median (95 %-KI)	8,5 [8,0; 23,0]	15,0 [8,0; 49,0]
75. Perzentil - Q3 (95 %-KI)	46,0 [15,0; NB]	49,0 [43,0; NB]
HR - Hazard Ratio [4]	1,393	
95 %-KI für den HR	[0,8339; 2,3256]	
p-Wert für den HR	0,2056	

Ansprechen für die Subskala ist definiert als Zunahme um ≥ 6,3 Punkte entsprechend 15 % des gesamten Skalenbereichs 0 – 42. Patientinnen, die keine Responder sind, werden am Tag der letzten verfügbaren Bewertung zensiert.

[1] Anzahl der Patientinnen mit nicht fehlenden Subskalenwerten zu Baseline. Der Baselinewert wird als die letzte nicht fehlende Bewertung vor der ersten Dosis der Studienmedikation definiert.

[2] Anzahl der Patientinnen mit Ansprechen nach Subskalenwerten. Der Nenner für den Prozentsatz ist die Anzahl der auswertbaren Patientinnen.

[3] Kaplan-Meier-Schätzer

[4] Aus dem Cox-Proportional-Hazards-Modell mit Faktoren für Antidepressiva zu Baseline, Baselinewert der Subskala und Behandlung.

Für die Bedeutung von Abkürzungen siehe Abkürzungsverzeichnis.

Quelle: [72]

In der Gruppe der Patientinnen mit einem Beginn der PPD im 3. Trimester zeigte sich ein statistisch signifikant schnelleres Ansprechen auf der Subskala „Bedürfnisse der Mutter“ bei Patientinnen im Vergleichsarm. In der Gruppe der Patientinnen mit einem Beginn der PPD nach

der Geburt lag zwar ein numerischer Trend zugunsten von Zuranolon vor, jedoch wurde kein statistisch signifikanter Unterschied zwischen den Behandlungsgruppen erreicht. Die entsprechenden Kaplan-Meier-Plots finden sich in Anhang 4-H.

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.

4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche

Das Datum der Recherche soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen. Liegt die Recherche im EU-Dossier mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt, ist die Recherche zu aktualisieren. Die aktualisierte Recherche ist im Dossier darzustellen.

In den Fällen, in denen keine neuen Quellen identifiziert werden, kann auf das Rechercheergebnis im EU-Dossier verwiesen werden. Sofern neue Quellen identifiziert werden, sind die neuen Erkenntnisse entsprechend der Vorgaben aufzuarbeiten.

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 sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie analog Abschnitt 4.3.1.1 zur Verfügung (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms et cetera).*** *Benennen Sie sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie*

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Recherche*
- *Studien aus der Suche in Studienregistern/ Studienergebnisdatenbanken*
- *Studien aus der Suche auf der Internetseite des Gemeinsamen Bundesausschusses*
- *Resultierender Studienpool aus den einzelnen Suchschritten*

Sofern Angaben zur Informationsbeschaffung zu Studien für indirekte Vergleiche im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

4.3.2.1.2 Charakteristika der Studien für indirekte Vergleiche

Charakterisieren Sie nachfolgend die Studien, die für indirekte Vergleiche identifiziert wurden und bewerten Sie darüber hinaus deren Ähnlichkeit. Begründen Sie darauf basierend den Ein- beziehungsweise Ausschluss von Studien für die von Ihnen durchgeführten indirekten Vergleiche. Bewerten Sie das Verzerrungspotenzial der für indirekte Vergleiche herangezogenen Studien. ***Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2 und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.***

Sofern Informationen zu den Charakteristika der Studien für indirekte Vergleiche im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

4.3.2.1.3 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.

Sofern ein Überblick über die patientenrelevanten Endpunkte aus indirekten Vergleichen im EU-Dossier hinterlegt ist und dieser Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-100: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche

Studie	**<Mortalität>**	**<Gesundheits-bezogene Lebensqualität>**	**<Endpunkt>**	**<Endpunkt>**	**<Endpunkt>**
<Studie 1>	nein	ja	ja	ja	nein

4.3.2.1.3.1 <Endpunkt xxx> – 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 drei 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 4.3.1.3.1.1.**

Geben Sie für den im vorliegenden Abschnitt präsentierten Endpunkt einen Überblick über die in den Studien verfügbaren Vergleiche. Beispielhaft wäre folgende Darstellung denkbar:

Tabelle 4-101: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden

Anzahl Studien	**Studie**	**Intervention**	**<Vergleichs-therapie 1>**	**<Vergleichs-therapie 2>**	**<Vergleichs-therapie 3>**
1	<Studie 1>	●		●	●
2	<Studie 2>	●		●	
	<Studie 3>	●		●	
1	<Studie 4>		●	●	●
et cetera	et cetera	et cetera	et cetera		

Stellen Sie zusätzlich die Netzwerkstruktur des indirekten Vergleichs grafisch dar.

Sofern Informationen zur Netzwerkstruktur des indirekten Vergleichs im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

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

Sofern Informationen zur Operationalisierung des jeweiligen Endpunkts im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-102: Operationalisierung von <Endpunkt xxx>

Studie	**Operationalisierung**
<Studie 1>	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-F. 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-F.

Tabelle 4-103: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
<Studie 1>	<hoch/ niedrig>	<ja/nein/ unklar>	<ja/nein/ unklar>	<ja/nein/ unklar>	<ja/ nein>	<hoch/ niedrig>

Begründen Sie für jede Studie die abschließende Einschätzung.

Sofern eine Bewertung des Verzerrungspotenzials auf Studienebene im EU-Dossier hinterlegt ist und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen. Die obenstehende Tabelle 4-19 ist im Dossier unabhängig von den Angaben im EU-Dossier auszufüllen.

Nicht zutreffend.

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

Sofern Ergebnisse für den jeweiligen Endpunkt im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-104: Ergebnisse für <Endpunkt xxx> aus RCT für indirekte Vergleiche

Studie	Tabellarische Präsentation in geeigneter Form (Anforderungen siehe Erläuterung in Abschnitt 4.3.1.3.1.1)
<Studie 1>	

Nicht zutreffend.

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 der üblichen Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Stellen Sie die Ergebnisse der paarweisen Meta-Analysen dar. 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 inkonsistente Ergebnisse.*

Machen Sie darüber hinaus Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Sofern eine Darstellung der Ergebnisse des indirekten Vergleichs entsprechend der oben beschriebenen Vorgaben im EU-Dossier hinterlegt ist und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

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.

4.3.2.1.3.2 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 4.3.1.3.2.***

Sofern Informationen beziehungsweise Ergebnisse zu durchgeführten Subgruppenanalysen auf Basis indirekter Vergleiche aus RCT im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

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

Das Datum der Recherche soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen. Liegt die Recherche im EU-Dossier mehr als drei

Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt, ist die Recherche zu aktualisieren. Die aktualisierte Recherche ist im Dossier darzustellen.

In den Fällen, in denen keine neuen Quellen identifiziert werden, kann auf das Rechercheergebnis im EU-Dossier verwiesen werden. Sofern neue Quellen identifiziert werden, sind die neuen Erkenntnisse entsprechend der Vorgaben aufzuarbeiten.

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 (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms et cetera.).*** *Benennen Sie*

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Recherche*
- *Studien aus der Suche in Studienregistern/Studienergebnisdatenbanken*
- *Studien aus der Suche auf der Internetseite des Gemeinsamen Bundesausschusses*
- *Resultierender Studienpool aus den einzelnen Suchschritten*

Sofern Angaben zur Informationsbeschaffung zu nicht randomisierten vergleichenden Studien im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

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.***

Sofern Informationen zu den Charakteristika der nicht randomisierten vergleichenden Studien im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Beschreiben Sie die Verzerrungsaspekte der nicht randomisierten vergleichenden Studie auf Studienebene mithilfe des Bewertungsbogens in Anhang 4-F. 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-F.

Tabelle 4-105: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien

Studie	Zeitliche Parallelität der Gruppen	Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren	Verblindung		Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte
			Patient	Behandelnde Personen		
<Studie 1>	<ja/nein/ unklar>	<ja/nein/ unklar>	<ja/nein/ unklar>	<ja/nein/ unklar>	<ja/nein/ unklar>	<ja/nein>

Beschreiben Sie zusammenfassend die Bewertungsergebnisse zu Verzerrungsaspekten auf Studienebene.

Sofern eine Bewertung des Verzerrungspotenzials auf Studienebene im EU-Dossier hinterlegt ist und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen. Die obenstehende Tabelle 4-21 ist im Dossier unabhängig von den Angaben im EU-Dossier auszufüllen.

Nicht zutreffend.

4.3.2.2.3 Ergebnisse aus nicht randomisierten vergleichenden Studien

Geben Sie in der folgenden Tabelle einen Überblick über die patientenrelevanten Endpunkte, auf denen Ihre Bewertung des medizinischen Nutzens und Zusatznutzens aus nicht randomisierten vergleichenden Studien beruht. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Zeile. Fügen Sie für jede Studie eine neue Zeile ein.

Sofern ein Überblick über die patientenrelevanten Endpunkte aus nicht randomisierten vergleichenden Studien im EU-Dossier hinterlegt ist und dieser Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-106: Matrix der Endpunkte in den eingeschlossenen nicht randomisierten vergleichenden Studien

Studie	<Mortalität>	<Gesundheits-bezogene Lebensqualität>	<Endpunkt>	<Endpunkt>	<Endpunkt>
<Studie 1>	nein	ja	ja	ja	nein

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.

Sofern Informationen zur Operationalisierung des jeweiligen Endpunkts im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-107: 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-F. 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-F.

Tabelle 4-108: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien

Studie	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte
<Studie 1>	<ja/nein/unklar>	<ja/nein/unklar>	<ja/nein/unklar>	<ja/nein>

Beschreiben Sie zusammenfassend die Bewertungsergebnisse zu Verzerrungsaspekten auf Endpunktebene.

Sofern eine Bewertung des Verzerrungspotenzials für den jeweiligen Endpunkt im EU-Dossier hinterlegt ist und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen. Die obenstehende Tabelle 4-24 ist im Dossier unabhängig von den Angaben im EU-Dossier auszufüllen.

Nicht zutreffend.

Stellen Sie die Ergebnisse der nicht randomisierten vergleichenden Studien gemäß den Anforderungen des TREND- beziehungsweise des STROBE-Statements dar. Machen Sie dabei auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Sofern Ergebnisse zu nicht randomisierten vergleichenden Studien im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossier zu verweisen.

Nicht zutreffend.

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* 4.3.1.3.2.**

Sofern Informationen beziehungsweise Ergebnisse zu durchgeführten Subgruppenanalysen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

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

Das Datum der Recherche soll nicht mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen. Liegt die Recherche im EU-Dossier mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt, ist die Recherche zu aktualisieren. Die aktualisierte Recherche ist im Dossier darzustellen.

In den Fällen, in denen keine neuen Quellen identifiziert werden, kann auf das Rechercheergebnis im EU-Dossier verwiesen werden. Sofern neue Quellen identifiziert werden, sind die neuen Erkenntnisse entsprechend der Vorgaben aufzuarbeiten.

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 sowohl für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie analog Abschnitt 4.3.1.1 zur Verfügung (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms et cetera).*** *Benennen Sie für das zu bewertende Arzneimittel als auch für die zweckmäßige Vergleichstherapie*

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Recherche*
- *Studien aus der Suche in Studienregistern/Studienergebnisdatenbanken*
- *Studien aus der Suche auf der Internetseite des Gemeinsamen Bundesausschusses*
- *Resultierender Studienpool aus den einzelnen Suchschritten*

Sofern Angaben zur Informationsbeschaffung nach Untersuchungen, die nicht in den Abschnitten 4.3.1, 4.3.2.1 und 4.3.2.2 aufgeführt sind, im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

4.3.2.3.2 Charakteristika der weiteren Untersuchungen

Charakterisieren Sie nachfolgend die weiteren Untersuchungen und bewerten Sie deren Verzerrungsaspekte.

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.

Sofern Angaben zu den Charakteristika der weiteren Untersuchungen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

4.3.2.3.3 Ergebnisse aus weiteren Untersuchungen

Geben Sie in der folgenden Tabelle einen Überblick über die patientenrelevanten Endpunkte, auf denen Ihre Bewertung des medizinischen Nutzens und Zusatznutzens aus weiteren Untersuchungen beruht. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Zeile. Fügen Sie für jede Studie eine neue Zeile ein.

Sofern Angaben zu den Ergebnissen aus weiteren Untersuchungen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-109: Matrix der Endpunkte in den eingeschlossenen weiteren Untersuchungen

Studie	**<Mortalität>**	**<Gesundheits-bezogene Lebensqualität>**	**<Endpunkt>**	**<Endpunkt>**	**<Endpunkt>**
<Studie 1>	nein	ja	ja	ja	nein

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.

Sofern Informationen zur Operationalisierung des jeweiligen Endpunkts im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Tabelle 4-110: Operationalisierung von <Endpunkt xxx> – weitere Untersuchungen

Studie	**Operationalisierung**
<Studie 1>	

Bewerten Sie die Verzerrungsaspekte für den in diesem Abschnitt beschriebenen Endpunkt. 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.

Sofern eine Bewertung der Verzerrungsaspekte für den jeweiligen Endpunkt im EU-Dossier hinterlegt ist und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein soll, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

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. Machen Sie darüber hinaus Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Sofern Ergebnisse der weiteren Untersuchungen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus weiteren Untersuchungen fortlaufend in einem eigenen Abschnitt dar.

4.3.2.3.3.2 Subgruppenanalysen – weitere Untersuchungen

Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen aus weiteren Untersuchungen. ***Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.2.***

Sofern Informationen beziehungsweise Ergebnisse zu durchgeführten Subgruppenanalysen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

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.

Zuranolon wird angewendet zur Behandlung von Erwachsenen mit PPD nach Entbindung. Zur Bewertung des Zusatznutzens von Zuranolon in diesem Anwendungsbereich wurden zwei Teilpopulationen mit unterschiedlichen zVTs definiert.

Für die Teilpopulation a) Patientinnen mit **leichter** depressiver Symptomatik nach der Geburt konnten keine Studien identifiziert und somit keine Daten vorgelegt werden.

Für Teilpopulation b) Patientinnen mit **mittelschwerer bis schwerer** depressiver Symptomatik stehen mit den RCTs ROBIN und SKYLARK zwei relevante Studien zur Verfügung. Beide sind randomisierte, kontrollierte klinische Studien der Phase 3. In der Studie ROBIN sowie auch in SKYLARK wird das Verzerrungspotential aus Studie- und Endpunktebene als niedrig eingestuft. Die zur Bewertung herangezogenen Endpunkte sind patientenrelevant und valide.

Für die vorliegende Nutzenbewertung wurden diese beiden Studien aufgrund ihrer hinreichenden Ähnlichkeit im Studiendesign und in der Studienpopulation in einer Meta-Analyse zusammengefasst. Dadurch liegen Ergebnisse mit höchster Ergebnissicherheit (Evidenzstufe Ia) vor.

Insgesamt kann auf Basis der Meta-Analyse aus den Studien ROBIN und SKYLARK ein signifikanter Zusatznutzen für die Teilpopulation b) mit mittelschwerer bis schwerer PPD nach Entbindung abgeleitet werden.

4.4.2 Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß

Führen Sie die in den Abschnitten 4.3.1 und 4.3.2 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. Berücksichtigen Sie dabei auch die Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext. Liegt ein Zusatznutzen vor, beschreiben Sie worin der Zusatznutzen besteht.

Stellen Sie die Wahrscheinlichkeit des Zusatznutzens dar, das heißt, 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 gegebenenfalls nachgewiesene Unterschiede zwischen verschiedenen Patientengruppen.

Zur Bewertung und Quantifizierung des medizinischen Nutzens und Zusatznutzens werden im vorliegenden Modul 4A Ergebnisse patientenrelevanter Endpunkte der Nutzendimensionen Mortalität, Morbidität, gesundheitsbezogene Lebensqualität und Sicherheit der Zulassungsstudien ROBIN und SKYLARK herangezogen.

Beide Studien untersuchten die Wirksamkeit und Sicherheit von Zuranolon zur Behandlung mittelschwerer bis schwerer PPD als randomisierte, doppelblinde, kontrollierte Parallelgruppenstudien mit einer 14-tägigen Behandlungsdauer und einer Nachbeobachtung bis Tag 45. Es wurden Patientinnen mit einem HAMD-17-Wert von ≥ 26 eingeschlossen, was auf eine schwere Depression hinweist. Zusätzliche Auswertungen nach MADRS beschreiben ein Spektrum von mittelschweren bis schweren Symptomen. Die Studienpopulationen sind demografisch und krankheitsspezifisch vergleichbar.

Obwohl die Datenlage zur PPD in Deutschland begrenzt ist, stimmen das Durchschnittsalter der Patientinnen und die Diagnosemethoden mit deutschen Standards überein. Daher wird von einer Übertragbarkeit auf den deutschen Versorgungskontext ausgegangen.

Die Ableitung des Zusatznutzens von Zuranolon für die Teilpopulation b) (mittelschwer – schwer) beruht maßgeblich auf den Ergebnissen zur Mortalität, Morbidität und Sicherheit aus einer Meta-Analyse der beiden Studien ROBIN und SKYLARK. Zusätzlich zu Wirksamkeit und Sicherheit liegen aus der Studie ROBIN Daten für die Lebensqualität vor. Die Patientinnen der Teilpopulation a) (mild) waren in den Studien nicht eingeschlossen, da diese oft durch therapeutische Interventionen oder unterstützende Maßnahmen wie körperliche Aktivität und nicht medikamentös behandelt werden.

Die vorliegenden Daten, bestehend aus einer Meta-Analyse der höchsten Evidenzstufe Ia sowie ergänzend statistisch signifikante Ergebnisse auch auf Ebene der Einzelstudien (siehe Anhang 4-H und Anhang 4-I), verdeutlichen eindrucksvoll den Zusatznutzen der Zuranolon+PiT-Therapie bei der mittelschweren – schweren PPD.

Insbesondere der schnelle Wirkungseintritt bereits zu Beginn der Behandlung ermöglicht eine rasche und für die Patientinnen deutlich wahrnehmbare Reduktion der depressiven Symptomatik und trägt maßgeblich zur Verbesserung der Lebensqualität betroffener Mütter bei. Damit trägt Zuranolon der besonderen Akuität der PPD Rechnung, die in einer Phase hoher psychischer und funktioneller Vulnerabilität der Mutter auftritt und eine zeitnahe therapeutische Intervention erfordert.

Im Folgenden fassen Tabelle 4-111 und Tabelle 4-112 die Ergebnisse zusammen, die anschließend bezüglich der medizinischen Bedeutung und Ausmaß des Zusatznutzens in Kontext gesetzt werden.

Schnell einsetzende Wirkung stellt einen Zusatznutzen dar

Die PPD stellt eine Erkrankung dar, die in eine Phase maximaler psychischer, funktioneller und biologischer Vulnerabilität der Mutter fällt. Vor diesem Hintergrund ist ein rascher Wirkungseintritt ein zentrales Kriterium für die Wahl einer geeigneten Therapie [2; 3]. Studien zeigen, dass die Symptome einer PPD bei einem Großteil der Patientinnen innerhalb von 3 bis 6 Monaten abklingen [4; 5]. Vor diesem Hintergrund fügt sich eine 14-tägige Therapie mit akutem, klar zeitlich begrenztem Behandlungsverlauf wie Zuranolon gut in das klinische Bild der Erkrankung ein. PPD ist einerseits eine akut auftretende und zeitkritische Erkrankung, erfordert andererseits nicht in allen Fällen eine langfristige Pharmakotherapie [4; 5]. Zuranolon

ist aufgrund seiner Eigenschaft als Rapid-Acting Antidepressant (RAAD) und seiner spezifischen Entwicklung für die Pathophysiologie der PPD als einmalige 14-tägige orale Kurzzeittherapie konzipiert.

Die Ergebnisse in Tabelle 4-111 zeigen, dass die Kombination von Zuranolon mit PiT im Vergleich zu PiT allein eine schnelle und statistisch signifikante Verbesserung der depressiven Symptomatik bewirkt, und somit die Akutphase drastisch verkürzt. Schon an Tag 3 war der Unterschied zwischen den Behandlungsarmen für mehrere Endpunkte statistisch signifikant.

Tabelle 4-111: Signifikante Endpunkte zum ersten Messzeitpunkt

Endpunkt	**Art der Analyse**	**Meta-Analyse Zuranolon+PiT (N = 176) vs. PiT (N = 177) Effektmaß [95 %-KI], p-Wert**	**Ableitung des Ausmaßes des Zusatznutzens**
Morbidität			
Depressive Symptomatik (HAMD-17, MADRS) - Symptomverbesserung und Remission unter Zuranolon			
HAMD-17-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen Tag 3	RR: 1,723 [1,186; 2,503], 0,0043	Beträchtlich
HAMD-17-Remission	Anteil der Studienteilnehmerinnen mit Ansprechen Tag 3	RR: 2,414 [1,515; 5,064], 0,0197	Beträchtlich
MADRS-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen Tag 8	RR: 1,423 [1,125; 1,799], 0,0032	Beträchtlich
Patientenberichtete Endpunkte (EPDS) - Verbesserung der selbstberichteten PPD-Symptome unter Zuranolon			
EPDS-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen Tag 3	RR: 1,454 [1,031; 2,049], 0,0326	Gering
Angstsymptomatik (HAM-A) - Reduktion begleitender Angstsymptome unter Zuranolon			
HAM-A-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen Tag 3	RR: 1,304 [1,009; 1,686], 0,0423	Gering
Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.			

Insbesondere die schnelle Ansprechrate bei der HAMD-17 und MADRS sowie das frühzeitige Erreichen einer Remission bei der HAMD-17 in **beträchtlichem Ausmaß** unterstreichen die Bedeutung der Behandlung für Patientinnen mit mittelschwerer bis schwerer PPD. Die schnelle

Wirkung ist entscheidend, um die zeitkritischen Auswirkungen der Erkrankung auf die Mutter-Kind-Beziehung und die familiäre Dynamik zu minimieren. Neben den eindrücklichen Ergebnissen bereits am ersten Messzeitpunkt, verdeutlichen auch die Ergebnisse zum Ende der Kurzzeitbehandlung, am Tag 15, und zum Ende der Beobachtung an Tag 45 die schnelle Verbesserung der Symptome und den deutlichen Vorteil der Zuranolon+PiT-Behandlung.

Tabelle 4-112: Zusammenfassung der Ergebnisse

Endpunkt	**Art der Analyse**	**Meta-Analyse** **Zuranolon+PiT (N = 176) vs. PiT (N = 177)** **Effektmaß [95 %-KI], p-Wert**		**Ableitung des Ausmaßes des Zusatznutzens**	
		Tag 15 Ende der Kurzzeitbehandlung	Tag 45 Ende der Beobachtung	Tag 15	Tag 45
Mortalität					
Anzahl der Todesfälle		In beiden Studien wurden keine Todesfälle beobachtet.			
Morbidität					
Depressive Symptomatik (HAMD-17, MADRS) - Schnellere und häufigere Symptomverbesserung und Remission unter Zuranolon					
HAMD-17-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,507 [1,219; 1,863], 0,0002	RR: 1,269 [1,051; 1,532], 0,0130	Beträchtlich	Gering
	Zeit bis zum Ansprechen	HR: 1,854 [1,4372; 2,3920], < 0,0001		Erheblich	
HAMD-17-Remission	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,773 [1,222; 2,574], 0,0026	RR: 1,634 [1,213; 2,200], 0,0012	Beträchtlich	Beträchtlich
	Zeit bis zur Remission	HR: 1,753 [1,2818; 2,3962], 0,0004		Beträchtlich	
MADRS-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,528 [1,233; 1,894], 0,0001	RR: 1,315 [1,086; 1,592], 0,0050	Beträchtlich	Gering
	Zeit bis zum Ansprechen	HR: 1,497 [1,1630; 1,9258], 0,0017		Beträchtlich	
MADRS-Remission	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,690 [1,258; 2,271], 0,0005	RR: 1,409 [1,089; 1,823], 0,0091	Beträchtlich	Gering
	Zeit bis zur Remission	HR: 1,683 [1,2676; 2,2350], 0,0003		Beträchtlich	
Patientenberichtete Endpunkte (EPDS, PHQ-9) - Schnellere und häufigere Verbesserung der selbstberichteten PPD-Symptome unter Zuranolon					
EPDS-Ansprechen	Anteil der	RR: 1,169 [1,018; 1,342],	RR: 1,162 [1,017; 1,328],	Gering	Gering

Endpunkt	Art der Analyse	Meta-Analyse Zuranolon+PiT (N = 176) vs. PiT (N = 177) Effektmaß [95 %-KI], p-Wert		Ableitung des Ausmaßes des Zusatznutzens	
	Studienteilnehmerinnen mit Ansprechen	0,0272	0,0269		
	Zeit bis zum Ansprechen	HR: 1,244 [0,9933; 1,5575], 0,0572		Nicht belegt	
PHQ-9-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,177 [1,025; 1,352], 0,0209	RR: 1,148 [1,001; 1,315], 0,0480	Gering	Gering
	Zeit bis zum Ansprechen	HR: 1,306 [1,0426; 1,6348], 0,0201		Gering	
Angstsymptomatik (HAM-A) - Reduktion begleitender Angstsymptome unter Zuranolon					
HAM-A-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,323 [1,129; 1,552], 0,0006	RR: 1,178 [1,014; 1,369], 0,0333	Beträchtlich	Gering
	Zeit bis zum Ansprechen	HR: 1,356 [1,0738; 1,7128], 0,0105		Gering	
Lebensqualität					
SF-36 und BIMF (nur in Studie ROBIN erhoben) - Häufigere Verbesserung körperlicher Schmerzen und sozialer Funktion unter Zuranolon					
SF-36 PCS-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,500 [0,718; 3,133], 0,2802	RR: 1,543 [0,659; 3,615], 0,0322	Nicht belegt	Nicht belegt
	Zeit bis zum Ansprechen	HR: 1,619 [0,7953; 3,2953], 0,1841		Nicht belegt	
SF-36 MCS-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,058 [0,835; 1,341], 0,6393	RR: 1,223 [0,995; 1,502], 0,0555	Nicht belegt	Nicht belegt
	Zeit bis zum Ansprechen	HR: 1,051 [0,7375; 1,4967], 0,7845		Nicht belegt	
SF-36 „Körperliche Schmerzen”-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,204 [0,877; 1,652], 0,2513	RR: 1,461 [1,051; 2,032], 0,0243	Nicht belegt	Gering
	Zeit bis zum Ansprechen	HR: 1,229 [0,8214; 1,8390], 0,3159		Nicht belegt	

Endpunkt	**Art der Analyse**	**Meta-Analyse** **Zuranolon+PiT (N = 176) vs. PiT (N = 177)** **Effektmaß [95 %-KI], p-Wert**		**Ableitung des Ausmaßes des Zusatznutzens**	
SF-36 “Soziale Funktionsfähigkeit”-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,149 [0,899; 1,470], 0,2671	RR: 1,218 [1,011; 1,467], 0,0378	Nicht belegt	Gering
	Zeit bis zum Ansprechen	HR: 1,102 [0,7751; 1,5675], 0,5879		Nicht belegt	
SF-36 “Psychisches Wohlbefinden”-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,149 [0,918; 1,439], 0,2253	RR: 1,240 [1,028; 1,497], 0,0249	Nicht belegt	Gering
	Zeit bis zum Ansprechen	HR: 1,022 [0,7245; 1,4427], 0,8999		Nicht belegt	
BIMF-Gesamtwert-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,616 [1,040; 2,513], 0,0330	RR: 1,223 [0,914; 1,636], 0,1753	Gering	Nicht belegt
	Zeit bis zum Ansprechen	HR: 1,040 [0,7028; 1,5387], 0,8449		Nicht belegt	
BIMF-“Kompetenz der Mutter”-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,142 [0,725; 1,800], 0,5668	RR: 1,373 [0,921; 2,046], 0,1194	Nicht belegt	Nicht belegt
	Zeit bis zum Ansprechen	HR: 0,986 [0,6289; 1,5444], 0,9492		Nicht belegt	
BIMF-“Bedürfnisse der Mutter”-Ansprechen	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,054 [0,747; 1,487], 0,7654	RR: 1,165 [0,905; 1,501], 0,2364	Nicht belegt	Nicht belegt
	Zeit bis zum Ansprechen	HR: 0,912 [0,6241; 1,3316], 0,6322		Nicht belegt	
Relevante Ergebnisse von Kliniker*innen bewertet bezüglich Schweregrad (CGI-S) der PPD und Verbesserung des Zustands (CGI-I)					
CGI-S-Ansprechen [1]	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 1,348 [1,113; 1,633], 0,0023	RR: 1,271 [1,073; 1,505], 0,0054	Beträchtlich	Gering
	Zeit bis zum Ansprechen	HR: 1,688 [1,3209; 2,1562], < 0,0001		Beträchtlich	
CGI-I-Ansprechen [1]	Anteil der Studienteilnehmerinnen mit	RR: 1,393 [1,151; 1,686], 0,0007	RR: 1,181 [0,992; 1,406], 0,0615	Beträchtlich	Nicht belegt

Endpunkt	**Art der Analyse**	**Meta-Analyse** **Zuranolon+PiT (N = 176) vs. PiT (N = 177)** **Effektmaß [95 %-KI], p-Wert**	**Ableitung des Ausmaßes des Zusatznutzens**
	Ansprechen		
	Zeit bis zum Ansprechen	HR: 1,660 [1,3042; 2,1135], < 0,0001	Beträchtlich
Sicherheit [2]			
C-SSRS (Frage 1 – 10)	Anteil der Studienteilnehmerinnen mit Ansprechen	RR: 0,95 [0,388; 2,339], 0,9157 / RR: 0,43 [0,135; 1,367], 0,1525	Nicht belegt
UE		RR: 1,21 [1,009; 1,449], 0,0396	Ergänzend dargestellt
SOC Erkrankungen des Nervensystems		RR: 1,66 [1,230; 2,248], 0,0010	Geringerer
PT Somnolenz		RR: 2,84 [1,569; 5,142], 0,0006	Geringerer Transient zu Beginn der Behandlung (vorwiegend Tag 1-2)
PT Sedierung		RR: 14,57 [1,946; 109,126], 0,0091	Geringerer Transient zu Beginn der Behandlung (vorwiegend Tag 1-2)
Schwere UE		RR: 1,46 [0,419; 5,074], 0,5540	Nicht belegt
SUE		RR: 2,91 [0,306; 27,748], 0,3521	Nicht belegt
Therapieabbruch aufgrund von UE		RR: 2,43 [0,478; 12,351], 0,2848	Nicht belegt

[1] Das CGI-S-Ansprechen war definiert als Reduktion von ≥ 1,05 Punkte (15 % der Skalenspannweite 0 – 7) von Baseline. Das CGI-I-Ansprechen war definiert als Antwort „stark verbessert“ und „sehr stark verbessert“ der 7 möglichen Antworten. Alle Analysen zu CGI-Endpunkten befinden sich in Anhang 4-G.

[2] Die Sicherheitsanalysen wurden für die Sicherheits-Population durchgeführt: Zuranolon+PiT (N = 176) und PiT (N = 171).

Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.

Mortalität

Die Mortalität wird in den Studien durch die Anzahl der während des Studienzeitraums aufgetretenen Todesfälle erfasst. Eine metaanalytische Zusammenfassung ergab sich nicht, da in beiden Studien keine Todesfälle auftraten.

Morbidität

Die Morbidität wird durch verschiedene Instrumente gemessen. Der primäre Endpunkt beider Studien war die HAMD-17, welche die Schwere depressiver Symptome anhand von 17 Items bewertet, die sowohl psychopathologische als auch somatische Aspekte wie Schlafstörungen oder Suizidgedanken umfassen. Höhere Werte auf der Skala (0 – 52) deuten auf schwerere Symptome hin. Die Ergebnisse für die HAMD-17 zeigen, dass die Kombination von Zuranolon mit PiT im Vergleich zu PiT allein statistisch signifikante Verbesserungen bei der Ansprechrate erzielte. Das Ansprechen war *a priori* definiert als Reduktion von ≥ 50 % gegenüber dem Gesamtwert zu Baseline. Zum Ende der 14-tägigen Kurzzeitbehandlung mit Zuranolon betrug das Relative Risiko (RR) 1,507 mit einem p-Wert von 0,0002 und einem 95 %-Konfidenzintervall (KI) von [1,219; 1,863], was auf einen **beträchtlichen** Zusatznutzen hinweist. Zusätzlich zeigen die Analysen zur Zeit bis zum HAMD-17-Ansprechen einen statistisch signifikanten Vorteil zugunsten von Zuranolon und PiT. Patientinnen unter Zuranolon sprachen deutlich schneller an (Median 8 vs. 20 Tage), was einer signifikant erhöhten Rate des Ansprechens über die Zeit um 85,4 % entspricht (HR = 1,854 [1,4372; 2,3920], p < 0,0001). Dieser Effekt entspricht einem schnellen Ansprechen in **erheblichem** Ausmaß und blieb bis zum Ende der Beobachtung an Tag 45 bestehen.

Die MADRS findet häufig Anwendung in der klinischen Praxis und konzentriert sich auf zentrale depressive Merkmale wie Traurigkeit und Suizidgedanken. Sie misst Veränderungen der Symptomatik mit 10 Items auf einer Skala von 0 – 60. Das MADRS-Ansprechen war auch *a priori* definiert als Reduktion von ≥ 50 % gegenüber dem Gesamtwert zu Baseline. Im Zuranolon+PiT-Arm sprachen statistisch signifikant mehr Patientinnen an als im Vergleichsarm. Zum Ende der Behandlung lag die Wahrscheinlichkeit, eine relevante Verbesserung, also ein Ansprechen auf der MADRS zu erreichen bei der Behandlung mit Zuranolon in Kombination mit PiT um 52,8 % höher als im Vergleichsarm (RR: 1,528 [1,233; 1,894], p = 0,0001). Dies gilt als **beträchtlicher** Zusatznutzen. Auch hier wurde ein Ansprechen statistisch signifikant – um 49,7 % (8,5 vs. 17 Tage im Median) – schneller im Zuranolon+PiT-Arm als im Vergleichsarm erreicht (HR: 1,497 [1,1630; 1,9258], p = 0,0017), was einem ebenfalls **beträchtlichen** Zusatznutzen entspricht. Diese schnell eintretende, deutlich überlegene Wirksamkeit von Zuranolon+PiT gegenüber der Vergleichstherapie wird durch den anhaltenden, statistisch signifikanten Effekt untermauert, der bis zum Abschluss der Beobachtungsphase an Tag 45 nachweisbar ist.

Auf den Instrumenten HAMD-17 und MADRS wurde nicht nur das Ansprechen, sondern auch die Remission analysiert. Eine Remission war auf der HAMD-17 definiert als Wert von ≤ 7 Punkten und bei der MADRS ≤ 10 Punkten. Übereinstimmend zeigten beide Instrumente zum Ende der Behandlung einen statistisch signifikanten Vorteil zugunsten der Zuranolon+PiT-Behandlung. Zum Ende der Behandlung lag die Wahrscheinlichkeit, eine Remission nach

HAMD-17 bzw. MADRS zu erreichen bei der Behandlung mit Zuranolon in Kombination mit PiT um 77,3 % bzw. 69 % höher als im Vergleichsarm (RR: 1,773 [1,222; 2,574], p = 0,0026 bzw. RR: 1,690 [1,258; 2,271], p = 0,0005). Die Dauer bis zur Remission unterschied sich dabei auch statistisch signifikant und wurde im Zuranolon+PiT-Arm schneller erreicht (HAMD-17 um 75,3 %, 29 vs. 50 Tage; MADRS um 68,3 %, 16 vs. 56 Tage), als im Vergleichsarm (HR: 1,753 [1,2818; 2,3962], p = 0,0004 bzw. HR: 1,683 [1,2676; 2,2350], p = 0,0003). Mit einer statistisch signifikanten Überlegenheit gegenüber dem Vergleichsarm bezüglich der Remission und dem schnellen Einsetzen dieser – beides mit dem Ausmaß **beträchtlich** – lindert Zuranolon die Symptome der PPD gemessen mittels HAMD-17 und MADRS und ermöglicht Patientinnen schnell eine Remission zu erreichen. Die statistisch signifikante Überlegenheit der Zuranolon+PiT-Behandlung blieb bis zum Ende der Beobachtung mit einem **beträchtlichen** bzw. **geringen** Ausmaß für die Remission im HAMD-17 bzw. MADRS erhalten.

Die Edinburgh Postnatal Depression Scale (EPDS) ist ein selbstberichteter Fragebogen, der speziell für die Erkennung und Bewertung der PPD entwickelt wurde. Mit 10 Items und einer Skala von 0 – 30 bezieht sie sich auf emotionale Symptome. Das Ansprechen, heißt eine klinisch relevante Verbesserung, war für die EPDS – in Ermangelung prädefinierter Grenzen – definiert als Zunahme um ≥ 15 % der Skalenspannweite. Für das EPDS-Ansprechen zeigte sich zu Behandlungsende (RR: 1,169 [1,018; 1,342], p = 0,0272) sowie zum Ende der Beobachtung (RR: 1,162 [1,017; 1,328], p = 0,0269) ein statistisch signifikanter Vorteil der Zuranolon+PiT-Behandlung gegenüber dem Vergleichsarm. Dies entspricht einem Zusatznutzen in **geringem** Ausmaß. Die Zeit bis zum EPDS-Ansprechen zeigte für beide Behandlungsgruppen keinen statistisch signifikanten Unterschied.

Der Patient Health Questionnaire (PHQ)-9, ein weiteres Selbstbeurteilungsinstrument, erfasst die Schwere depressiver Symptome anhand von 9 Items (Skala 0 – 27) und legt besonderen Wert auf Symptome wie Anhedonie und Suizidgedanken. Eine relevante Verbesserung der Symptome war definiert als Zunahme um ≥ 15 % der Skalenspannweite und wurde als Ansprechen ausgewertet. Für das PHQ-9-Ansprechen zeigte sich zu Behandlungsende ein statistisch signifikanter Vorteil der Zuranolon+PiT-Behandlung gegenüber dem Vergleichsarm (RR: 1,177 [1,025; 1,352], p = 0,0209). Die Nachhaltigkeit der Verbesserung und Überlegenheit gegenüber dem Vergleichsarm zeigte sich auch zum Ende der Beobachtung (RR: 1,148 [1,001; 1,315], p = 0,0480). Das Ansprechen wurde statistisch signifikant – um 30,6 % (5 vs. 7 Tage zum 25. Perzentil) – schneller im Zuranolon+PiT-Arm als im Vergleichsarm erreicht (HR: 1,306 [1,0426; 1,6348], 0,0201). Dies entspricht einem Zusatznutzen in **geringem** Ausmaß.

Zu den typischen Symptomen einer PPD gehören Angststörungen, insbesondere Panikattacken und übersteigerte Sorgen um das Wohl des Kindes bzw. für das Wohle des Kindes nicht ausreichend sorgen zu können. Die Hamilton Rating Scale for Anxiety (HAM-A) als weiterer Morbiditätsendpunkt misst die Schwere von Angstzuständen mit 14 Items (Skala von 0 – 56 Punkten), die sowohl psychische als auch somatische Angst bewerten. Das klinisch relevante Ansprechen wird über eine 15 %-ige Verbesserung auf der Skalenspannweite der HAM-A angezeigt. Im Zuranolon+PiT-Arm sprachen statistisch signifikant mehr Patientinnen

an als im Vergleichsarm. Zum Ende der Behandlung lag die Wahrscheinlichkeit, eine relevante Verbesserung bezüglich der Angstsymptomatik zu erreichen bei der Behandlung mit Zuranolon in Kombination mit PiT um 32,3 % höher als im Vergleichsarm (RR: 1,323 [1,129; 1,552], p = 0,0006). Dies entspricht einem **beträchtlichen** Zusatznutzen. Auch hier wurde ein Ansprechen statistisch signifikant – um 35,6 % (3 vs. 4 Tage zum 25. Perzentil) – schneller im Zuranolon+PiT-Arm als im Vergleichsarm erreicht (HR: 1,356 [1,0738; 1,7128], p = 0,0105). Diese schnell eintretende, deutlich überlegene Wirksamkeit von Zuranolon gegenüber der Vergleichstherapie wird bis zum Ende der Beobachtungsphase beibehalten (RR: 1,178 [1,014; 1,369], p = 0,0333) und entspricht einem **geringen** Zusatznutzen.

Relevante Ergebnisse von Kliniker*innen bewertet

Die Clinical Global Impression-Skalen (CGI) sind standardisierte Instrumente, die häufig in klinischen Studien eingesetzt werden, um den Schweregrad einer Erkrankung sowie die Verbesserung des Zustands von Patient*innen zu bewerten. Zum Gesamtbild in Bezug auf Schweregrad der PPD bzw. Veränderung des Zustands der Patientin kann die Einschätzung von Kliniker*innen einen wertvollen Beitrag leisten. Die zuvor dargestellten Endpunkte werden hierdurch nicht nur ergänzt, sondern in ihrer Robustheit unterstützt.

Die CGI-S-Skala (CGI-Severity) dient in diesem Studiensetting dazu, den Gesundheitszustand der PPD-Patientinnen zum Zeitpunkt der Bewertung einzuschätzen. Dabei wird eine 7-Punkte-Likert-Skala verwendet, die von „normal, nicht krank“ (1) bis „extrem krank“ (7) reicht. Die Bewertung erfolgt relativ zur bisherigen klinischen Erfahrung der Bewertungsperson mit Patientinnen, die dieselbe Diagnose haben, und bildet die Grundlage für spätere Beurteilungen der Verbesserung. Der Gesundheitszustand gemessen mittels CGI-S verbessert sich im Zuranolon+PiT-Arm statistisch signifikant bei mehr Patientinnen als im Vergleichsarm. Zum Ende der Behandlung haben Patientinnen unter Zuranolon eine 34,8 % höhere Wahrscheinlichkeit eine klinisch relevante Verbesserung zu erreichen als unter der Vergleichstherapie (RR: 1,348 [1,113; 1,633], p = 0,0023). Diese Verbesserung wird auch statistisch signifikant schneller erreicht (HR: 1,688 [1,3209; 2,1562], p < 0,0001) Dies entspricht einem **beträchtlichen** Effekt.

Die CGI-I-Skala (CGI-Improvement) misst hingegen die Verbesserung des Zustands eines Patienten nach der Behandlung im Vergleich zu Baseline. Auch hier wird eine 7-Punkte-Likert-Skala verwendet, die von „sehr stark verbessert“ (1) bis „sehr stark verschlechtert“ (7) reicht. Im CGI-I zeigte sich ebenfalls für statistisch signifikant mehr Patientinnen eine Verbesserung im Zuranolon+PiT-Arm als im Vergleichsarm. Zum Ende der Behandlung haben Patientinnen unter Zuranolon eine ca. 40 % höhere Wahrscheinlichkeit die Verbesserung zu erreichen als unter der Vergleichstherapie (RR: 1,393 [1,151; 1,686], p = 0,0007). Diese Verbesserung wird auch statistisch signifikant schneller erreicht (HR: 1,660 [1,3042; 2,1135], < 0,0001). Hieraus kann in Übereinstimmung mit den Ergebnissen des CGI-S ein **beträchtlicher** Effekt abgeleitet werden.

Der statistisch signifikante Unterschied zwischen den Behandlungsgruppen ist auch nachhaltig und wurde zum Ende der Beobachtung in beiden CGI-Endpunkten bestätigt.

Lebensqualität

Die Lebensqualität wird durch den Short Form (SF)-36 und den Barkin Index of Maternal Functioning (BIMF) erfasst. Beide Instrumente wurden nur in der Studie ROBIN erhoben, daher können hier nur die Ergebnisse auf Ebene einer RCT dargestellt werden.

Der SF-36 ist ein standardisierter Fragebogen mit 36 Items, der die allgemeine gesundheitsbezogene Lebensqualität mit 8 Subskalen misst, die in körperliche und psychische Summenscores unterteilt werden können. Der BIMF hingegen bewertet die mütterliche Funktionsfähigkeit in den ersten 12 Monaten nach der Geburt. Mit 20 Items und einer Skala von 0 – 120 deckt er Bereiche wie Selbstfürsorge, Mutter-Kind-Interaktion und soziale Unterstützung ab. Beide Instrumente sind validiert und gelten als patientenrelevant, da sie zentrale Aspekte der Lebensqualität und der Bewältigung des Alltags erfassen. Es wurden Responderanalysen bezüglich des Ansprechens durchgeführt, wobei für alle Skalen eine relevante Verbesserung definiert war als Zunahme um ≥ 15 % des gesamten Skalenbereich.

Statistisch signifikant positive Effekte auf die Lebensqualität zeigten sich bei drei der acht Subskalen des SF-36. Es konnte am Ende der Beobachtung für „Körperliche Schmerzen“ ein statistisch signifikanter Vorteil zugunsten von Zuranolon+PiT gegenüber der Vergleichstherapie beobachtet werden (RR: 1,461 [1,051; 2,032], p = 0,0243). Ähnlich positive Effekte waren auch auf den Subskalen „Soziale Funktionsfähigkeit“ (RR: 1,218 [1,011; 1,467], p = 0,0378) und „Psychisches Wohlbefinden“ (RR: 1,240 [1,028; 1,497], p = 0,0249) zu sehen. Für den Gesamtwert des BIMF wurden bereits zum Ende der Zuranolon-Behandlung statistisch signifikant mehr Patientinnen mit Ansprechen im Zuranolon+Pit-Arm beobachtet als im Vergleichsarm (RR: 1,616 [1,040; 2,513], p = 0,0330). Den statistisch signifikanten Ergebnissen zugunsten von Zuranolon stehen keine gegenteiligen Effekte auf irgendeiner Skala des SF-36 oder BIMF entgegen. Zusammenfassend kann für die Lebensqualität ein Zusatznutzen in **geringem** Ausmaß entnommen werden.

Sicherheit

Erfassung von Suizidalität mittels Columbia Suicide Severity Rating Scale (C-SSRS)

In den Studien ROBIN und SKYLARK wurden suizidale Gedanken bei wenigen Patientinnen beobachtet, Aufgrund der insgesamt sehr geringen Ereigniszahlen konnte kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen festgestellt werden.

Unerwünschte Ereignisse

Im Zuranolon+PiT-Arm trat bei 63,6 % der Patientinnen mindestens ein UE auf, während im Vergleichsarm bei 52,6 % der Patientinnen mindestens ein UE beobachtet wurde. Innerhalb der SOC „Erkrankungen des Nervensystems“ wurden statistisch signifikant häufiger UE im Zuranolon+PiT-Arm festgestellt als im Vergleichsarm (RR: 1,66 [1,230; 2,248], p = 0,0010). Nebenwirkungen aus dieser SOC sind für Zuranolon bekannt und werden auch in der Fachinformation genannt [1]. Die PTs „Somnolenz“ und „Sedierung“ traten im Zuranolon+PiT-Arm signifikant häufiger auf (RR: 2,84 [1,569; 5,142], p = 0,0006 bzw. RR: 14,57 [1,946; 109,126], p = 0,0091). Die Ereignisse waren mild bis moderat und traten, wie in Tabelle 4-113 dargestellt, vorwiegend in den ersten 1-2 Tagen auf [6; 7].

Tabelle 4-113: Inzidenz von UE nach Zeit des Auftretens

UE **Zeitintervall**	**Zuranolon+PiT** **(N = 176)**	**PiT** **(N = 171)**
Somnolenz [1], n (%)		
1 – 2 Tage	32 (18,2)	12 (7,0)
3 – 7 Tage	10 (5,7)	1 (0,6)
8 – 15 Tage	4 (2,3)	1 (0,6)
> 15 Tage	0	0
Sedierung [2]		
1 – 2 Tage	10 (5,7)	0
3 – 7 Tage	5 (2,8)	1 (0,6)
8 – 15 Tage	2 (1,1)	0
> 15 Tage	0	0

[1]: Somnolenz umfasst die PTs Somnolenz und Hypersomnie.

[2]: Sedierung umfasst den PT Sedierung

UE werden basierend auf den Zeitintervallen ihres Auftretens klassifiziert, entsprechend der individuellen Expositionsdauer jeder Teilnehmerin zum Zeitpunkt des Auftretens des UE. Die Zeit bis zum Auftreten wird berechnet als Datum des UE-Beginns – Datum der ersten Dosis + 1. Für Teilnehmerinnen, die denselben Ereignistyp mehrfach und in verschiedenen Zeitintervallen erlebten, wird das Ereignis in jedem Intervall, in dem es auftrat, einmal gezählt. UE mit unvollständigen Beginn-Daten werden von diesen Analysen ausgeschlossen.

Quelle: [7]

Alle Abkürzungen befinden sich im Abkürzungsverzeichnis.

Sowohl Somnolenz als auch Sedierung gehören zu den bekannten, in der Fachinformation genannten, sehr häufigen Nebenwirkungen, die jedoch zumeist transient zu Beginn der Behandlung auftreten und im Zusammenhang mit dem Wirkmechanismus von Zuranolon stehen [1].

Hinsichtlich des Auftretens schwerer UE zeigte sich kein statistisch signifikanter Unterschied zwischen den Behandlungsarmen. Ebenso gab es keinen signifikanten Unterschied im Auftreten von SUE zwischen den Gruppen. Diese Ereignisse waren insgesamt selten.

Die Therapieabbrüche aufgrund von UE waren nicht signifikant unterschiedlich und insgesamt selten mit 2,8 % im Zuranolon+PiT-Arm und 1,2 % im Vergleichsarm.

Insgesamt zeigte Zuranolon in der Indikation PPD ein gut verträgliches Sicherheitsprofil. Die Mehrheit der UE war von leichter bis moderater Schwere. Häufig berichtete UE betrafen das zentrale Nervensystem, darunter Somnolenz und Sedierung, die transient zum Beginn der Behandlung beobachtet wurden.

Schlussfolgerungen zum Zusatznutzen und zum therapeutisch bedeutsamen Zusatznutzen

Die PPD ist eine akute, ernsthafte Erkrankung, deren depressive Symptomatik nicht nur die betroffene Mutter und die Entwicklung des Säuglings erheblich beeinträchtigt, sondern auch die gesamte familiäre Dynamik, inklusive Partner*innen und Geschwisterkinder, mit kurz- und langfristigen Konsequenzen belastet.

Bei mittelschweren bis schweren Formen der PPD sind die Symptome wie anhaltende Traurigkeit, häufiges Weinen, deutliche Ängstlichkeit, Schlafstörungen, Konzentrationsschwierigkeiten und möglicherweise Appetitveränderungen oft so ausgeprägt, dass sie die alltäglichen Funktionen der Mutter erheblich einschränken. Das Leben der Betroffenen mit mittelschwerer bis schwerer PPD kann akut durch Selbstmordgedanken oder gar Selbstmordversuche bedroht sein. Symptome der Mutter schränken die Fähigkeit, sich um sich selbst und den Säugling zu kümmern deutlich ein und erfordern eine engmaschige Beobachtung und eine angemessene Behandlung. Die Auswirkungen auf die Mutter-Kind-Interaktion können zudem langfristige Folgen für die Entwicklung des Säuglings haben, einschließlich emotionaler und kognitiver Beeinträchtigungen. Endpunkte, die die Symptome der Mutter bewerten, sind somit nicht nur für die Mutter selbst von Bedeutung, sondern auch für die langfristige Gesundheit des Säuglings. Darüber hinaus können die depressiven Symptome der Mutter negative Konsequenzen im gesamten familiären Umfeld nach sich ziehen, einschließlich eines erhöhten Depressionsrisikos für den Partner oder die Partnerin sowie emotionaler und sozialer Entwicklungsstörungen bei Geschwisterkindern [8-10]. Daher ist es unerlässlich, alle Endpunkte in diesem Kontext als schwerwiegend zu betrachten, um die umfassenden Auswirkungen der Erkrankung auf die gesamte Familie zu erkennen und zu adressieren.

Alle Ergebnisse bestätigen: Zuranolon zeigt eine schnelle und anhaltende Wirksamkeit bei der Behandlung von mittelschwerer bis schwerer PPD. Diese schnelle Wirkung ist besonders wertvoll, da eine PPD nicht nur die Lebensqualität der Betroffenen erheblich beeinträchtigt, sondern auch schwerwiegende Folgen für die Familie haben kann, einschließlich eines erhöhten Risikos für Suizid [11; 12], einer gestörten Mutter-Kind-Bindung und kindlicher Mangelernährung im ersten Lebensjahr [3]. Im Vergleich zu herkömmlichen Antidepressiva, die oft Wochen bis Monate benötigen, um eine Wirkung zu zeigen, bietet Zuranolon durch seinen raschen Wirkungseintritt eine dringend benötigte Entlastung in einer kritischen Phase [13-15]. Dies beschleunigt nicht nur die Genesung, sondern mindert auch die langfristigen sozialen und gesundheitlichen Folgen einer PPD. Bereits nach der zweiten täglichen Dosis geht es den Patientinnen deutlich besser, was die HAMD-17-, EPDS- und HAM-A-Ergebnisse an Tag 3 mit einem statistisch signifikanten Vorteil gegenüber dem Vergleichsarm zeigen (siehe Tabelle 4-111).

Die Ergebnisse zum Ende der Behandlung sowie auch zum Ende der Beobachtung aus der Meta-Analyse belegen signifikante Verbesserungen.

Schnelle Wirkung mit beträchtlichem Zusatznutzen für Patientinnen und Familien

Unter Zuranolon+PiT zeigen die Patientinnen eine besonders schnelle und statistisch signifikante Verbesserung der depressiven Symptomatik, insbesondere durch hohe Ansprech- und Remissionsraten bei HAMD-17 und MADRS. Diese schnelle Wirkung ist entscheidend für die Patientin selbst sowie auch für die Mutter-Kind-Beziehung, die unter den Folgen der Erkrankung massiv beeinträchtigt sein können. Insgesamt stellt die frühzeitige Verbesserung der Symptome einen zentralen Vorteil in **beträchtlichem** Ausmaß dar.

Morbidität: beträchtlicher Zusatznutzen

Signifikant mehr Patientinnen unter Zuranolon gegenüber der Vergleichstherapie erreichten am Ende der Behandlung eine 50%-ige Verbesserung der Depressionssymptome gemessen mittels HAMD-17 und MADRS. Dies wird als „beträchtlicher Zusatznutzen“ bewertet. Die TTE-Analysen zeigten einen schnellen Wirkeintritt von beträchtlichem Ausmaß für MADRS sowie erheblichem Ausmaß für HAMD-17, was besonders in kritischen Phasen der Erkrankung, wie bei Suizidalität, von Bedeutung ist. Die signifikante Überlegenheit gegenüber der Vergleichstherapie und damit der Zusatznutzen bleibt bis zum Ende der Beobachtung erhalten. Die Responderanalysen sowie auch TTE-Analysen der HAMD-17 sowie auch der MADRS bezüglich einer Remission unter Zuranolon zeigen ebenfalls signifikante Vorteile gegenüber der Vergleichstherapie in beträchtlichem Ausmaß. Langfristige Daten und Real-World-Analysen unterstützen die Stabilität der Remission über den klinischen Studienzeitraum hinaus. Die Ergebnisse zeigen, dass Patientinnen, die eine Remission erreichen, in der Regel über längere Zeiträume stabil bleiben, was durch longitudinale Studien und reale Versicherungsdaten gestützt wird [6]. Dies unterstreicht die Bedeutung einer Kurzzeitbehandlung mit Zuranolon als schnelle und effektive Therapieoption für Patientinnen mit PPD, die eine nachhaltige Symptomlinderung bietet.

Auch für die Instrumente EPDS, PHQ-9 und HAM-A zeigte sich zum Ende der Behandlung ein deutlicher Zusatznutzen. Der Effekt war über die TTE-Analysen für PHQ-9 und HAM-A signifikant schneller eingetreten und blieb bei allen drei Endpunkten bis zum Ende der Behandlung als signifikante Überlegenheit sichtbar.

Folglich ergibt sich für Zuranolon in der Endpunktkategorie Morbidität ein **beträchtlicher Zusatznutzen.**

Einschätzung von Kliniker*innen: Schnelle Abnahme der Symptome und der Krankheitsschwere

Die CGI-Daten liefern eine klinische Einschätzung der Reduktion des Schweregrads der PPD. Die klinische Einschätzung, wie sie durch Skalen wie CGI-I und CGI-S erfolgt, ermöglicht es, eine objektive Verbesserung der Patientin im Kontext der klinischen Praxis zu bewerten. Diese Einschätzungen sind oft entscheidend für die Therapieanpassung und das Management der Erkrankung. Die Ergebnisse zeigen statistisch signifikante Überlegenheit von Zuranolon und ergänzen die patientenberichteten Endpunkte und die Ergebnisse der Meta-Analyse, insbesondere in Bezug auf die schnell einsetzende Symptomlinderung und die Reduktion des Schweregrads der Erkrankung.

Lebensqualität: geringer Zusatznutzen

Zusätzlich zur Morbidität zeigte sich der Zusatznutzen auch für die Lebensqualität. Patientinnen berichten statistisch signifikant häufiger Verbesserungen in den SF-36-Subskalen „körperlicher Schmerz", „soziale Funktion" und „mentale Gesundheit". Dies wird als **geringer Zusatznutzen** eingestuft.

Sicherheit: transiente milde bis moderate Nebenwirkungen mit rascher Abnahme

Das Sicherheitsprofil von Zuranolon ist insgesamt gut. Schwere oder schwerwiegende UE sind insgesamt sehr selten. Die im Vergleich zum Kontrollarm häufiger auftretenden UE, wie Somnolenz und Sedierung waren mild bis moderat und traten vorwiegend lediglich in den ersten 1 – 2 Tagen auf und nahmen dann ab [6]. Diese Ereignisse sind gut behandelbar, bzw. werden gut toleriert, was durch die niedrige Rate von Therapieabbrüchen aufgrund von UE bestätigt wird.

Zusammenfassung

Insgesamt belegen die vorgelegten Daten in Form der Meta-Analyse als höchster Evidenzstufe Ia und der Einzelstudien, dass durch die Anwendung von Zuranolon nicht nur die Symptome der PPD signifikant gelindert werden, sondern auch ein wichtiger Beitrag zur Verbesserung der Lebensqualität der betroffenen Mütter und zur Förderung einer gesunden Bindung zwischen Mutter und Säugling geleistet wird. Die vorgelegten Daten unterstreichen die schnelle und über die gesamte Studie anhaltende Wirksamkeit einer Zuranolon+PiT-Behandlung bei der mittelschweren bis schweren PPD und resultieren in einem **Beleg für einen beträchtlichen Zusatznutzen.**

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 4.3.2 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 4-114: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens

Bezeichnung der Patientengruppen	Ausmaß des Zusatznutzens
a) Patientinnen mit **leichter** PPD nach der Geburt*	Nicht belegt
b) Patientinnen mit **mittelschwerer bis schwerer** PPD nach der Geburt	Beträchtlich
* Zu dieser Patientenpopulation liegen keine Daten vor, da diese oft durch therapeutische Interventionen oder unterstützende Maßnahmen wie körperliche Aktivität und nicht medikamentös behandelt wird.	

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.

Sofern diesbezüglich Informationen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

4.5.2 Begründung für die Vorlage nicht randomisierter vergleichender Studien und weiterer Untersuchungen

Sofern mit dem Dossier nicht randomisierte 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 beziehungsweise Untersuchungen behandelten Fragestellungen Studien höchster Evidenzstufe (randomisierte klinische Studien) durchzuführen oder zu fordern.

Sofern diesbezüglich Informationen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

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.

Sofern diesbezüglich Informationen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

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 beziehungsweise Aussagen zu patientenrelevanten Endpunkten zulassen.

Eine Validierung von Surrogatendpunkten bedarf in der Regel einer Meta-Analyse von Studien, in denen sowohl Effekte auf den Surrogatendpunkt als auch Effekte auf den interessierenden patientenrelevanten Endpunkt untersucht wurden (Burzykowski 2005[20], Molenberghs 2010[21]).

Diese Studien müssen bei Patientenkollektiven und Interventionen durchgeführt worden sein, die Aussagen für das dem vorliegenden Antrag zugrundeliegende 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 STE (Burzykowski 2006[22]) 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 2006[23]) 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

[20] Burzykowski T (Ed.): The evaluation of surrogate endpoints. New York: Springer; 2005.
[21] 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.
[22] Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006; 5(3): 173-186.
[23] Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. Stat Med 2006; 25(2): 183-203.

– verwendete Methodik
– entsprechende Ergebnisse (abhängig von der Methode)
– Untersuchungen zur Robustheit
– gegebenenfalls 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 beziehungsweise Aussagen zu patientenrelevanten Endpunkten zulassen.

Sofern Informationen zu Surrogatendpunkten im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Nicht zutreffend.

4.6 Referenzliste

Geben Sie nachfolgend alle Quellen (zum Beispiel Publikationen, Studienberichte, Studienregistereinträge), die Sie im vorliegenden Dokument zitiert haben (als fortlaufend nummerierte Liste). Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (zum Beispiel Vancouver oder Harvard). Geben Sie bei Fachinformationen, sofern vorhanden einen Link (zum Beispiel: https://www.fachinfo.de/) und immer den Stand des Dokuments an.

Sollten zu Nachweisen aus dem EU-Dossier, die Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, in den vorhergehenden Abschnitten Quellen im EU-Dossier hinterlegt sein, ist auf diese zu verweisen. Hierfür sind die Vorgaben zur Aufbereitung von Verweisen in Modul 5 in den Abschnitten 1.3 und 4.1 des Dokumentes zur Erstellung und Einreichung eines Dossiers (Anlage II.1) zu beachten.

Beispielhafte Zitierempfehlungen

Zeitschriftenartikel:

Flinn IW, Miller CB, Ardeshna KM et al. DYNAMO: A Phase II Study of Duvelisib (IPI-145) in Patients With Refractory Indolent Non-Hodgkin Lymphoma. J Clin Oncol 2019; 37(11): 912-922. https://doi.org/10.1200/jco.18.00915.

Online Quelle:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden; Version 7.0 [online]. 2023 [Zugriff: 16.02.2024]. URL: https://www.iqwig.de/methoden/allgemeine-methoden_version-7-0.pdf.

Monografie:

Braun J, Müller-Wieland D, Renz-Polster H, Krautzig S. Basislehrbuch Innere Medizin. München: Elsevier; 2018.

Buchbeitrag

Felser G, Klemperer D. Psychologische Aspekte von Interessenkonflikten. In: Lieb K, Klemperer D, Ludwig WD (Ed). Interessenkonflikte in der Medizin; Hintergründe und Lösungsmöglichkeiten. Berlin: Springer; 2011. S. 27-45.

Unveröffentlichter Studienbericht

Bristol Myers Squibb. A Phase 3 Randomized, Double-blind, Multi-center Study of Adjuvant Nivolumab versus Placebo in Subjects with High Risk Invasive Urothelial Carcinoma; study CA209274; Primary Clinical Study Report [unveröffentlicht]. 2020.

1. Biogen Netherlands B.V. 2025. Fachinformation Zurzuvae™ - Stand: 10/2025.
2. Pearlstein, T., Howard, M., Salisbury, A. & Zlotnick, C. 2009. Postpartum depression. *Am. J. Obstet. Gynecol.,* 200**,** 357–64.
3. Slomian, J., Honvo, G., Emonts, P., Reginster, J. Y. & Bruyère, O. 2019. Consequences of maternal postpartum depression: A systematic review of maternal and infant outcomes. *Women's health (London, England),* 15**,** 1745506519844044.
4. Campbell, S. B. & Cohn, J. F. 1997. The timing and chronicity of postpartum depression: Implications for infant development. *Postpartum depression and child development.* New York, NY, US: Guilford Press.
5. Glavin, K., Smith, L., Sørum, R. & Ellefsen, B. 2010. Redesigned community postpartum care to prevent and treat postpartum depression in women--a one-year follow-up study. *J. Clin. Nurs.,* 19**,** 3051–62.
6. European Medicines Agency (EMA) 2025. Question 88 and Applicant's Response.
7. Biogen 2025. 2.7.4 Summary of Clinical Safety.
8. Goodman, J. H. 2004. Paternal postpartum depression, its relationship to maternal postpartum depression, and implications for family health. *J. Adv. Nurs.,* 45**,** 26–35.
9. Nilova, V., Ward, L. & Hall, P. 2017. Women's experiences of parenting toddlers following postnatal depression. *Aust. J. Psychol.,* 69**,** 192–9.
10. Burgess, A. 2011. Fathers' roles in perinatal mental health: causes, interactions and effects. *New Digest 53, NCT, the Fatherhood Institute.*
11. Yu, H., Shen, Q., Bränn, E., Yang, Y., Oberg, A. S., Valdimarsdóttir, U. A. & Lu, D. 2024. Perinatal Depression and Risk of Suicidal Behavior. *JAMA Netw Open,* 7**,** e2350897.
12. Savitz, D. A., Stein, C. R., Ye, F., Kellerman, L. & Silverman, M. 2011. The epidemiology of hospitalized postpartum depression in New York State, 1995-2004. *Ann. Epidemiol.,* 21**,** 399–406.
13. Bundesärztekammer (BÄK), Kassenärztliche Bundesvereinigung (KBV) & Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) 2022. Nationale VersorgungsLeitlinie Unipolare Depression – Langfassung, Version 3.2. 2022
14. Sage Therapeutics 2022. Clinical Study Report: A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY EVALUATING THE EFFICACY AND

SAFETY OF SAGE-217 IN THE TREATMENT OF ADULTS WITH SEVERE POSTPARTUM DEPRESSION - 217-PPD-301 (SKYLARK) - Version 1.

15. Sage Therapeutics 2019. Clinical Study Report: A Multicenter, Randomized, Double-Blind, Parallel-Group, Placebo-Controlled Study Evaluating the Efficacy, Safety, and Pharmacokinetics of SAGE-217 in the Treatment of Adult Female Subjects with Severe Postpartum Depression - 217-PPD-201 (ROBIN).
16. Gemeinsamer Bundesausschuss (G-BA) 2025. Niederschrift (finale Fassung) zum Beratungsgespräch gemäß § 8 AM-NutzenV - Beratungsanforderung 2025-B-099 - Zuranolon zur Behandlung von Erwachsenen nach der Geburt mit einer postpartalen Depression (PPD).
17. Sage Therapeutics Inc. 2018. Protocol: A Multicenter, Randomized, Double-Blind, Parallel-Group, Placebo-Controlled Study Evaluating the Efficacy, Safety, and Pharmacokinetics of SAGE-217 in the Treatment of Adult Female Subjects With Severe Postpartum Depression - Protocol Number: 217-PPD-201 (ROBIN) - IND Number 132,131 - Amendment 8 - Version 9.0.
18. Sage Therapeutics, I. 2018. Statistical Analysis Plan - Methods - Protocol 217-PPD-201 (ROBIN): A MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PARALLEL-GROUP, PLACEBO-CONTROLLED STUDY EVALUATING THE EFFICACY, SAFETY, AND PHARMACOKINETICS OF SAGE-217 IN THE TREATMENT OF ADULT FEMALE SUBJECTS WITH SEVERE POSTPARTUM DEPRESSION - Version 1.0.
19. Sage Therapeutics Inc. 2021. Clinical Protocol: A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY EVALUATING THE EFFICACY AND SAFETY OF SAGE-217 IN THE TREATMENT OF ADULTS WITH SEVERE POSTPARTUM DEPRESSION - PROTOCOL NUMBER: 217-PPD-301 (SKYLARK) - EUDRACT NUMBER: 2020-001424-34 - Version 2.
20. Sage Therapeutics Inc. 2022. Statistical Analysis Plan Methods - Protocol Number 217-PPD-301 (SKYLARK): A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY EVALUATING THE EFFICACY AND SAFETY OF SAGE-217 IN THE TREATMENT OF ADULTS WITH SEVERE POSTPARTUM DEPRESSION - Version: Final Version 1.0 - Version Date of SAP: 11 April 2022.
21. Hamilton, M. 1960. A rating scale for depression. *J Neurol Neurosurg Psychiatry,* 23**,** 56–62.
22. DGPPN, BÄK, Awmf & Kbv 2015. S3-Leitlinie/Nationale VersorgungsLeitlinie Unipolare Depression, Version 5, 2. Auflage.
23. AG Diabetes und Psychologie, Deutsche Diabetes Gesellschaft (DDG) & Diabetes und Psychologie e. V. 2022. Hamilton Rating Scale for Depression (HAMD). https://www.diabetes-psychologie.de/downloads/Beschreibung_HAMD.pdf.
24. Gerbasi, M. E., Eldar-Lissai, A., Acaster, S., Fridman, M., Bonthapally, V., Hodgkins, P., Kanes, S. J. & Meltzer-Brody, S. 2020. Associations between commonly used patient-reported outcome tools in postpartum depression clinical practice and the Hamilton Rating Scale for Depression. *Archives of women's mental health,* 23**,** 727–35.
25. Ji, S., Long, Q., Newport, D. J., Na, H., Knight, B., Zach, E. B., Morris, N. J., Kutner, M. & Stowe, Z. N. 2011. Validity of depression rating scales during pregnancy and the postpartum period: impact of trimester and parity. *J. Psychiatr. Res.,* 45**,** 213–9.
26. Trajković, G., Starčević, V., Latas, M., Leštarević, M., Ille, T., Bukumirić, Z. & Marinković, J. 2011. Reliability of the Hamilton Rating Scale for Depression: A meta-analysis over a period of 49years. *Psychiatry Res.,* 189**,** 1–9.

27. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) 2009. Selektive Serotonin- und Noradrenalin Wiederaufnahmehemmer (SNRI) bei Patienten mit Depressionen.
28. Frank, E., Prien, R. F., Jarrett, R. B., Keller, M. B., Kupfer, D. J., Lavori, P. W., Rush, A. J. & Weissman, M. M. 1991. Conceptualization and rationale for consensus definitions of terms in major depressive disorder. Remission, recovery, relapse, and recurrence. *Arch. Gen. Psychiatry,* 48**,** 851–5.
29. Gemeinsamer Bundesausschuss (G-BA) 2022. Niederschrift (finale Fassung) zum Beratungsgespräch gemäß § 8 AM-NutzenV - Beratungsanforderung 2022-B-094 - Zuranolon zur Behandlung von erwachsenen Frauen mit einer postpartalen Depression.
30. Montgomery, S. A. & Asberg, M. 1979. A new depression scale designed to be sensitive to change. *The British journal of psychiatry : the journal of mental science,* 134**,** 382–9.
31. Weyer, G. 2015. Internationale Skalen für Psychiatrie. Hogrefe Verlag 2015.
32. Maust, D., Cristancho, M., Gray, L., Rushing, S., Tjoa, C. & Thase, M. E. 2012. Psychiatric rating scales. *Handb. Clin. Neurol.,* 106**,** 227–37.
33. Kearns, N. P., Cruickshank, C. A., McGuigan, K. J., Riley, S. A., Shaw, S. P. & Snaith, R. P. 1982. A comparison of depression rating scales. *The British journal of psychiatry : the journal of mental science,* 141**,** 45–9.
34. Maier, W. & Philipp, M. 1985. Comparative analysis of observer depression scales. *Acta Psychiatr. Scand.,* 72**,** 239–45.
35. Hawley, C. J., Gale, T. M. & Sivakumaran, T. 2002. Defining remission by cut off score on the MADRS: selecting the optimal value. *J. Affect. Disord.,* 72**,** 177–84.
36. Riedel, M., Möller, H. J., Obermeier, M., Schennach-Wolff, R., Bauer, M., Adli, M., Kronmüller, K., Nickel, T., Brieger, P., Laux, G., Bender, W., Heuser, I., Zeiler, J., Gaebel, W. & Seemüller, F. 2010. Response and remission criteria in major depression--a validation of current practice. *J. Psychiatr. Res.,* 44**,** 1063–8.
37. Keller, M. B. 2003. Past, present, and future directions for defining optimal treatment outcome in depression: remission and beyond. *JAMA,* 289**,** 3152–60.
38. Montgomery, S. A., Huusom, A. K. & Bothmer, J. 2004. A randomised study comparing escitalopram with venlafaxine XR in primary care patients with major depressive disorder. *Neuropsychobiology,* 50**,** 57–64.
39. Gemeinsamer Bundesausschuss (G-BA) 2021. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie (AM RL): Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V Esketamin (Depression, akute Kurzzeitbehandlung, Kombinationstherapie).
40. Gemeinsamer Bundesausschuss (G-BA) 2023. Tragende Gründe zum Beschluss des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie: Anlage XII – Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a des Fünften Buches Sozialgesetzbuch (SGB V) Esketamin Neubewertung nach Fristablauf (Depression, therapieresistent, in Kombination mit SSRI oder SNRI)) - Vom 21. September 2023.
41. Cox, J. L., Holden, J. M. & Sagovsky, R. 1987. Detection of postnatal depression. Development of the 10-item Edinburgh Postnatal Depression Scale. *The British journal of psychiatry : the journal of mental science,* 150**,** 782–6.
42. Sultan, P., Ando, K., Elkhateb, R., George, R. B., Lim, G., Carvalho, B., Chitneni, A., Kawai, R., Tulipan, T., Blake, L., Coker, J. & O'Carroll, J. 2022. Assessment of Patient-

Reported Outcome Measures for Maternal Postpartum Depression Using the Consensus-Based Standards for the Selection of Health Measurement Instruments Guideline: A Systematic Review. *JAMA Netw Open,* 5**,** e2214885.
43. Spitzer, R. L., Kroenke, K. & Williams, J. B. 1999. Validation and utility of a self-report version of PRIME-MD: the PHQ primary care study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA,* 282**,** 1737–44.
44. Kroenke, K., Spitzer, R. L. & Williams, J. B. 2001. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.,* 16**,** 606–13.
45. Levis, B., Benedetti, A. & Thombs, B. D. 2019. Accuracy of Patient Health Questionnaire-9 (PHQ-9) for screening to detect major depression: individual participant data meta-analysis. *BMJ,* 365**,** l1476.
46. Clark, D. B. & Donovan, J. E. 1994. Reliability and validity of the Hamilton Anxiety Rating Scale in an adolescent sample. *J. Am. Acad. Child Adolesc. Psychiatry,* 33**,** 354–60.
47. Vaccarino, A. L., Evans, K. R., Sills, T. L. & Kalali, A. H. 2008. Symptoms of anxiety in depression: assessment of item performance of the Hamilton Anxiety Rating Scale in patients with depression. *Depress. Anxiety,* 25**,** 1006–13.
48. Zimmerman, M., Martin, J., Clark, H., McGonigal, P., Harris, L. & Holst, C. G. 2017. Measuring anxiety in depressed patients: A comparison of the Hamilton anxiety rating scale and the DSM-5 Anxious Distress Specifier Interview. *J. Psychiatr. Res.,* 93**,** 59–63.
49. Hamilton, M. 1959. The assessment of anxiety states by rating. *Br. J. Med. Psychol.,* 32**,** 50–5.
50. Ware, J., Jr., Kosinski, M. & Keller, S. D. 1996. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med. Care,* 34**,** 220–33.
51. Ware, J. E., Jr. & Sherbourne, C. D. 1992. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med. Care,* 30**,** 473–83.
52. 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. *Med. Care,* 31**,** 247–63.
53. Barkin, J. L., Wisner, K. L., Bromberger, J. T., Beach, S. R., Terry, M. A. & Wisniewski, S. R. 2010. Development of the Barkin Index of Maternal Functioning. *J Womens Health (Larchmt),* 19**,** 2239–46.
54. Gemeinsamer Bundesausschuss (G-BA) 2024. Nutzenbewertung von Tofersen, Neues Anwendungsgebiet: Behandlung von Erwachsenen mit amyotropher Lateralsklerose, die mit einer Mutation im Superoxid-Dismutase-1-Gen assoziiert ist.
55. Mundt, J. C., Greist, J. H., Gelenberg, A. J., Katzelnick, D. J., Jefferson, J. W. & Modell, J. G. 2010. Feasibility and validation of a computer-automated Columbia-Suicide Severity Rating Scale using interactive voice response technology. *J. Psychiatr. Res.,* 44**,** 1224–8.
56. Posner, K., Brown, G. K., Stanley, B., Brent, D. A., Yershova, K. V., Oquendo, M. A., Currier, G. W., Melvin, G. A., Greenhill, L., Shen, S. & Mann, J. J. 2011. The Columbia-Suicide Severity Rating Scale: initial validity and internal consistency findings from three multisite studies with adolescents and adults. *Am. J. Psychiatry,* 168**,** 1266–77.
57. The Cochrane Collaboration 2025. About IPD meta-analyses [Online]. Verfügbar unter: https://methods.cochrane.org/ipdma/about-ipd-meta-analyses.

58. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) 2025. Allgemeine Methoden - Version 8.0 - vom 19.12.2025.
59. Deligiannidis, K. M., Citrome, L., Huang, M. Y., Acaster, S., Fridman, M., Bonthapally, V., Lasser, R. & Kanes, S. J. 2023. Effect of Zuranolone on Concurrent Anxiety and Insomnia Symptoms in Women With Postpartum Depression. *J. Clin. Psychiatry,* 84.
60. Deligiannidis, K. M., Meltzer-Brody, S., Gunduz-Bruce, H., Doherty, J., Jonas, J., Li, S., Sankoh, A. J., Silber, C., Campbell, A. D., Werneburg, B., Kanes, S. J. & Lasser, R. 2021. Effect of Zuranolone vs Placebo in Postpartum Depression: A Randomized Clinical Trial. *JAMA Psychiatry,* 78**,** 951–9.
61. Deligiannidis, K. M., Meltzer-Brody, S., Maximos, B., Peeper, E. Q., Freeman, M., Lasser, R., Bullock, A., Kotecha, M., Li, S., Forrestal, F., Rana, N., Garcia, M., Leclair, B. & Doherty, J. 2023. Zuranolone for the Treatment of Postpartum Depression. *Am. J. Psychiatry,* 180**,** 668–75.
62. clinicaltrials.gov (CT) Registereintrag 2016. A Multicenter, Randomized, Double-Blind, Parallel-Group, Placebo-Controlled Study Evaluating the Efficacy, Safety, and Pharmacokinetics of SAGE-217 in the Treatment of Adult Female Subjects With Severe Postpartum Depression. https://clinicaltrials.gov/study/NCT02978326.
63. clinicaltrials.gov (CT) Registereintrag 2020. A Randomized, Double-Blind, Placebo-controlled Study Evaluating the Efficacy and Safety of SAGE-217 in the Treatment of Adults With Severe Postpartum DepressionA Multicenter, Randomized, Double-Blind, Parallel-Group, Placebo-Controlled Study Evaluating the Efficacy, Safety, and Pharmacokinetics of SAGE-217 in the Treatment of Adult Female Subjects With Severe Postpartum Depression. https://clinicaltrials.gov/study/NCT04442503.
64. EU-CTR Registereintrag 2020. A Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Efficacy and Safety of Sage-217 in the Treatment of Adults with Severe Postpartum Depression. *2020-001424-34.* https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2020-001424-34.
65. European Medicines Agency (EMA) 2025. CHMP D180 List of outstanding issues.
66. Müller, M. J., Himmerich, H., Kienzle, B. & Szegedi, A. 2003. Differentiating moderate and severe depression using the Montgomery-Asberg depression rating scale (MADRS). *J. Affect. Disord.,* 77**,** 255–60.
67. Leucht, S., Fennema, H., Engel, R. R., Kaspers-Janssen, M. & Szegedi, A. 2018. Translating the HAM-D into the MADRS and vice versa with equipercentile linking. *J. Affect. Disord.,* 226**,** 326–31.
68. Biogen GmbH 2026. Zusatzanalysen zum Schweregrad der PPD in ROBIN und SKYLARK.
69. Statistisches Bundesamt (Destatis) 2025. Daten zum durchschnittlichen Alter der Mutter bei Geburt insgesamt und 1. Kind nach Bundesländern.
70. Research Network in Peripartum Depression Disorder (Riseup-PPD) 2024. Evidence-based practice guidelines for prevention, screening and treatment of peripartum depression.
71. American College of Obstetricians and Gynecologists (ACOG) 2023. Treatment and Management of Mental Health Conditions During Pregnancy and Postpartum: ACOG Clinical Practice Guideline No. 5. *Obstetrics and gynecology,* 141**,** 1262–88.
72. Biogen GmbH 2025. Zusatzanalysen-Dokument für die G-BA Einreichung.

Anhang 4-A: Suchstrategien – bibliografische Recherche

Geben Sie nachfolgend die Suchstrategien für die bibliografische(n) Recherche(n) an, und zwar getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche et cetera). Für jede durchsuchte Datenbank ist die verwendete Strategie separat darzustellen. Geben Sie dabei zunächst jeweils den Namen der durchsuchten Datenbank (zum Beispiel Medline), die verwendete Suchoberfläche (zum Beispiel NML, Ovid et cetera), das Datum der Suche, das Zeitsegment (zum Beispiel: „1946 to 2024 week 07“) und die gegebenenfalls verwendeten Suchfilter (mit Angabe einer Quelle) an. Listen Sie danach die Suchstrategie einschließlich der resultierenden Trefferzahlen auf. Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel (eine umfassende Suche soll Freitextbegriffe und Schlagwörter enthalten).

Wird im Falle einer vorangegangenen klinischen Bewertung nach der Verordnung (EU) 2021/2282 auf die Recherche im EU-Dossier durch Verweis Bezug genommen und liegt die Recherche im EU-Dossier mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt, ist die Recherche zu aktualisieren. Die aktualisierte Recherche ist im Dossier darzustellen.

Datenbankname	Medline
Suchoberfläche	Ovid
Datum der Suche	23.08.2024
Zeitsegment	ALL <1946 to August 23, 2024>
Suchfilter	Lefebvre [Quelle[24]] – Cochrane Highly Sensitive Search Strategy for identifying randomized trials in MEDLINE: sensitivity- and precision-maximizing version (2023 revision)

#	**Suchbegriffe**	Ergebnis
1	exp Myocardial Infarction/ or Acute Coronary Syndrome/	215276
2	exp Percutaneous Coronary Intervention/	67684
3	((acute adj1 coronary adj1 syndrome*) or (myocardial adj1 infarction)).ti,ab.	243348
4	(percutaneous adj1 coronary adj1 intervention*).ti,ab.	46088
5	or/1-4	354183
6	ticagrelor*.mp.	4448
7	exp randomized controlled trial/	621265
8	controlled clinical trial.pt.	95592
9	(randomized or placebo or randomly).ab.	1146056
10	clinical trials as topic.sh.	203134
11	trial.ti.	316232
12	or/7-11	1622239
13	exp animals/ not humans.sh.	5251605
14	12 not 13	1494779
15	and/5-6,14	1144

[24] Das Zitat zu dem hier beispielhaft angegebenen Suchfilter lautet wie folgt: Lefebvre C, Glanville J, Briscoe S et al. Cochrane Handbook for Systematic Reviews of Interventions; Version 6.4; Technical Supplement to Chapter 4: Searching for and selecting studies [online]. 2024 [Zugriff: 21.02.2024]. URL: https://training.cochrane.org/handbook/current/chapter-04-technical-supplement-searching-and-selecting-studies. Hinweis: Für die Suche in der Cochrane-Datenbank „Cochrane Central Register of Controlled Trials (Clinical Trials)“ sollte kein Studienfilter verwendet werden.

Anhang 4-A1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Datenbankname	Medline
Suchoberfläche	PubMed
Datum der Suche	16.04.2026
Zeitsegment	Keine Einschränkungen
Suchfilter	Cochrane Highly Sensitive Search Strategy for identifying randomized trials in MEDLINE: sensitivity- and precision-maximizing version (2008 revision); PubMed format

Zeile	Suche	Treffer
#1	Zuranolone[tiab] OR SAGE-217[tiab] OR SAGE217[tiab] OR Zurzuvae[tiab] OR BIIB125[tiab] OR "zuranolone" [Supplementary Concept]	176
#2	Depression, Postpartum[Mesh]	9100
#3	("postpart*"[tiab] OR "postnatal"[tiab] OR "peripart*"[tiab]) AND ("Depression"[MeSH Terms] OR "Depression"[tiab])	18779
#4	#2 OR #3	20280
#5	("randomized controlled trial"[pt] OR "controlled clinical trial"[pt] OR "randomized"[tiab] OR "placebo"[tiab] OR "clinical trials as topic"[MeSH Terms:noexp] OR "randomly"[tiab] OR "trial"[ti]) NOT ("animals"[MeSH] NOT "humans"[MeSH])	1652010
#6	#1 AND #4 AND #5	36

Datenbankname	Cochrane Library
Suchoberfläche	Cochrane Library
Datum der Suche	16.04.2026
Zeitsegment	Keine Einschränkungen
Suchfilter	Kein Filter

Zeile	Suche	Treffer
#1	Zuranolone:ti,ab OR SAGE-217:ti,ab OR SAGE217:ti,ab OR Zurzuvae:ti,ab OR BIIB125	99
#2	[mh "Depression, Postpartum"]	1167
#3	(postpart*:ti,ab OR postnatal:ti,ab OR peripart*:ti,ab) AND ("Depression":ti,ab)	3904
#4	#2 OR #3	4054
#5	#1 AND #4	54

Cochrane Reviews	1
Cochrane Protocols	0
Trials	53
Editorials	0
Special Collections	0
Clinical Answers	0

Anhang 4-A2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend.

Anhang 4-A3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend.

Anhang 4-A4: Suche nach weiteren Untersuchungen

Nicht zutreffend.

Anhang 4-B: Suchstrategien – Suche in Studienregistern/Studienergebnisdatenbanken

Geben Sie nachfolgend die Suchstrategien für die Suche(n) in Studienregistern/ Studienergebnisdatenbanken an. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche et cetera) wie unten angegeben. Für jede/s durchsuchte Studienregister/Studienergebnisdatenbank ist eine separate Strategie darzustellen. Geben Sie dabei jeweils den Namen des durchsuchten Studienregisters/Studienergebnisdatenbank (zum Beispiel clinicaltrials.gov), die Internetadresse, unter der das/die Studienregister/Studienergebnisdatenbank erreichbar ist (zum Beispiel http://www.clinicaltrials.gov), das Datum der Suche, die verwendete Suchstrategie und die resultierenden Treffer an. Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel.

Wird im Falle einer vorangegangenen klinischen Bewertung nach der Verordnung (EU) 2021/2282 auf die Recherche im EU-Dossier durch Verweis Bezug genommen und liegt die Recherche im EU-Dossier mehr als drei Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt, ist die Recherche zu aktualisieren. Die aktualisierte Recherche ist im Dossier darzustellen.

Studienregister/Studienergebnisdatenbank	ClinicalTrials.gov
Internetadresse	http://www.clinicaltrials.gov
Datum der Suche	23.08.2024
Eingabeoberfläche	Basic Search
Suchstrategie	(ticagrelor OR AZD-6140) AND (acute coronary syndrome OR myocardial infarction) [Other terms]
Treffer	452

Anhang 4-B1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Studienregister	ClinicalTrials.gov (CT)
Internetadresse	https://clinicaltrials.gov/
Datum der Suche	16.04.2026
Suchstrategie	**Intervention:** Zuranolone OR SAGE-217 OR SAGE217 OR Zurzuvae OR BIIB125 **Study type:** Interventional Studies
Treffer	22

Studienregister	EU Clinical Trials Register (EUCTR)
Internetadresse	https://www.clinicaltrialsregister.eu/ctr-search/search
Datum der Suche	16.04.2026
Suchstrategie	Zuranolone OR SAGE-217 OR SAGE217 OR Zurzuvae OR BIIB125
Treffer	2

Studienregister	Clinical Trials (CTIS)
Internetadresse	https://euclinicaltrials.eu/search-for-clinical-trials/?lang=en
Datum der Suche	16.04.2026
Suchstrategie	**Contain any of these terms:** Zuranolone, SAGE-217, SAGE217, Zurzuvae, BIIB125
Treffer	0

Anhang 4-B2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend.

Anhang 4-B3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend.

Anhang 4-B4: Suche nach weiteren Untersuchungen

Nicht zutreffend.

Anhang 4-C: Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Recherche)

Listen Sie nachfolgend die im Volltext gesichteten und ausgeschlossenen Dokumente aus der/den bibliografischen Recherche(n) auf. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche et cetera) wie unten angegeben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (zum Beispiel Vancouver oder Harvard) und nummerieren Sie die Zitate fortlaufend. Geben Sie jeweils einen Ausschlussgrund an und beziehen Sie sich dabei auf die im Abschnitt 4.2.2 genannten Ein- und Ausschlusskriterien.

Sofern Informationen zu im Volltext gesichteten und ausgeschlossenen Dokumente aus der/den bibliografischen Recherche(n) im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Anhang 4-C1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Nicht zutreffend.

Anhang 4-C2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend.

Anhang 4-C3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend.

Anhang 4-C4: Suche nach weiteren Untersuchungen

Nicht zutreffend.

Anhang 4-D: Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern/Studienergebnisdatenbanken)

Listen Sie nachfolgend die durch die Studienregistersuche(n)/Studienergebnisdatenbanksuche(n) identifizierten, aber ausgeschlossenen Registereinträgen auf. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche et cetera) wie unten angegeben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (zum Beispiel Vancouver oder Harvard) und nummerieren Sie die Zitate fortlaufend. Geben Sie jeweils einen Ausschlussgrund an und beziehen Sie sich dabei auf die im Abschnitt 4.2.2 genannten Ein- und Ausschlusskriterien.

Sofern Informationen zu durch die Studienregistersuche(n)/Studienergebnisdatenbanksuche(n) identifizierten, aber ausgeschlossenen Registereinträgen im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Anhang 4-D1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Nummer	Studien ID	Titel	Jahr	Link	Ein-/Ausschluss
Clinicaltrials.gov					
1	NCT02978781	A Phase 2a, Double-Blind, Placebo-Controlled, Randomized Withdrawal Study Evaluating the Efficacy, Safety, Tolerability, and Pharmacokinetics of SAGE-217 in the Treatment of Subjects With Essential Tremor	2016	https://clinicaltrials.gov/study/NCT02978781	A1 - Patientenpopulation
2	NCT03000530	A Phase 2, Two-Part (Open-Label Followed by Double-Blind) Study Evaluating the Safety, Tolerability, Pharmacokinetics, and Efficacy of SAGE-217 in the Treatment of Adult Subjects With Moderate to Severe Major Depressive Disorder	2016	https://clinicaltrials.gov/study/NCT03000530	A1 - Patientenpopulation
3	NCT02978326	A Multicenter, Randomized, Double-Blind, Parallel-Group, Placebo-Controlled Study Evaluating the Efficacy, Safety, and Pharmacokinetics of SAGE-217 in the Treatment of Adult Female Subjects With Severe Postpartum Depression	2016	https://clinicaltrials.gov/study/NCT02978326	eingeschlossen

4	NCT03000569	A Phase 2, Two-Part Study to Evaluate the Safety, Tolerability, Pharmacokinetics, and Efficacy of SAGE-217 in Subjects With Parkinson's Disease	2016	https://clinicaltrials.gov/study/NCT03000569	A1 - Patienten-population
5	NCT03284931	A Randomized, Double-Blind, Multiple Dose, 3-way Crossover, Exploratory Study to Assess the Pharmacodynamic Effects of SAGE-217 Capsules in Healthy Adults Using a 5-hour Phase Advance Model of Insomnia	2017	https://clinicaltrials.gov/study/NCT03284931	A1 - Patienten-population
6	NCT03672175	A Phase 3, Multicenter, Double-Blind, Randomized, Placebo-Controlled Study Evaluating the Efficacy of SAGE-217 in the Treatment of Adult Subjects With Major Depressive Disorder	2018	https://clinicaltrials.gov/study/NCT03672175	A1 - Patienten-population
7	NCT03771664	A Randomized, Double-blind, Placebo-controlled Study of the Safety, Tolerability, and Efficacy of SAGE-217 Compared to Placebo in Adult Subjects With Comorbid Major Depressive Disorder and Insomnia	2018	https://clinicaltrials.gov/study/NCT03771664	A1 - Patienten-population
8	NCT03692910	A 2-Part Study (Open-label Followed by Double-blind, Randomized, Placebo-controlled, Parallel Group) of the Safety, Tolerability, Pharmacokinetics, and Efficacy of SAGE-217 in the Treatment of Subjects With Bipolar I/II Disorder With a Current Major Depressive Episode	2018	https://clinicaltrials.gov/study/NCT03692910	A1 - Patienten-population
9	NCT04007367	A Phase 3, Randomized, Double-Blind, Placebo-controlled Study of the Efficacy and Safety of SAGE-217 With a Fixed, Repeated Treatment Regimen on Relapse Prevention in Adults With Major Depressive Disorder	2019	https://clinicaltrials.gov/study/NCT04007367	A1 - Patienten-population
10	NCT03864614	A Phase 3, Open-Label, 1-Year Study of the Safety, Tolerability, and Need for Re-Treatment With SAGE-217 in Adult Subjects With Major Depressive Disorder	2019	https://clinicaltrials.gov/study/NCT03864614	A1 - Patienten-population

11	NCT04442503	A Randomized, Double-Blind, Placebo-controlled Study Evaluating the Efficacy and Safety of SAGE-217 in the Treatment of Adults With Severe Postpartum Depression	2020	https://clinicaltrials.gov/study/NCT04442503	eingeschlossen
12	NCT04442490	A Phase 3, Multicenter, Double-Blind, Randomized, Placebo-Controlled Study Evaluating the Efficacy of SAGE-217 in the Treatment of Adult Subjects With Major Depressive Disorder	2020	https://clinicaltrials.gov/study/NCT04442490	A1 - Patienten-population
13	NCT04476030	A Phase 3, Randomized, Double-Blind Study Comparing the Efficacy and Safety of SAGE-217 Plus an Antidepressant Versus Placebo Plus an Antidepressant in Adults With Major Depressive Disorder	2020	https://clinicaltrials.gov/study/NCT04476030	A1 - Patienten-population
14	NCT05655507	A Phase 1, Study to Evaluate the Pharmacokinetics and Safety of Zuranolone in Adolescents (12 to 17 Years of Age) With Major Depressive Disorder	2022	https://clinicaltrials.gov/study/NCT05655507	A1 - Patienten-population
15	NCT06759558	Allopregnanolone (Zuranolone) in Post-stroke Depression	2024	https://clinicaltrials.gov/study/NCT06759558	A1 - Patienten-population
16	NCT05005338	An Open Label Phase 1 Study to Determine the Absolute Bioavailability of Single and Repeated Daily Oral Administration of BIIB122 (DNL151) Utilizing a Microtracer Dose of Intravenously Administered [14C] BIIB122 ([14C] DNL151) to Healthy Subjects	2021	https://clinicaltrials.gov/study/NCT05005338	A1 - Patienten-population
17	NCT05119790	An Open Label, Phase 1 Study of Mass Balance, Pharmacokinetics, Metabolite Profile, and Metabolite Identification of a Single Oral Dose of [14C]-BIIB122 ([14C] DNL151) in Healthy Male Subjects	2021	https://clinicaltrials.gov/study/NCT05119790	A1 - Patienten-population

18	NCT05229562	An Open-Label, Parallel-Group, Phase 1 Study to Evaluate the Pharmacokinetics, Safety, and Tolerability of BIIB122 in Healthy Adult Japanese, Chinese, and Caucasian Participants	2022	https://clinicaltrials.gov/study/NCT05229562	A1 - Patienten-population
19	NCT05348785	A Phase 2b, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Determine the Efficacy and Safety of BIIB122 in Participants With Parkinson's Disease	2022	https://clinicaltrials.gov/study/NCT05348785	A1 - Patienten-population
20	NCT05418673	A Phase 3, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Determine the Efficacy and Safety of BIIB122/DNL151 in Participants With Parkinson's Disease and Pathogenic LRRK2 Variants	2022	https://clinicaltrials.gov/study/NCT05418673	A1 - Patienten-population
21	NCT06602193	A Phase 2a, Multicenter, Randomized, Double-Blind, Placebo-Controlled Study to Evaluate the Safety and Pharmacodynamic Effects of BIIB122 in Participants With LRRK2-Associated Parkinson's Disease (LRRK2-PD)	2024	https://clinicaltrials.gov/study/NCT06602193	A1 - Patienten-population
22	NCT06264440	A Phase 1, Open-Label, Crossover Study to Assess the Potential for Gastric pH-Dependent Drug-Drug Interactions of BIIB122 With a Proton Pump Inhibitor in Healthy Participants	2024	https://clinicaltrials.gov/study/NCT06264440	A1 - Patienten-population
EU-CTR					
1	2019-002640-25	A Phase 3, Randomized, Double-Blind, Placebo-Controlled Study of the Efficacy and Safety of Sage-217 with Fixed, Repeated Treatment Regimen on Relapse Prevention in Adults with Major Depressive Disorder	2019	https://www.clinicaltrialsregister.eu/ctr-search/search?query=eudract_number:2019-002640-25	A1 - Patienten-population
2	2020-001424-34	A Randomized, Double-Blind, Placebo-Controlled Study Evaluating the Efficacy and Safety of Sage-217 in the Treatment of	2020	https://www.clinicaltrialsregister.eu/ctr-search/search?q	eingeschlossen

		Adults with Severe Postpartum Depression		uery=eudract_number:2020-001424-34	

Anhang 4-D2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend.

Anhang 4-D3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend.

Anhang 4-D4: Suche nach weiteren Untersuchungen

Nicht zutreffend.

Anhang 4-E: Methodik der eingeschlossenen Studien – RCT

Beschreiben Sie nachfolgend die Methodik jeder eingeschlossenen, in Abschnitt 4.3.1.1.5 genannten Studie. Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle 4-115 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 4-115 inklusive eines Flow-Charts für den Patientenfluss.

Sofern Informationen zur Methodik jeder eingeschlossenen, in Abschnitt 4.3.1.1.5 beziehungsweise 4.3.1.2 genannten Studie im EU-Dossier hinterlegt sind und diese Grundlage der Nutzenbewertung nach § 35a SGB V sein sollen, ist auf die entsprechenden Abschnitte des EU-Dossiers zu verweisen.

Studie: ROBIN, Teil B

Tabelle 4-115 (Anhang): Studiendesign und -methodik für Studie ROBIN

Item[a]	Charakteristikum	Studieninformation
Studienziel		
2 b	Genaue Ziele, Fragestellung und Hypothesen	**Primäres Ziel:** Das primäre Ziel der Studie war es, zu bestimmen, ob die Behandlung mit Zuranolon depressive Symptome bei Patientinnen mit schwerer PPD im Vergleich zu Placebo reduziert. Dies wurde anhand der Veränderung des HAMD-Gesamtwerts vom Ausgangswert bis Tag 15 bewertet. **Sekundäre Ziele:** - Bewertung der Reduktion depressiver Symptome durch Zuranolon im Vergleich zu Placebo anhand: • Veränderung des HAMD-Gesamtwerts gegenüber dem Score zu Baseline zu allen anderen Zeitpunkten. • HAMD-Ansprechen. • HAMD-Remission. • Veränderung im MADRS-Gesamtwert zu Tag 15 gegenüber dem Score zu Baseline. • Verbesserung/ Ansprechen via CGI-I. • Veränderungen von der Baseline in den HAMD-Subskalen und einzelnen Item-Scores am Tag 15 und allen anderen Zeitpunkten - Bewertung der Reduktion von Angstsymptomen durch Zuranolon im Vergleich zu Placebo anhand der Veränderung des HAM-A-Gesamtwerts. - Bewertung der Sicherheit und Verträglichkeit von Zuranolon im Vergleich zu Placebo durch die Häufigkeit von UE, Vitalzeichen, Laborwerten, EKG-Parametern und der C-SSRS. **Explorative Ziele:** - Die Nutzung von Gesundheitsressourcen (Healthcare Resource Utilization; HCRU) zu Beginn und am Tag 45 zu bewerten. - Untersuchung der Auswirkungen von Zuranolon im Vergleich zu Placebo (an Tag 15 und zu allen anderen Zeitpunkten) auf: • Subjektiv berichtete depressive Symptome (EPDS; PHQ-9). • Mütterliches Verhalten (Barkin Index of Maternal Functioning [BIMF]). • Allgemeinen Gesundheitsstatus (SF-36). - Das PK-Ziel dieser Studie war es, das PK-Profil von Zuranolon in Plasmaproben und die Konzentration von

Item[a]	Charakteristikum	Studieninformation
		Zuranolon in der Muttermilch, wenn möglich, nach der Verabreichung von Zuranolon zu bewerten.
Methoden		
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	Die ROBIN-Studie ist als multizentrische, randomisierte, doppelblinde, parallel-gruppierte, placebokontrollierte Studie konzipiert. Das Zuteilungsverhältnis der Randomisierung betrug 1 : 1, wobei die Teilnehmerinnen entweder Zuranolon oder Placebo erhielten. Die Studie bestand aus einer Screening-Periode, einer Behandlungsperiode und einer Follow-Up-Periode. Während der Screening-Periode wurde festgestellt ob die Teilnehmerinnen gemäß DSM-5 SCIDI eine PPD aufweisen. Außerdem wurden die Teilnehmerinnen hinsichtlich Ein- und Ausschlusskriterien überprüft. Wenn die Teilnehmerinnen die Kriterien der Screening-Periode erfüllten, konnten sie in die Behandlungsperiode übergehen und wurden 1 : 1 in den Zuranolon-Arm oder Placebo-Arm randomisiert. Die orale Einnahme von 30 mg Zuranolon oder Placebo erfolgte einmal täglich über einen Zeitraum von 14 Tagen. In diesem Zeitraum wurden Wirksamkeits- und Sicherheitsendpunkte regelmäßig erhoben, und Blutproben wurden zur Analyse von Zuranolon, Metaboliten von Zuranolon und HPβCD gemäß dem Zeitplan entnommen. In der Follow-Up-Periode fanden 1 Woche (entspricht: 21 ± 1 Tage nach Beginn der Verabreichung) und 1 Monat (entspricht: 45 ± 3 Tage nach Beginn der Verabreichung) nach Ende der Verabreichung des Studienmedikaments Nachsorgeuntersuchungen statt.
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	Das ursprüngliche Protokoll in der **Version 1.0 vom 18. Oktober 2016** wurde 8-mal geändert. Die erste Teilnehmerin unterzeichnete das Einwilligungsformular (Informed Consent Form; ICF) am 04. Januar 2017. Im Folgenden werden die relevanten Änderungen jeder Protokollversion zusammengefasst. **Protokolländerung 1, Version 2.0, 13. Dezember 2016** - Der Zeitpunkt der primären, sekundären und explorativen Endpunkte wurde von Tag 14 auf Tag 15 verschoben. - Ein Kriterium zum Ausschluss für die gleichzeitige Anwendung starker CYP3A4-Inhibitoren wurde hinzugefügt. - Ein ± 30-Minuten-Zeitfenster für die Einnahme des Studienmedikaments wurde eingeführt, mit der Anweisung, die Dosis auszulassen, wenn die Einnahme mehr als 60 Minuten außerhalb des geplanten Zeitpunkts liegen würde. **Protokolländerung 2, Version 3.0, 17. März 2017** - Teil B wurde hinzugefügt, mit der Verabreichung von Zuranolon-Kapseln (30 mg) einmal täglich abends in

Item[a]	Charakteristikum	Studieninformation
		einem vollständig ambulanten Setting. Teil A wurde geschlossen, sobald Zuranolon-Kapseln an den Studienzentren verfügbar waren. - Die Verwendung von Psychopharmaka zur Behandlung depressiver Symptome, wie Antidepressiva oder atypische Antipsychotika, wurde eingeschränkt, wenn deren Einnahme innerhalb von 30 Tagen vor der Einwilligung begonnen wurde und/oder nicht in stabiler Dosierung eingenommen wurden. - Die Einschränkung für den Beginn von Psychopharmaka wurde von „zwischen Screening und Abschluss der Tag-15-Bewertungen“ auf „innerhalb von 30 Tagen vor dem Screening bis zum Abschluss der Tag 15-Bewertungen“ erweitert. - Ein Ausschlusskriterium für Teilnehmerinnen mit positivem Urindrogentest beim Screening wurde hinzugefügt. - Die Nachbeobachtungsperiode wurde von 75 Tagen auf 45 Tage verkürzt. **Protokolländerung 3, Version 4.0, 06. Juni 2017** - Die Screening-Periode wurde von 7 auf 14 Tage verlängert. **Protokolländerung 4, Version 5.0, 31. August 2017** - Die Verabreichung des Studienmedikaments wurde von „in der Klinik oder zu Hause mit einem medizinischen Fachpersonal“ auf „Selbstverabreichung durch die Teilnehmer“ geändert, wobei die Medikamenteneinnahme durch einen abendlichen Telefonanruf überwacht wurde. Dies führte zu einer Reduzierung des Besuchsplans und der Häufigkeit der Bewertungen während der Behandlungsperiode. - Die Screening-Periode wurde von 14 auf 28 Tage verlängert, um die Terminplanung Die Teilnehmerin zu erleichtern und die Überprüfung der Screening-Daten vor der Randomisierung zu verbessern. - Die Stanford Sleepiness Scale (SSS) wurde als Sicherheits-/ Dosisanpassungsbewertung entfernt, und stattdessen wurden Anweisungen zur Dosisanpassung basierend auf der Verträglichkeit hinzugefügt. (Rationale: Sedierung/Somnolenz wird während der gesamten Studie als unerwünschtes Ereignis erfasst; daher ist die SSS nicht als separater Endpunkt zur Bewertung der Sicherheit erforderlich. - Der Beck Anxiety Inventory (BAI) wurde als Wirksamkeitsendpunkt entfernt. (Rationale: Der BAI wird nicht durchgeführt, um die Redundanz bei der Bewertung

Item[a]	Charakteristikum	Studieninformation
		von Angstsymptomen zu verringern [d. h. die HAM-A-Skala ist ausreichend].) - Die Pulsoximetrie wurde als Sicherheitsbewertung entfernt. (Rationale: Sedierung/Somnolenz wird während der gesamten Studie als unerwünschtes Ereignis erfasst; daher ist die Pulsoximetrie nicht als separater Endpunkt zur Bewertung der Sicherheit erforderlich.) - Eine Option zur Erhöhung der Stichprobengröße wurde hinzugefügt, falls nach Abschluss von 14 auswertbaren Teilnehmern am Tag 15 eine verblindete Schätzung der gemeinsamen Standardabweichung (basierend auf der mittleren Veränderung von der Baseline bis Tag 15 im HAMD-Gesamtwert) mehr als die angenommene SD (10 Punkte) beträgt. **Protokolländerung 5, Version 6.0, 01. Dezember 2017** - Die Stichprobengröße wurde von 32 auf 60 Teilnehmerinnen erhöht, um die Power zur Erkennung eines Unterschieds im primären Endpunkt zwischen Zuranolon und Placebo von 72 % auf 80 % zu steigern. - Die Option zur Erhöhung der Stichprobengröße durch Berechnung einer verblindeten Schätzung der gemeinsamen SD im primären Endpunkt wurde entfernt **Protokolländerung 6, Version 7.0, 15. Februar 2018** - Die Stichprobengröße wurde von 60 auf 140 Teilnehmerinnen erhöht, um die Power zur Erkennung eines Unterschieds im primären Endpunkt zwischen Zuranolon und Placebo von 80 % auf 90 % zu steigern. - Alle Details zu Teil A der Studie wurden entfernt, da dieser geschlossen wurde und zum Zeitpunkt der Änderung keine aktiven Teilnehmerinnen mehr hatte. - Die Einschränkung für den Beginn von Psychopharmaka wurde von „innerhalb von 30 Tagen vor dem Screening“ auf „innerhalb von 30 Tagen vor Tag 1“ reduziert. - Die Phase der Studie wurde von „Phase 2a“ auf „Phase 3“ aktualisiert. **Protokolländerung 7, Version 8.0, 14. Juni 2018** - Aktualisierung des Prozesses zur Erstellung des Randomisierungsschemas und der Verpackung des Studienmedikaments. - Ein Ausschlusskriterium für Teilnehmerinnen, die während der Studienzeit eine elektive Operation planen, wurde hinzugefügt. - Die Einschränkung des Konsums von Grapefruitsaft, Grapefruit, Sevilla-Orangen oder Johanniskraut wurde von „innerhalb von 30 Tagen vor Tag 1“ auf „innerhalb von 14 Tagen vor Tag 1“ reduziert.

Item[a]	Charakteristikum	Studieninformation
		- Die Einschränkung starker CYP3A4-Induktoren wurde auf jegliche CYP-Induktoren erweitert, wie Rifampin, Carbamazepin, Ritonavir, Enzalutamid, Efavirenz, Nevirapin, Phenytoin, Phenobarbital oder Johanniskraut, innerhalb von 14 Tagen oder 5 Halbwertszeiten (je nachdem, was länger ist) vor der ersten Dosis des Studienmedikaments und während der gesamten Studie. **Protokolländerung 8, Version 9.0, 02. Juli 2018** - Hauptzweck war die Korrektur der Nummerierung der Einschluss- und Ausschlusskriterien; es wurden keine weiteren Änderungen vorgenommen. **Änderungen im SAP:** Diese Änderungen wurden nur im SAP vorgenommen: Definition des Brustmilchsets zur Analyse der Konzentrationen in der Brustmilch: - Protokolländerung 8, Version 9.0 (Abschnitt 14.6.2.1) definiert die obere Grenze für ein TEAE als 7 Tage nach der letzten Dosis des Studienmedikaments, während Abschnitt 13.4 TEAEs als alle AE definiert, die nach der ersten Verabreichung des Studienmedikaments auftreten. Auf Anfrage von Sage definiert der SAP TEAEs als alle AE, die nach der ersten Verabreichung des Studienmedikaments auftreten. - Das Protokoll enthält keine MADRS-Response und Remission als sekundäre Endpunkte, diese wurden jedoch zur Analyse angefordert. - Subgruppenanalysen ausgewählter Wirksamkeitsendpunkte wurden von Sage angefordert. - Das Protokoll verwendet die "All Randomized Set" als Datensatz für die Analyse, um die Zusammenfassungen zur Patientinnenverteilung, Demografie und Ausgangscharakteristika zu analysieren. Auf Anfrage von Sage wurden jedoch die Patientinnenverteilung für alle eingeschriebenen Patientinnen einbezogen, und die Demografie sowie die Ausgangscharakteristika basierten auf den Patientinnen im "Safety Set" und "Efficacy Set". Nur ein Patient wurde in Teil A eingeschrieben und behandelt, bevor die Einschreibung geschlossen wurde. Daher wurden die in Teil A erhobenen Daten lediglich aufgelistet. - Änderungen an den ursprünglichen und nachfolgenden Schätzungen der Stichprobengröße: Jede Änderung, in der es eine Veränderung der gewünschten Studienpower und des Placebo-adjustierten Behandlungseffekts für diese Studie gab, wurde durch klinische Informationen außerhalb dieser Studie geleitet. Diese Informationen umfassten (1) Kommentare der Behörde während des

Item[a]	Charakteristikum	Studieninformation
		Meetings für Brexanolon, (2) Studienergebnisse aus dem Brexanolon-Programm (547-PPD-202B und 202C) und (3) Studienergebnisse aus dem Zuranolon-Programm für Major Depression (MDD) (Studie 217-MDD-201B). Die ursprüngliche Berechnung der Stichprobengröße verwendete ein Alpha-Niveau von 10 %, eine Power von 80 % mit einem Behandlungseffekt und einer SD von jeweils 10 Punkten (basierend auf beobachteten Daten aus der Studie 547-PPD-202A). Nach Gesprächen mit der Behörde bezüglich der Durchbruchtherapiebezeichnung für Brexanolon wurden Kommentare in die aktuelle Studie integriert, um ein Alpha von 5 % zu verwenden, was zu einer Verringerung der Power auf 72 % führte. Folglich wurde nach der Verfügbarkeit der Studienergebnisse aus der Studie 217-MDD-201 Teil B, die einen placebo-adjustierten Behandlungseffekt von 7,5 Punkten (kleiner als 10 Punkte) zeigten, die aktuelle Stichprobengröße erhöht, um die Studienpower auf dem vorgegebenen Niveau von 80 % zu halten. Schließlich nutzte Sage die Gelegenheit, die Studienpower auf 90 % zu erhöhen, basierend auf einem placebo-adjustierten Behandlungseffekt von 4 Punkten, SD = 7 Punkte (basierend auf gepoolten Daten der 90-mg-Dosis aus den Studien 547-PPD-202A, -202B und -202C). Diese Änderung wurde in der Protokolländerung 6, Version 7.0 des Protokolls und des SAP beschrieben. - Sensitivitätsanalysen wurden hinzugefügt, falls mehr als 10 % der Teilnehmerinnen HAMD-Werte am Tag 15 fehlen. - Ein PP-Set wurde für Sensitivitätsanalysen des primären Endpunkts hinzugefügt, falls erforderlich. - Es wurde festgelegt, dass die PK-Analyse außerhalb des SAP abgeschlossen und außerhalb des klinischen Studienberichts berichtet wird. Zusätzlich besagt das Protokoll, dass das All Randomized Set verwendet wird, um die Verteilung Die Teilnehmerin, demografische Daten und Basismerkmale zu analysieren; jedoch wurde die Verteilung Die Teilnehmerin auf alle eingeschlossenen Teilnehmerinnen ausgeweitet, und die demografischen Daten sowie die Basismerkmale wurden unter Verwendung des Safety Sets und des Efficacy Sets analysiert.
4	Probanden / Patienten	
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	**Einschlusskriterien:** 1. Die Probandin hat vor der Durchführung von studienspezifischen Verfahren das ICF unterschrieben. 2. Die Probandin ist eine nicht-stationär behandelte Frau im Alter von 18 bis 45 Jahren, einschließlich. 3. Die Probandin ist in gutem körperlichem Gesundheitszustand und hat keine klinisch signifikanten

Item[a]	Charakteristikum	Studieninformation
		Befunde, wie vom Prüfarzt bei der körperlichen Untersuchung, dem 12-Kanal-EKG oder klinischen Laboruntersuchungen festgestellt. 4. Die Probandin stimmt zu, die Studienanforderungen einzuhalten. 5. Die Probandin muss entweder die Sekretion von Milch zum Zeitpunkt des Screenings eingestellt haben oder, falls sie noch in der laktierenden Phase ist oder aktiv stillt, zustimmen, das Stillen für den Zeitraum von kurz vor der Einnahme des Studienmedikaments bis Tag 21 vorübergehend einzustellen, damit eine 7-tägige Auswaschphase nach der letzten Dosis des Studienmedikaments eingehalten wird. 6. Die Probandin muss einen negativen Schwangerschaftstest beim Screening und am Tag 1 vor Beginn der Verabreichung des Studienmedikaments haben. 7. Die Probandin hatte eine schwere depressive Episode, die nicht früher als im dritten Trimester und nicht später als in den ersten 4 Wochen nach der Entbindung begann, und erfüllt die Kriterien für eine schwere depressive Episode gemäß DSM-5, diagnostiziert durch ein DSM-5 SCID-I. 8. Die Probandin hat einen HAMD-Gesamtwert von ≥ 26 beim Screening und am Tag 1 (vor der Randomisierung). 9. Die Probandin ist ≤ 6 Monate nach der Entbindung. 10. Die Probandin ist gewillt den Beginn anderer Antidepressiva oder Angstmedikamente sowie neuer Pharmakotherapie-Regime, einschließlich bedarfsorientierter Benzodiazepin-Angstlöser, bis nach dem Ende der Behandlungsperiode und dem Abschluss aller Bewertungen am Tag 15 zu verschieben. 11. Die Probandin hat kein nachweisbares Hepatitis-B-Oberflächenantigen (HBsAg), keine nachweisbaren Antikörper gegen das Hepatitis-C-Virus (HCV), nachweisbare Anti-HCV-Antikörper, aber eine negative Viruslast und keine nachweisbaren Antikörper gegen das humane Immundefizienz Viren (HIV) beim Screening. 12. Die Probandin stimmt zu, während der Teilnahme an der Studie und für 30 Tage nach der letzten Dosis des Studienmedikaments eine der folgenden Methoden zur Empfängnisverhütung zu verwenden, es sei denn, sie sind chirurgisch steril: • Kombinierte (östrogen- und gestagenhaltige) orale, intravaginale oder transdermale hormonelle Verhütung, die mit der Hemmung des Eisprungs verbunden ist. • Orale, injizierbare oder implantierbare gestagenhaltige hormonelle Verhütung, die mit der Hemmung des Eisprungs verbunden ist.

Item[a]	Charakteristikum	Studieninformation
		• Intrauterines Gerät. • Intrauterines hormonfreisetzendes System. • Bilaterale Tubenokklusion. • Partner mit Vasektomie. **Ausschlusskriterien:** Probandinnen werden ausgeschlossen, wenn sie eines der folgenden Ausschlusskriterien erfüllen: 1. Die Probandin hat eine aktuelle Anamnese mit aktiven klinisch signifikanten Manifestationen von metabolischen, hepatischen, renalen, hämatologischen, pulmonalen, kardiovaskulären, gastrointestinalen, muskuloskelettalen, dermatologischen, urogenitalen, neurologischen oder Augen-, Ohren-, Nase- und Rachenstörungen oder eine andere akute oder chronische Erkrankung, die nach Meinung des Prüfarztes die Fähigkeit der Probandin, an dieser klinischen Studie teilzunehmen oder diese abzuschließen, einschränken würde. 2. Die Probandin hat eine bekannte Allergie gegen die Zuranolon-Kapsel oder deren Hilfsstoffe. 3. Die Probandin hat aktive Psychosen gemäß der Beurteilung des Prüfarztes. 4. Die Probandin hat einen Suizidversuch im Zusammenhang mit der aktuellen Episode der PPD unternommen. 5. Die Probandin hat eine medizinische Historie von Anfällen. 6. Die Probandin hat eine medizinische Historie von bipolaren Störungen, Schizophrenie und/oder schizoaffektiven Störungen. 7. Die Probandin hat eine Anamnese mit aktiver Alkoholabhängigkeit oder Drogenabhängigkeit (einschließlich Benzodiazepinen) in den 12 Monaten vor dem Screening. 8. Die Probandin hatte innerhalb von 30 Tagen vor dem Screening eine Exposition gegenüber einem anderen Prüfpräparat oder Gerät. 9. Die Probandin hat zuvor an einer klinischen Studie zu SAGE-547 oder Zuranolon teilgenommen. 10. Probandinnen, die sich für die Studie vorstellen und derzeit Psychopharmaka erhalten, die zur Behandlung depressiver Symptome eingesetzt werden, wie Antidepressiva, atypische Antipsychotika usw., die seit mindestens 30 Tagen vor Tag 1 nicht mehr in derselben Dosis eingenommen wurden. (Probandinnen, die sich für die Studie anmelden und diese Medikamente innerhalb von 30 Tagen vor Beginn der Einnahme des Studienmedikaments abgesetzt haben, können zugelassen werden, wenn sie die

Item[a]	Charakteristikum	Studieninformation
		Medikamente länger als 5 Halbwertszeiten bis zum Beginn der Einnahme des Studienmedikaments abgesetzt haben). 11. Verwendung von bekannten starken Inhibitoren des Cytochrom P450 (CYP)3A4 innerhalb von 14 Tagen oder 5 Halbwertszeiten (je nachdem, was länger ist) oder Konsum von Grapefruitsaft, Grapefruit, Sevilla-Orangen oder Produkten, die diese enthalten, innerhalb von 14 Tagen vor der ersten Dosis des Studienmedikaments und während der gesamten Studie. 12. Die Probandin hat einen positiven Urindrogentest beim Screening. 13. Verwendung von CYP-Induktoren, wie Rifampin, Carbamazepin, Ritonavir, Enzalutamid, Efavirenz, Nevirapin, Phenytoin, Phenobarbital oder Johanniskraut, innerhalb von 14 Tagen oder 5 Halbwertszeiten (je nachdem, was länger ist) vor der ersten Dosis des Studienmedikaments und während der gesamten Studie. 14. Die Probandin plant, während der Teilnahme an der Studie eine elektive Operation durchzuführen.
4b	Studienorganisation und Ort der Studiendurchführung	**Studienorganisation:** Sage Therapeutics, Inc. **Ort der Studiendurchführung:** Die Studie wurde an 33 Zentren in den Vereinigten Staaten durchgeführt. Acht weitere Zentren in den Vereinigten Staaten wurden aktiviert, haben jedoch keine Teilnehmerinnen gescreent. Vier Studienzentren hatten während der Studie einen Wechsel des Hauptprüfers.
5	Interventionen Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	**Part A** Da nur eine Patientin in Part A der Studie teilgenommen hat und Sage-217 erhielt, werden die Informationen zur Intervention hier nur kurz zusammengefasst. Zuranolon-oral Lösung war als 6 mg/ml Lagerlösung in wässriger Form erhältlich, die 40 % 2-Hydroxypropyl-β-Cyclodextrin und 0,0025 % Sucralose enthielt. Diese Lösung wurde weiter mit sterilem Wasser zur Injektion verdünnt, um die ausgewählten Dosen zu erreichen. Die 6 mg/ml Lagerlösung wurde aus Zuranolon-Wirkstoffpulver und Hilfsstoffen hergestellt und anschließend am klinischen Standort zur Dosisvorbereitung gemischt. Die klinische Dosis wurde durch Abwiegen der entsprechenden Menge der Lagerlösung und Verdünnen auf ein Volumen von 40 ml mit Wasser zubereitet. Verfügbare Dosen umfassten 20 mg und 15 mg. Die Chargennummer des Studienmedikaments für die Zuranolon-oral Lösung war 16-0027. Die Placebo-oral Lösung wurde keiner Teilnehmerinnen verabreicht. **Part B:** Am Tag 1 der Behandlungsperiode wurden die geeigneten Teilnehmerinnen im Verhältnis 1 : 1 auf Zuranolon 30 mg einmal

Item[a]	Charakteristikum	Studieninformation
		täglich (QD) oder Placebo randomisiert. Das Studienmedikament wurde jeden Abend um etwa 20 Uhr verabreicht. Gemäß Protokolländerung 2, Version 3.0, und Protokolländerung 3, Version 4.0, wurde das Studienmedikament unter Aufsicht eines medizinischen Fachpersonals entweder in der Klinik oder zu Hause bei den Teilnehmerinnen verabreicht. Gemäß Protokolländerung 4, Version 5.0, und danach wurde das Studienmedikament von den Teilnehmerinnen selbst verabreicht, und die Einhaltung wurde durch einen Follow-up-Anruf vom Studienzentrum jeden Abend überwacht. Während der Behandlungsperiode konnten Teilnehmerinnen in Teil B, die 30 mg QD nicht vertrugen, für den Rest der Behandlungsperiode 20 mg QD erhalten. Teilnehmerinnen, die UE auf der Dosisstufe von 20 mg QD erlebten, sollten nach Ermessen des Prüfarztes aus der Studienbehandlung ausgeschlossen werden. Zuranolon-Kapseln waren als harte Gelatinekapseln erhältlich, die ein weißes bis cremefarbenes Pulver enthielten. Neben dem Wirkstoff Zuranolon enthalten die Zuranolon-Kapseln Croscarmellose-Natrium, Mannit, silicifiziertes mikrokristalline Cellulose und Natriumstearylfumarat und als Hilfsstoffe. Die Kapseln waren in Stärken von 10 mg, 20 mg und 30 mg erhältlich, um Behandlungsdosen von 20 mg und 30 mg bereitzustellen. Die entsprechenden Placebo-Kapseln waren harte Gelatinekapseln, die nur die oben genannten Hilfsstoffe für die aktive Kapselbehandlung enthielten. Die Teilnehmerinnen erhielten 2 Kapseln pro Dosis (Zuranolon oder Placebo). Vor der Protokolländerung 7, Version 8.0, wurde die Dosis von 30 mg Zuranolon als 1 x 20 mg und 1 x 10 mg Kapseln verabreicht; mit der Änderung 7 und danach (wie in Abschnitt 9.8.1 des klinischen Studienberichts beschrieben) wurde das Studienmedikament als 1 x 30 mg und 1 x Placebo-Kapseln verabreicht. Die Dosis von 20 mg wurde in allen Protokollversionen als 2 x 10 mg Kapseln verabreicht. Die Chargennummern für Zuranolon-Kapseln waren 17-0003 (10 mg), 17-0084 (20 mg), 17-0004 (20 mg) und 17-0085 (30 mg).
6	Zielkriterien	
6a	Klar definierte primäre und sekundäre Zielkriterien, Erhebungszeitpunkte, ggf. alle zur Optimierung der Ergebnisqualität verwendeten Erhebungsmethoden (z. B.	**Primärer Endpunkt:** Das primäre Ziel der Studie ist es, zu bestimmen, ob die Behandlung mit Zuranolon depressive Symptome bei Patientinnen mit schwerer PPD im Vergleich zu Placebo reduziert. Dies wird anhand der Veränderung des HAMD-Gesamtwerts vom Baselinewert bis Tag 15 bewertet. **Sekundäre Endpunkte**

Item[a]	Charakteristikum	Studieninformation
	Mehrfachbeobachtungen, Training der Prüfer) und ggf. Angaben zur Validierung von Erhebungsinstrumenten	- Bewertung der Reduktion depressiver Symptome bzw. Angstsymptome durch Zuranolon im Vergleich zu Placebo anhand: • Der Veränderung des HAMD-Gesamtwerts zu allen anderen Zeitpunkten gegenüber dem Score zu Baseline. • Dem HAMD-Ansprechen, definiert über eine ≥ 50 %-ige Reduktion des HAMD-Gesamtwerts zu Baseline. • HAMD-Remission, definiert als ein HAMD-Gesamtwert ≤ 7. • Veränderung von der Baseline bis Tag 15 im MADRS-Gesamtwert. • Die Veränderung des CGI-S-Gesamtwerts zu Tag 15 und anderen Zeitpunkten gegenüber der Baseline. • Die Veränderung des MADRS-Gesamtwerts und den individuellen Item Scores zu Tag 15 und zu anderen Zeitpunkten gegenüber der Baseline • CGI-I-Response, definiert als „sehr stark verbessert“ oder „stark verbessert“ gemäß Fragebogen. • Veränderung des HAM-A-Gesamtwerts zu Tag 15 und zu anderen Zeitpunkten gegenüber der Baseline. - Bewertung der Sicherheit und Verträglichkeit von Zuranolon im Vergleich zu Placebo durch: • die Häufigkeit und Schwere von UE. • Beobachtete Werte und Veränderungen gegenüber der Baseline der Vitalzeichen, Laborwerten, EKG-Parametern und der C-SSRS. **Explorative Endpunkte:** - HCRU Daten zu Baseline und Tag 45. - Veränderung des EPDS-Gesamtwerts zu Tag 15 und anderen Zeitpunkten gegenüber der Baseline. - Veränderung des PHQ-9-Gesamtwerts zu Tag 15 und anderen Zeitpunkten gegenüber der Baseline. - Veränderung des BIMF-Gesamtwerts und dessen Subskalen zu Tag 15 und anderen Zeitpunkten gegenüber der Baseline. - Veränderung des SF-36 Subskala und Komponenten-Gesamtwerts zu Tag 15 und anderen Zeitpunkten gegenüber der Baseline.
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	- Die SSS wurde als Sicherheits-/Dosisanpassungsbewertung entfernt, und stattdessen wurden Anweisungen zur Dosisanpassung basierend auf der Verträglichkeit hinzugefügt. (Rationale: Sedierung/Somnolenz wird während der gesamten Studie als unerwünschtes Ereignis erfasst; daher ist die SSS nicht als separater Endpunkt zur Bewertung der Sicherheit erforderlich.)

Item[a]	Charakteristikum	Studieninformation
		- Der BAI wurde als Wirksamkeitsendpunkt entfernt. (Rationale: Der BAI wird nicht durchgeführt, um die Redundanz bei der Bewertung von Angstsymptomen zu verringern [d.h. die HAM-A-Skala ist ausreichend].) - Die Pulsoximetrie wurde als Sicherheitsbewertung entfernt. (Rationale: Sedierung/Somnolenz wird während der gesamten Studie als unerwünschtes Ereignis erfasst; daher ist die Pulsoximetrie nicht als separater Endpunkt zur Bewertung der Sicherheit erforderlich.) - Es wurde festgelegt, dass die PK-Analyse außerhalb des SAP abgeschlossen und außerhalb des klinischen Studienberichts berichtet wird.
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	Unter der Annahme eines zweiseitigen Tests mit einem Alpha-Niveau von 0,05 würde eine Stichprobengröße von etwa 65 auswertbaren Teilnehmern pro Behandlungsgruppe eine Power von 90 % bieten, um einen placebo-adjustierten Behandlungseffekt von etwa 4 Punkten im primären Endpunkt, der Veränderung im HAMD-Gesamtwert am Tag 15 gegenüber der Baseline, bei einer angenommenen SD von 7 Punkten zu erkennen. Unter der Annahme einer Abbruchrate von 10 % und eines Randomisierungsverhältnisses von 1 : 1 werden etwa 72 randomisierte Teilnehmerinnen pro Behandlungsgruppe benötigt, um 130 auswertbare Teilnehmerinnen zu erhalten. Auswertbare Teilnehmerinnen sind definiert als die randomisierten Teilnehmer, die das Studienmedikament erhalten haben und einen gültigen Ausgangswert sowie mindestens eine Nachuntersuchung des HAMD-17 haben. Zusätzliche Teilnehmerinnen könnten randomisiert werden, wenn die Abbruchrate höher als 10 % ist.
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	**Zwischenanalysen** Es wurden keine Interimsanalysen geplant oder durchgeführt. **Kriterien für einen vorzeitigen Studienabbruch** Sage Therapeutics kann diese Studie oder einen Teil der Studie jederzeit aus Sicherheitsgründen beenden, einschließlich des Auftretens von UE oder anderer Befunde, die ein inakzeptables Risiko für die Teilnehmerinnen nahelegen, oder aus administrativen Gründen. Im Falle einer Studienbeendigung wird Sage Therapeutics dem Prüfarzt eine schriftliche Benachrichtigung zukommen lassen. Die Prüfzentren müssen umgehend ihre Ethikkommission (Institutional Review Board; IRB) benachrichtigen und die Withdrawal-Verfahren für die teilnehmenden Patientinnen einleiten.
8	Randomisierung, Erzeugung der Behandlungsfolge	

Item[a]	Charakteristikum	Studieninformation
8a	Methode zur Generierung der zufälligen Zuteilung	Teilnehmerinnen, die die Einschlusskriterien erfüllen, werden am Tag 1 randomisiert im Verhältnis 1 : 1 einer Behandlung mit Zuranolon oder Placebo zugewiesen. Den Teilnehmerinnen wurde nacheinander eine Randomisierungsnummer aus dem Randomisierungsschema zugewiesen, das dem entblindeten pharmazeutischen Personal zur Verfügung gestellt wird (vor Protokollversion 8.0) oder die Randomisierung wurde zentral über eine Interactive Response Technology (IRT) durchgeführt (Protokollversion 8.0 und danach).
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	Teilnehmerinnen, die die Einschlusskriterien erfüllen, werden gemäß einem Randomisierungsschema im Verhältnis 1 : 1 zufällig Zuranolon oder Placebo zugewiesen. Sobald festgestellt wurde, dass eine Teilnehmerin die Eignungskriterien erfüllt, wird ihr eine zufällige Nummer aus dem Randomisierungsschema zugewiesen, die dem entblindeten pharmazeutischen Personal zur Verfügung gestellt wird (vor Protokoll 6.0) oder über ein IRT (Protokolländerung 6.0 und danach). Die Randomisierungsnummern bestehen aus einer 4-stelligen Zahl, die mit einer 3 beginnt (z. B. „3001“) für die papierbasierte (manuelle) Randomisierung oder mit einer 4 beginnt (z. B. „4001“) für die IRT-Randomisierung, die für Teilnehmerinnen erfolgt, die unter Änderung 6.0 oder danach eingeschrieben wurden. Die Randomisierung der Teilnehmerinnen (einschließlich Studienabschnitt, Teilnehmeridentifikator, Randomisierungsnummer und zugewiesene Behandlung) wird in einer Liste bereitgestellt.
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	Dies war eine doppelblinde Studie. Teilnehmerinnen, die die Einschlusskriterien erfüllten, wurden im Verhältnis 1 : 1 zufällig Zuranolon oder Placebo zugewiesen. Teilnehmerinnen, Kliniker*innen und das Studienteam waren über die Zuteilung zur Behandlung verblindet. Vor der Protokolländerung 7, Version 8.0, waren das verantwortliche Apothekenpersonal, das für die Abgabe des Studienmedikaments verantwortlich war, über das Randomisierungsschema informiert. Unter Protokolländerung 7, Version 8.0 und danach, wurde die Randomisierung zentral über ein IRT-System durchgeführt, und alle Mitarbeiter vor Ort, einschließlich des Apothekenpersonals, waren verblindet; das entblindete Apothekenpersonal, das vor Änderung 7 für die Abgabe des Studienmedikaments verantwortlich war, waren nach der Einführung des IRT-Systems nicht mehr an der Durchführung der Studie beteiligt. Elektronische Kopien des Randomisierungsschemas wurden dem Labor, das die bioanalytischen Analysen durchführte, zur Verfügung gestellt, um eine begrenzte Analyse von Proben von Teilnehmerinnen, die Placebo erhielten, zu ermöglichen.

Item[a]	Charakteristikum	Studieninformation
		Im Falle eines medizinischen Notfalls darf das verantwortliche Apothekenpersonal den tatsächlichen Inhalt des Studienmedikaments dem Prüfarzt/Prüfärztin offenlegen, die auch Sage über den Notfall informieren sollten. In allen Fällen, in denen die Zuteilung des Studienmedikaments für eine Teilnehmerin entblindet wird, müssen relevante Informationen (einschließlich des Grundes für die Entblindung) in den Unterlagen der Teilnehmerin und im elektronischen Fallberichtformular (eCRF) dokumentiert werden. Wenn die Teilnehmerin oder das Personal des Studienzentrums (außer das verantwortliche Apothekenpersonal) entblindet wurde, wird die Teilnehmerin aus der Studie ausgeschlossen. Es trat kein medizinischer Notfall ein.
10	Randomisierung, Durchführung Wer hat die Randomisierungsliste erstellt, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	Das Schema zur Randomisierung wurde von einem unabhängigen Statistiker erstellt. Die Zuteilung zur Behandlungsgruppe (Zuranolon oder Placebo) basiert auf einem Randomisierungsschema. Die Randomisierungspläne wurden streng vertraulich behandelt und waren bis zum Zeitpunkt der Entblindung nur autorisiertem Personal zugänglich.
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?	Probandinnen (a, b), Kliniker*innen (b, c) und das Studienteam (b, c) waren über die Zuteilung zur Behandlung verblindet. Vor der Protokolländerung 7, Version 8.0, war das verantwortliche Apothekenpersonal, das für die Abgabe des Studienmedikaments verantwortlich war, über das Randomisierungsschema informiert. Unter Protokolländerung 7, Version 8.0 und danach, wurde die Randomisierung zentral über ein IRT-System durchgeführt, und alle Mitarbeitenden vor Ort, einschließlich des Apothekenpersonals, waren verblindet; das entblindete verantwortliche Apothekenpersonal, das vor Änderung 7 für die Abgabe des Studienmedikaments verantwortlich war, waren nach der Einführung des IRT-Systems nicht mehr an der Durchführung der Studie beteiligt. Elektronische Kopien des Randomisierungsschemas wurden dem Labor, das die bioanalytischen Analysen durchführte, zur Verfügung gestellt, um eine begrenzte Analyse von Proben von Teilnehmern, die Placebo erhielten, zu ermöglichen.
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Zuranolon-Kapseln waren als harte Gelatinekapseln (inklusive Hilfsstoffe) erhältlich, die ein weißes bis cremefarbenes Pulver enthielten. Die passenden Placebo-Kapseln waren harte Gelatinekapseln, die nur die entsprechenden Hilfsstoffe enthielten
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	Die Daten für den 1 Teilnehmerinnen in Teil A sind aufgeführt; statistische Analysen wurden nur für Teil B durchgeführt. Das Safety Set besteht aus allen Teilnehmerinnen in Teil B, die das Studienmedikament erhalten haben. Das Safety Set wurde

Item[a]	Charakteristikum	Studieninformation
		verwendet, um beschreibende Zusammenfassungen der Sicherheitsdaten gemäß der erhaltenen Behandlung bereitzustellen. Wenn eine Teilnehmerin während der Studie eine Dosis von Zuranolon eingenommen hat, wurde sie im Zuranolon-Arm für die erhaltene Behandlung betrachtet. Historische Daten (z. B. Demografie und Basismerkmale, vorherige Medikation, psychiatrische Vorgeschichte) wurden gemäß der erhaltenen Behandlung (Zuranolon oder Placebo) zusammengefasst. Das Efficacy Set bestand aus allen Teilnehmerinnen in Teil B, die das Studienmedikament erhalten haben und einen gültigen Ausgangswert sowie mindestens eine Nachuntersuchung zur Wirksamkeit aufweisen konnten. Das Efficacy Set wurde verwendet, um alle Daten zur Wirksamkeit, Demografie und zu den Basismerkmalen gemäß der randomisierten Behandlungsgruppe zu analysieren. Das Per Protocol (PP) Set umfasst Teilnehmerinnen im Efficacy Set ohne signifikante Protokollverletzungen oder Abweichungen, die die Wirksamkeit beeinflussen könnten. Diese Überprüfung der wesentlichen Wirksamkeitsprotokollabweichungen wurde abgeschlossen, und die Entscheidung, ob die Abweichung die Wirksamkeit beeinflusst, wurde vor der Entblindung der Datenbank dokumentiert. Teilnehmerinnen wurden gemäß der randomisierten Behandlung klassifiziert. Dieses Analyse-Set sollte für Sensitivitätsanalysen des primären Endpunkts verwendet werden, wenn mindestens 5 % der Teilnehmerinnen signifikante Protokollabweichungen aufwiesen, die als wirksamkeitsbeeinflussend bestimmt wurden, was jedoch nicht eintrat. Das PK Set besteht aus allen Teilnehmern im Safety Set mit ausreichenden Plasmaproben mit quantifizierbaren Konzentrationen von Zuranolon. Das PK Set wurde verwendet, um die Konzentrationen von Zuranolon für Teilnehmer, die nur Zuranolon erhielten, zusammenzufassen. Das Breast Milk Set besteht aus allen Teilnehmern im Safety Set, die eine optionale Einwilligung zur Sammlung von Brustmilch gegeben haben und die ausreichende Brustmilchproben mit quantifizierbaren Konzentrationen von Zuranolon bereitgestellt haben. Das Breast Milk Set wurde verwendet, um die Konzentrationen von Zuranolon für Teilnehmer, die nur Zuranolon erhielten, zusammenzufassen. Für den primären Endpunkt, die Änderung vom Ausgangswert bis Tag 15 im HAMD-Gesamtwert, wurde der Unterschied zwischen den Behandlungsgruppen durch ein MMRM bewertet, mit Behandlung, Ausgangswert des HAMD-Gesamtwerts, vorheriger Antidepressiva-Nutzung, Bewertungszeitpunkt und der Interaktion zwischen Zeitpunkten und Behandlung als feste Effekte. Beobachtete Werte und Änderungen/prozentuale Änderungen vom Ausgangswert im HAMD-Gesamtwert und den Einzelwerten werden nach Behandlungsgruppe zusammengefasst.

Item[a]	Charakteristikum	Studieninformation
		Beim Vergleich des Besuchswerts mit dem Ausgangswert innerhalb jeder Behandlungsgruppe wird die Veränderung des LS-MW vom Ausgangswert zusammen mit dem zugehörigen 95%-KI dargestellt. Beim Vergleich der Behandlungen (Zuranolon minus Placebo) werden die LS-MD in der Änderung vom Ausgangswert zusammen mit dem zugehörigen 95%-KI und dem paarweisen p-Wert für die Behandlung dargestellt. Eine unstrukturierte (UN) Kovarianzstruktur wurde verwendet, um die Fehler innerhalb der Teilnehmerinnen zu modellieren. Die Kenward-Roger-Näherung wurde verwendet, um die Freiheitsgrade des Nenners zu schätzen. Wenn es ein Konvergenzproblem mit dem UN-Kovarianzmodell gab, wurden die Kovarianzstrukturen Toeplitz, „compound symmetry“ oder „Autoregressive (order 1)“ (AR[1]) verwendet, wobei diese Reihenfolge bis zur Erreichung der Konvergenz beibehalten wurde. Wenn die Kovarianzstruktur nicht UN war, wurde der Sandwich-Schätzer für die Varianz-Kovarianz-Matrix abgeleitet, wobei die EMPIRICAL-Option in der PROC MIXED-Anweisung in SAS verwendet wurde. Für sekundäre Analysen wurden die Änderungen vom Ausgangswert im HAMD-17 zu allen anderen Zeitpunkten, MADRS und HAM-A-Gesamtwerts mit MMRM-Methoden ähnlich dem primären Endpunkt analysiert. HAMD-Response, HAMD-Remission und CGI-I-Response wurden unter Verwendung einer verallgemeinerten Schätzgleichung (GEE; Generalized Estimating Equation) für binäre Antworten analysiert. Die Modelle für HAMD-Response und Remission umfassen Begriffe für Behandlung, Ausgangswert des HAMD-Gesamtwerts, vorherige Antidepressiva-Nutzung, Bewertungszeitpunkt und die Interaktion zwischen Zeitpunkten und Behandlung. Das Modell für die CGI-I-Response umfasst den Ausgangswert des CGI-S. Beim Vergleich der Behandlungen (Zuranolon minus Placebo) werden modellbasierte Punkteschätzungen (d. h. angepasste Odds Ratios) zusammen mit dem zugehörigen 95%-KI und dem paarweisen p-Wert für die Behandlung dargestellt. Eine UN-Kovarianzstruktur wurde verwendet, um die subjektinternen Fehler zu modellieren. Darüber hinaus wurde als unterstützende Analyse ein logistisches Regressionsmodell separat für jeden Besuch verwendet, das die oben genannten Antwortvariablen als abhängige Variable und die anderen Begriffe als erklärende Variablen umfasst. Die Firth penalized likelihood-Methode wurde verwendet, um Verzerrungen in den Parameterschätzungen zu reduzieren und um Nichtkonvergenz aufgrund von quasi-Trennung zu vermeiden. Andere Wirksamkeitsendpunkte, einschließlich EPDS, PHQ-9, BIMF und SF-36, wurden mit MMRM-Methoden ähnlich dem primären Endpunkt analysiert. Zusammenfassungen der HCRU werden nach Behandlung präsentiert. Unerwünschte Ereignisse wurden unter Verwendung des MedDRA Codierungssystems (v 19.1) kodiert. Die Inzidenz aller UE, UE nach maximalem Schweregrad und UE nach größtem Zusammenhang mit dem Studienmedikament wird nach MedDRA SOC, PT und

Item[a]	Charakteristikum	Studieninformation
		erhaltener Behandlung zusammengefasst. Darüber hinaus wird die Inzidenz aller UE nach PT und erhaltener Behandlung zusammengefasst. Beobachtete Werte und Änderungen vom Ausgangswert in den Vitalzeichen, klinischen Laborwerten, EKG-Parametern und suizidalen Gedanken unter Verwendung des C-SSRS wurden ebenfalls nach der erhaltenen Behandlung zusammengefasst.
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	Die MMRM-Analyse für die Änderung vom Ausgangswert im HAMD-Gesamtwert wird für die folgenden Untergruppen wiederholt: • Abstammung: Weiß, Schwarz oder Afroamerikanisch,, und Andere. • Alterskategorie: 18 – 24; 25 – 45. • Verwendung von Antidepressiva zu Beginn: Ja, Nein. • BMI-Kategorie: ≤ 18,4; 18,5 – 24,9; 25 – 29,9, ≥ 30 kg/m² • Beginn der PPD: 3. Trimester, innerhalb von 4 Wochen nach der Entbindung. • PPD in der Familienanamnese: Ja; Nein. Im MMRM-Modell für die Subgruppenanalyse wird die Verwendung von Antidepressiva aufgrund der geringen Anzahl von Teilnehmerinnen in der Antidepressiva-Gruppe in jeder Untergruppe aus dem Modell ausgeschlossen. Beobachtete Werte und Änderungen/prozentuale Änderungen vom Ausgangswert bis zur morgendlichen Bewertung am Tag 15 (Tag 15 – Ausgangswert und alle anderen Zeitpunkte) im HAMD-Gesamtwert werden nach Behandlungsgruppe und Subgruppe zusammengefasst. Darüber hinaus wird die LS-Mitteländerung vom Ausgangswert (und das 95%-KI des Mittelwerts) im HAMD-Gesamtwert am Tag 15 basierend auf den MMRM-Ergebnissen nach Behandlungsgruppe und Subgruppe dargestellt. Außerdem sind für Subgruppenanalysen (Kategorie: Verwendung von Antidepressiva zu Beginn: Ja, Nein) für folgende Analysen *a priori* im SAP definiert: - MMRM-Analyse für die Veränderung des MADRS-Gesamtwerts gegenüber der Baseline - GEE-Logisic Regression-Analyse des CGI-I-Ansprechens - MMRM-Analyse für die Veränderung des HAM-A-Gesamtwerts gegenüber Baseline
Resultate		

Item[a]	Charakteristikum	Studieninformation
13	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	In Part A wurde nur eine Probandin gescreent, die dann auch ihre Einverständniserklärung abgegeben hat, randomisiert wurde, die Studienmedikation erhielt und die Studie abgeschlossen hat. Sie blieb während der 14-tägigen Behandlungsphase stationär. In Part B, haben 275 Probandinnen ihre Einverständniserklärung abgegeben, davon wurden 153 (56,0 %) randomisiert. Insgesamt erhielten 151 Patientinnen mindestens eine Dosis der Studienmedikation.
13a	Anzahl der Studien Teilnehmerinnen 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	**Part A** a) Im Part A der ROBIN-Studie wurde eine Teilnehmerin in den Sage-217-Arm randomisiert. b) Diese eine Teilnehmerin erhielt Zuranolon. c) In der Analyse des primären Zielkriteriums wird die Teilnehmerin aus Part A nicht berücksichtigt. **Part B** a) Im Part B der ROBIN-Studie wurden 153 Teilnehmerinnen randomisiert, wobei 76 Teilnehmerinnen der Placebo-Gruppe und 77 Teilnehmerinnen der Sage-217-Gruppe zugeordnet wurden. b) 151 Teilnehmerinnen erhielten die Studienmedikation (73 Placebo; 78 Zuranolon) c) In der Analyse des primären Zielkriteriums wurden 150 Teilnehmerinnen, die dem Efficacy Set angehören, berücksichtigt.
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	**Part A** In Part A der Studie sind keine Patientinnen verloren oder ausgeschlossen worden. **Part B** In Part B wurden zwei Patientinnen randomisiert, die der Placebo-Gruppe bzw. der Sage-217-Gruppe zugeordnet waren, jedoch noch keine Studienmedikation erhielten. Ihre jeweiligen Gründe für den Abbruch der Studie waren Nichteinhaltung der Vorschriften bzw. der Rücktritt der Teilnehmerin.
14	Aufnahme / Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	Die erste Patientin wurde am 04.01.2017 in die Studie aufgenommen. Die letzte Patientin hat ihre letzte Visite am 11.12.2018 abgeschlossen. Die Nachbeobachtung erfolgte für 45 ± 3 Tage nach Beginn der Verabreichung des Studienmedikaments.
14b	Informationen, warum die Studie endete oder beendet wurde	Die Studie wurde wie geplant durchgeführt und analysiert (Studienbericht vom 19.09.2019).
a: nach CONSORT 2010.		

Stellen Sie für jede Studie den Patientenfluss in einem Flow-Chart gemäß CONSORT dar.

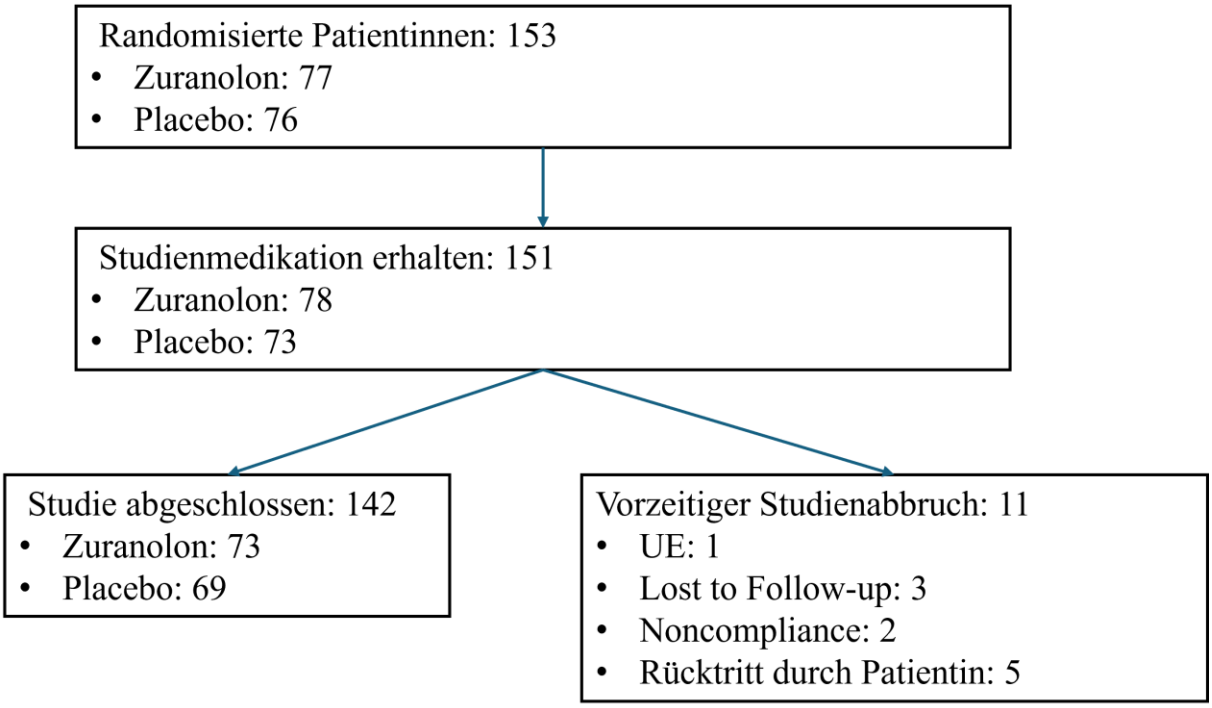


Abbildung 4-18: Patientenfluss der Studie ROBIN
Quelle: CSR Seite 65, Tabelle 8 [15]

Studie: SKYLARK

Tabelle 4-116 (Anhang): Studiendesign und -methodik für Studie SKYLARK

Item[a]	Charakteristikum	Studieninformation
Studienziel		
2 b	Genaue Ziele, Fragestellung und Hypothesen	**Primäres Ziel:** Festzustellen, ob die Behandlung mit Zuranolon die depressiven Symptome bei Erwachsenen mit PPD im Vergleich zu Placebo reduziert. **Sekundäre Ziele:** - Festzustellen, ob die Behandlung mit Zuranolon die Angstsymptome im Vergleich zu Placebo reduziert. - Die Selbstberichterstattung über depressive Symptome zu bewerten. - Die Sicherheit und Verträglichkeit von Zuranolon zu evaluieren. **Explorative Ziele:** - Die Plasma-Pharmakokinetik (PK) von Zuranolon zu bewerten.
Methoden		
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	Die SKYLARK-Studie wurde als randomisierte, doppel-blinde, placebo-kontrollierte, parallel-gruppierte Phase 3 Studie konzipiert, um die Wirksamkeit und Sicherheit von Zuranolon in Erwachsenen mit schwergradiger PPD zu evaluieren. Die Studie bestand aus einer Screening-Periode von bis zu 28 Tagen, einer 14-tägigen doppelblinden Behandlungsperiode und einer Follow-Up-Periode bis Tag 45. Während der Screening-Periode wurde die Diagnose Depression mittels dem DSM-5 SCID-5-CT gestellt. Anschließend wurden die Patientinnen in einem 1 : 1 Verhältnis in die Zuranolon-Gruppe oder Placebo-Gruppe randomisiert. Während der Behandlungsperiode nahmen die Patientinnen für 14 Tage jeden Abend ihre Studienmedikation ein, solange keine dosislimitierenden Bedenken bezüglich der Sicherheit oder Verträglichkeit bestanden. Teilnehmerinnen, die 50 mg nicht vertrugen, erhielten 40 mg für den Rest der Behandlungsperiode. Nach Ermessen des Prüfarztes wurden Teilnehmerinnen, die die Dosis von 40 mg nicht vertrugen, von der Verabreichung des Prüfpräparats ausgeschlossen. Während der Follow-Up Periode bis Tag 45, nach Beginn der Verabreichung des Studienmedikaments, wurden die Teilnehmerinnen weiterhin beobachtet.

Item[a]	Charakteristikum	Studieninformation
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	Das ursprüngliche Protokoll, datiert vom **01. April 2020**, hatte eine länderspezifische Änderung und wurde global einmal geändert. Die erste Teilnehmerin unterzeichnete das ICF am 08. Juni 2020. **Lokale Protokolländerung 1 (Version 1, UK), 18. September 2020** Der Hauptzweck der Protokolländerung 1 für das Vereinigte Königreich (UK), bestand darin, die von der britischen Arzneimittel- und Gesundheitsbehörde (Medicines and Healthcare Products Regulatory Agency; MHRA) bereitgestellten Kommentare zum Verfahren der Notfallentblindung und zu den Verfahren zur frühzeitigen Beendigung zu berücksichtigen; diese Änderung galt nur für Studien, die im Vereinigten Königreich durchgeführt wurden. Die Klarstellungen zum Protokoll sind unten aufgeführt: • Auf Anfrage der MHRA wurden alle Verweise auf die Durchführung des End-of-Treatment-Besuchs und des Early Termination-Besuchs am selben Tag im Falle einer frühzeitigen Beendigung entfernt. • Es wurde klargestellt, dass Teilnehmerinnen, die während der Behandlungsperiode die Teilnahme einstellen, alle Verfahren des End-of-Treatment-Besuchs durchlaufen sollten, während diejenigen, die die Teilnahme nach dem Ende der Behandlung einstellen, alle Verfahren des Early Termination-Besuchs durchlaufen sollten. • Die Verfahren zur Notfallentblindung wurden überarbeitet, um zu spezifizieren, dass Prüfer die Verblindung über das IRT-System ohne Rücksprache mit einem medizinischen Monitor oder Sage-Personal aufheben konnten. **Globale Protokolländerung 1 (Version 2), 29. Januar 2021** Der Hauptzweck der globalen Protokolländerung 1, bestand darin, zu ändern, wie Teilnehmerinnen, die die Studienmedikation abbrachen, in der Studie verfolgt werden sollten, sowie die beiden Protokollversionen (US [01. April 2020] und UK [18. September 2020]) in ein globales Protokoll zu integrieren. Weitere Änderungen wurden wie folgt umgesetzt: • COVID-19-Fragen wurden hinzugefügt, um Informationen zu Diagnose, Isolation und/oder Hospitalisierung aufgrund von COVID-19 im Rahmen der medizinischen Historie, der Erfassung UE und der Erfassung vorheriger/begleitender Medikamente/Verfahren während der gesamten Studie zu dokumentieren. • Die Eignungskriterien wurden erweitert, um Frauen einzuschließen, die bis zu 12 Monate nach der Entbindung waren (Änderung des Einschlusskriteriums #9). Das Kriterium für die Diagnose von PPD, einschließlich des Beginns der Symptome, blieb gemäß DSM-5 unverändert (Einschlusskriterium #7). Die Erweiterung auf 12 Monate

Item[a]	Charakteristikum	Studieninformation
		wurde eingeführt, um eine breitere Population von Teilnehmern zu erreichen, was mit DSM-5 übereinstimmt. • Es wurde klargestellt, dass ein Teilnehmerinnen mit einer Indexschwangerschaft, die zu einem neonatalen/infantilen Tod führte, ausgeschlossen wird (Änderung des Ausschlusskriteriums #6). • Details zum geschätzten Wert, der im Protokoll gemäß der Anfrage der FDA angegeben wurde, wurden hinzugefügt. • Die Definition einer Überdosierung wurde aus Abschnitt 12.4 entfernt, um mit der aktuellen Praxis von Sage und anderen Protokollen übereinzustimmen. Fälle von Überdosierung sollten gemäß den Angaben des Prüfers erfasst und als UE dokumentiert werden. • Die Anzahl der Standorte, an denen die Studie durchgeführt werden sollte, wurde erhöht. • Es wurde angegeben, dass Urin- und Serumschwangerschaftstests beim Screening durchgeführt werden sollten und dass der Urintest empfohlen wurde, um anderen Screening-Bewertungen vorauszugehen. Die Qualifikationskriterien sollten auf den Ergebnissen des Serumtests basieren. • Die Anforderungen an die Empfängnisverhütung wurden aktualisiert, um die Arten von bilateralen Tubenokklusionsverfahren zu spezifizieren, die als akzeptable Formen der Empfängnisverhütung angesehen wurden, und welche Verfahrensart mindestens 3 Monate nach dem Eingriff vor dem Screening erforderlich war. • Der Abschnitt über verbotene Medikamente wurde mit den als Ausschlusskriterien aufgeführten Medikamenten in Einklang gebracht. **Änderungen im SAP:** Die folgenden Punkte wurden im endgültigen SAP spezifiziert, waren jedoch nicht im klinischen Protokoll angegeben: • Es wurden Testsequenzen für wichtige sekundäre Endpunkte definiert, darunter die folgenden, einschließlich der folgenden (SAP Abschnitt 8.3.3.5): • Veränderung im HAMD-17-Gesamtwert am Tag 3 gegenüber Baseline. • Veränderung im HAMD-17-Gesamtwert am Tag 28 gegenüber Baseline. • Veränderung im HAMD-17-Gesamtwert am Tag 45 gegenüber Baseline. • Veränderung im CGI-S am Tag 15 gegenüber Baseline.

Item[a]	Charakteristikum	Studieninformation
		• Das Verfahren zur Anpassung der Multiplikation zur Kontrolle der Fehlerquote vom Typ I wurde im Detail beschrieben (SAP Abschnitt 8.3.3.5). • Die Definition des FAS wurde von „mindestens einer Nachuntersuchung des HAMD-17“ auf „alle randomisierten Teilnehmer, die die Studienmedikation mit gültigem Ausgangswert und mindestens 1 Nachuntersuchung des Gesamtwerts in mindestens einem der HAMD-17, HAM-A, MADRS, CGI-S, EPDS und PHQ-9 oder mindestens 1 Nachuntersuchungswert des CGI-I erhalten haben“ aktualisiert (SAP Abschnitt 7.2). • Die MADRS-Response und Remission wurden definiert und dem finalen SAP hinzugefügt (SAP Abschnitt 8.3.1.5). • Analysen zur Dauerhaftigkeit des Behandlungseffekts wurden dem finalen SAP hinzugefügt (SAP Abschnitt 8.3.4.1).
4	Probanden / Patienten	
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	**Einschlusskriterien:** 1. Die Probandin hat vor der Durchführung von studienspezifischen Verfahren ein ICF unterschrieben. 2. Die Probandin ist eine nicht-stationär behandelte Frau im Alter von 18 bis 45 Jahren, einschließlich. 3. Die Probandin ist in gutem körperlichem Gesundheitszustand und hat keine klinisch signifikanten Befunde, wie vom Prüfarzt bei der körperlichen Untersuchung, dem 12-Kanal-EKG oder klinischen Laboruntersuchungen festgestellt. 4. Die Probandin stimmt zu, die Studienanforderungen einzuhalten, einschließlich der Teilnahme an keiner Nachtschichtarbeit. 5. Die Probandin hat das Stillen eingestellt oder stimmt zu, ihrem Kind/ihren Kindern von kurz vor der Einnahme der Studienmedikation am Tag 1 bis 7 Tage nach der letzten Dosis der Studienmedikation keine Brustmilch zu geben. 6. Die Probandin muss beim Screening und am Tag 1 vor Beginn der Verabreichung der Studienmedikation einen negativen Schwangerschaftstest haben. 7. Die Probandin hatte eine MDD Episode, die nicht früher als im dritten Trimester und nicht später als in den ersten 4 Wochen nach der Entbindung begann, und erfüllt die Kriterien für eine schwere depressive Episode gemäß DSM-5, diagnostiziert durch SCID-5-CT. 8. Die Probandin hat einen HAMD-Gesamtwert von ≥ 26 beim Screening und am Tag 1 (vor der Randomisierung).

Item[a]	Charakteristikum	Studieninformation
		9. Die Probandin ist ≤ 12 Monate nach der Entbindung beim Screening und am Tag 1. 10. Probandinnen, die Antidepressiva einnehmen, müssen diese Medikamente in derselben Dosis seit mindestens 30 Tagen vor Tag 1 eingenommen haben. Teilnehmer, die die Einnahme von Antidepressiva innerhalb von 30 Tagen eingestellt haben, müssen dies länger als 5 Halbwertszeiten des Antidepressivums vor Tag 1 getan haben und werden in die „Keine Antidepressiva-Gruppe“ eingestuft. Teilnehmer, die Psychotherapie erhalten, müssen die Therapie seit mindestens 30 Tagen vor Tag 1 regelmäßig erhalten haben. 11. Die Probandin ist bereit, den Beginn anderer Antidepressiva oder Angstmedikamente sowie neuer Pharmakotherapie-Regime, einschließlich bedarfsorientierter Benzodiazepin-Angstlöser und Schlafmittel, bis nach Abschluss des Besuchs am Tag 45 zu verschieben. 12. Die Probandin stimmt zu, während der Teilnahme an der Studie und für 30 Tage nach der letzten Dosis der Studienmedikation mindestens 1 Methode der hochwirksamen Empfängnisverhütung gemäß Abschnitt 9.2.4 zu verwenden, es sei denn, sie ist chirurgisch steril (bilaterale Oophorektomie, Hysterektomie und/oder bilaterale Salpingektomie) oder hat keinen Geschlechtsverkehr, der ein Risiko einer Schwangerschaft birgt (Enthaltsamkeit nicht mit eingeschlossen). 13. Die Probandin stimmt zu, während der gesamten Studiendauer auf Drogenmissbrauch und Alkohol zu verzichten. **Ausschlusskriterien:** 1. Die Probandin ist derzeit einem signifikanten Suizidrisiko ausgesetzt, wie vom Prüfarzt beurteilt, oder hat einen Suizidversuch im Zusammenhang mit der aktuellen Episode der PPD unternommen. 2. Die Probandin hat eine aktuelle Anamnese mit aktiver klinisch signifikanter Manifestation von metabolischen, hepatischen, renalen, hämatologischen, pulmonalen, kardiovaskulären, gastrointestinalen, muskuloskelettalen, dermatologischen, urogenitalen, neurologischen oder Augen-, Ohren-, Nase- und Rachenstörungen oder einer anderen akuten oder chronischen Erkrankung, die nach Meinung des Prüfarztes die Fähigkeit des Teilnehmers, an dieser klinischen Studie teilzunehmen oder diese abzuschließen, einschränken würde. Ein Body-Mass-Index (BMI) ≤ 18 oder ≥ 45 kg/m² beim Screening ist ausschlusswürdig; ein BMI von 40 bis 44,9 kg/m², einschließlich, beim Screening unterliegt einer umfassenderen Bewertung medizinischer

Item[a]	Charakteristikum	Studieninformation
		Begleiterkrankungen (z. B. chronisch obstruktive Lungenerkrankung), begleitender Medikamente und vorheriger Verträglichkeit von sedierenden Mitteln. 3. Die Probandin hat beim Screening oder am Tag 1 einen klinisch signifikant abnormalen 12-Kanal-EKG-Befund. HINWEIS: Ein mittlerer QT-Intervall, berechnet mit der Fridericia-Methode von > 470 ms, ist die Grundlage für den Ausschluss aus der Studie. 4. Die Probandin hat eine bekannte Allergie gegen Zuranolon, Allopregnanolon oder verwandte Verbindungen. 5. Die Probandin hat aktive Psychosen gemäß der Beurteilung des Prüfarztes. 6. Die Indexschwangerschaft der Probandin führte zu einer Fehlgeburt, Totgeburt oder neonatalem/infantilem Tod; oder die Probandin hat die elterlichen Rechte aufgegeben (z. B. das Kind wurde zur Adoption freigegeben). 7. Die Probandin hat eine medizinische Historie von nicht-febrilen Anfällen. 8. Die Probandin hat eine medizinische Historie von bipolaren Störungen, Schizophrenie und/oder schizoaffektiven Störungen. 9. Die Probandin hat in den 12 Monaten vor dem Screening eine Anamnese mit milden, moderaten oder schweren Substanzgebrauchsstörungen (einschließlich Benzodiazepinen) gemäß DSM-5-Kriterien. 10. Die Probandin hatte innerhalb von 30 Tagen vor dem Screening Kontakt mit einem anderen Prüfmedikament oder -gerät. 11. Die Probandin hat zuvor an dieser Studie oder an einer anderen Studie teilgenommen, die Zuranolon oder SAGE-547 (Brexanolon), Ganaxolon, eine andere Verbindung, die Allopregnanolon enthält, oder wurde zuvor mit ZULRESSO™ (Brexanolon) behandelt. 12. Die Probandin hat eine Anamnese mit Schlafapnoe. 13. Die Probandin hat eine Vagusnervstimulation, Elektrokrampftherapie erhalten oder hat während der aktuellen PPD-Episode Ketamin eingenommen. 14. Die Probandin hat eine Magenbypassoperation, eine Schlauchmagen- oder LAP-Band-Operation gehabt oder hat andere verwandte Verfahren gehabt, die den gastrointestinalen Transit beeinträchtigen. 15. Die Probandin nimmt innerhalb von 28 Tagen vor Tag 1 Benzodiazepine, Barbiturate oder GABA(A)-Rezeptormodulatoren (z. B. Eszopiclon, Zopiclon, Zaleplon und Zolpidem) ein oder hat diese Mittel täglich oder nahezu täglich (≥ 4 Mal pro Woche) seit mehr als 1 Jahr verwendet. Die Probandin nimmt innerhalb von

Item[a]	Charakteristikum	Studieninformation
		30 Tagen vor Tag 1 ein Benzodiazepin oder GABA-Modulator mit einer Halbwertszeit von ≥ 48 Stunden (z. B. Diazepam) ein. 16. Die Probandin nimmt innerhalb von 14 Tagen vor Tag 1 nicht-GABA-haltige Schlafmittel die nicht GABA-modulieren (z. B. Melatonin, Benadryl [Antihistaminika], Trazodon) oder erste oder zweite Generation (typische/atypische) Antipsychotika ein. 17. Die Probandin wurde innerhalb des letzten Jahres vor dem Screening mit und/oder wegen irgendeiner Art von Krebs (außer Basalzellkarzinom und in situ Melanom) diagnostiziert und/oder behandelt. 18. Die Probandin nimmt innerhalb von 28 Tagen vor Tag 1 Psychostimulanzien (z. B. Methylphenidat, Amphetamin) oder Opioide regelmäßig oder nach Bedarf ein. 19. Verwendung von bekannten starken Inhibitoren des CYP3A4 innerhalb von 14 Tagen oder 5 Halbwertszeiten (je nachdem, was länger ist) oder Konsum von Grapefruitsaft, Grapefruit, Sevilla-Orangen oder Produkten, die diese enthalten, innerhalb von 14 Tagen vor der ersten Dosis des IP. 20. Die Probandin hat beim Screening oder am Tag 1 vor der Dosis einen positiven Urindrogentest und/oder einen positiven Alkoholtest. 21. Verwendung von starken CYP3A-Induktoren, wie Rifampin, Carbamazepin, Enzalutamid, Mitotan, Phenytoin oder Johanniskraut innerhalb von 14 Tagen oder 5 Halbwertszeiten (je nachdem, was länger ist) vor der ersten Dosis des IP. 22. Die Probandin plant, während der Teilnahme an der Studie eine elektive Operation durchzuführen. 23. Die Probandin ist ein Angehöriger des Sponsors, Prüfarztes, des Stellvertreters des Prüfarztes oder des Personals des Studienzentrums. 24. Die Probandin ist unfreiwillig in einer Strafanstalt oder einer anderen Einrichtung unter Gerichtsbeschluss untergebracht.
4b	Studienorganisation und Ort der Studiendurchführung	**Studienorganisation:** Sage Therapeutics, Inc **Orte der Studiendurchführung:** 78 Standorte in den USA, 9 Standorte in Spanien und 5 Standorte in UK.
5	Interventionen Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	Teilnehmerinnen verabreichten sich selbst Zuranolon (50 mg oder 40 mg [nur im Falle einer Dosisanpassung]) 1 x täglich oral gegen 20 Uhr mit Nahrung über einen Zeitraum von 14 Tagen. Die Dosen von 50 mg und 40 mg werden als 2 Kapseln pro Dosis verabreicht (50 mg, verabreicht als eine 30-mg-Kapsel und eine 20-mg-Kapsel, und 40 mg, verabreicht als zwei 20-mg-Kapseln). Placebo

Item[a]	Charakteristikum	Studieninformation
		wird ebenfalls als 2 Kapseln verabreicht, um die Verblindung aufrechtzuerhalten Zuranolon wurde als Hartgelatinekapseln bereitgestellt, die ein weißes bis cremefarbenes Pulver enthalten. Neben der angegebenen Menge des Wirkstoffs Zuranolon enthielten die aktiven Zuranolon-Kapseln Croscarmellose-Natrium, Mannitol, silicifiziertes mikrokristallines Cellulose, kolloidales Siliziumdioxid und Natriumstearylfumarat als Hilfsstoffe. Die Kapseln waren in den Dosierungen von 20 mg und 30 mg erhältlich. Die entsprechenden Placebo-Kapseln waren Hartgelatinekapseln, die nur die für die aktive Kapsel aufgeführten Hilfsstoffe enthielten. Die Chargennummern für Zuranolon waren 3690473 und 3694014 (30 mg) sowie 3690471 und 4040888 (20 mg). Die Chargennummer für die in dieser Studie verabreichten Placebo-Kapseln war 3692500. Nach Erhalt der Studienmedikation überprüfte das Prüfpersonal oder das verantwortliche Apothekenpersonal den Erhalt der Studienmedikation und dokumentierte diesen. Das zuständige Personal vor Ort gab die bereitgestellten, teilnehmerspezifischen Kits am Tag 1 und Tag 8 an die Teilnehmerinnen aus.
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	**Primärer Endpunkt:** Der primäre Wirksamkeitsendpunkt ist die Veränderung im HAMD-Gesamtwert zu Tag 15 gegenüber Baseline. **Sekundäre Endpunkte:** - Veränderung im HAMD-Gesamtwert zu Tag 3 gegenüber Baseline - Veränderung im HAMD-Gesamtwert zu Tag 28 gegenüber Baseline - Veränderung im HAMD-Gesamtwert zu Tag 45 gegenüber Baseline - Veränderung im CGI-S-Wert zu Tag 15 gegenüber Baseline - HAMD-Ansprechen zu Tag 15 und Tag 45, definiert über eine ≥ 50 %-ige Reduktion des HAMD-Gesamtwerts im Vergleich zu Baseline. - HAMD-Remission zu Tag 15 und Tag 45, definiert als ein HAMD-Gesamtwert ≤ 7. - CGI-I Ansprechen zu Tag 15, definiert als „sehr stark verbessert“ oder „stark verbessert“ gemäß Fragebogen. - Veränderung im HAM-A-Gesamtwert zu Tag 15 gegenüber Baseline.

Item[a]	Charakteristikum	Studieninformation
		- Veränderung im MADRS-Gesamtwert zu Tag 15 gegenüber Baseline. - Veränderung in der HAMD-Subskala zu Tag 15 gegenüber Baseline. - Veränderung in selbstberichteten Instrumenten zu depressiven Symptomen, bewertet durch den EPDS-Gesamtwert und PHQ-9. - Inzidenz von behandlungsbedingten UE. **Explorative Endpunkte:** - Veränderung in den Vitalzeichen, EKG, klinischen Laborparametern, C-SSRS-Antworten und der 20-Item Physician Withdrawal Checklist (PWC-20). - PK-Parameter (z. B. Clearance) und Expositionsschätzungen (z. B. Fläche unter der Kurve über ein Dosierungsintervall, maximale Plasmakonzentration), bewertet durch Population PK-Methoden.
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	Bei den folgenden Endpunkten wurde der Zeitpunkt im finalen SAP vom 11. April 2022 spezifiziert: - Es wurden Testsequenzen für wichtige sekundäre Endpunkte definiert, darunter die folgenden (Abschnitt 8.3.3.5 im SAP): o Veränderung im HAMD-17-Gesamtwert am Tag 3 gegenüber Baseline o Veränderung im HAMD-17-Gesamtwert am Tag 28 gegenüber Baseline o Veränderung im HAMD-17-Gesamtwert am Tag 45 gegenüber Baseline o Veränderung im CGI-S am Tag 15 gegenüber Baseline - Die MADRS-Response (≥ 50 %-ige Reduktion des MADRS-Gesamtwert im Vergleich zu Baseline) und Remission (MADRS-Gesamtwert ≤ 10) wurden definiert und dem finalen SAP hinzugefügt (SAP Abschnitt 8.3.1.5).
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	Unter der Annahme eines zweiseitigen Tests mit einem Alpha-Niveau von 0,05 würde eine Stichprobengröße von etwa 86 auswertbaren Teilnehmern pro Behandlungsgruppe eine statistische Power von 90 % bieten, um einen placebo-adjustierten Behandlungseffekt von etwa 4 Punkten im primären Endpunkt, der Veränderung im HAMD-Gesamtwert am Tag 15 gegenüber der Baseline, bei einer angenommenen SD von 8 Punkten zu erkennen. Unter der Annahme einer Abbruchrate von 10 % und eines Randomisierungsverhältnisses von 1 : 1 werden etwa 192 randomisierte Teilnehmerinnen (96 pro Behandlungsgruppe)

Item[a]	Charakteristikum	Studieninformation
		benötigt, um 86 auswertbare Teilnehmerinnen pro Behandlungsgruppe zu erhalten. Auswertbare Teilnehmerinnen sind definiert als die randomisierten Teilnehmerinnen, die das Studienmedikament erhalten haben und einen gültigen Ausgangswert sowie mindestens eine Nachuntersuchung des HAMD-17 haben. Zusätzliche Teilnehmerinnen könnten randomisiert werden, wenn die Abbruchrate höher als 10 % ist.
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	**Zwischenanalysen** Es wurden keine Interimsanalysen geplant oder durchgeführt. **Kriterien für einen vorzeitigen Studienabbruch** Sage Therapeutics kann diese Studie oder einen Teil der Studie jederzeit aus Sicherheitsgründen beenden, einschließlich des Auftretens von UE oder anderer Befunde, die ein inakzeptables Risiko für die Teilnehmerinnen nahelegen, oder aus administrativen Gründen. Im Falle einer Studienbeendigung wird Sage Therapeutics dem Prüfarzt eine schriftliche Benachrichtigung zukommen lassen. Die Prüfzentren müssen umgehend ihre IRB benachrichtigen und die Withdrawal-Verfahren für die teilnehmenden Teilnehmerinnen einleiten.
8	Randomisierung, Erzeugung der Behandlungsfolge	
8a	Methode zur Generierung der zufälligen Zuteilung	Die Randomisierungspläne werden von einem unabhängigen Statistiker erstellt. Die Einteilung in die Behandlungsgruppe (Zuranolon, 50 mg oder Placebo) erfolgt auf der Grundlage des Randomisierungsplans. Die Randomisierung erfolgt zentral über ein IRT-System.
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	Teilnehmerinnen, die die Einschlusskriterien erfüllen, werden in einer stratifizierten Weise randomisiert, basierend auf der Verwendung von Antidepressiva zum Ausgangszeitpunkt (Ja = aktuell und stabil; Nein = nicht behandelt oder abgesetzt (≥ 30 Tage oder > 5 Halbwertszeiten)). Die Randomisierung erfolgt innerhalb jeder Stratum im Verhältnis 1 : 1, um entweder 50 mg Zuranolon oder Placebo zu erhalten.
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	Teilnehmerinnen, die die Einschlusskriterien erfüllen, werden am Tag 1 im Verhältnis 1 : 1 zufällig entweder Zuranolon oder Placebo zugewiesen. Die Teilnehmerinnen, ihre behandelnden Ärztinnen/Ärzte und das Studienteam bleiben hinsichtlich der Behandlungszuweisung verblindet. Die Randomisierung erfolgt zentral über ein IRT. Die Randomisierungspläne werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt, wobei nur autorisiertes Personal bis zur Entblindung Zugriff darauf hat. Im Falle eines medizinischen Notfalls darf das Apothekenpersonal die tatsächliche Studienmedikation dem Prüfpersonal offenlegen, welches Sage über den Notfall informieren muss (siehe Abschnitt

Item[a]	Charakteristikum	Studieninformation
		13.6 für weitere Details zur Entblindung). Jede Entblindung muss mit den entsprechenden Informationen, einschließlich des Grundes, in den Patientinnen-/Patientenunterlagen und im eCRF dokumentiert werden. Falls die Teilnehmerin oder andere Studienmitarbeiterinnen/Studienmitarbeiter (außer Apothekenpersonal) entblindet werden, wird die Teilnehmerin aus der Studie ausgeschlossen.
10	Randomisierung, Durchführung Wer hat die Randomisierungsliste erstellt, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	Nach Erhalt der Studienmedikation inspizierte und dokumentierte das Prüfpersonal oder das verantwortliche Apothekenpersonal den Erhalt der Studienmedikation. Das dafür vorgesehene Personal der Prüfzentren gab die teilnehmerspezifischen Kits an die Teilnehmerinnen an Tag 1 und Tag 8 aus. Das Personal der Prüfzentren griff beim Screening-Besuch auf das IRT-System zu, um für jede Teilnehmerin eine ID zu erhalten. An Tag 1 griff das Personal erneut auf das IRT-System zu, um die berechtigte Teilnehmerin für die Studie zu randomisieren und die Medikamenten-ID für die auszugebende Studienmedikation zu erhalten. Die Medikamenten-ID sowie die Anzahl der ausgegebenen Kapseln wurden dokumentiert. Beim darauffolgenden Besuch zur Ausgabe der Studienmedikation griff die Prüferin oder Beauftragte auf das IRT-System zu, um die Medikamenten-ID für die an diesem Besuch auszugebende Studienmedikation zu erhalten. Die Medikamenten-ID, die Anzahl der ausgegebenen Kapseln sowie die Anzahl der von der Teilnehmerin zurückgegebenen Kapseln wurden dokumentiert. Falls Fehler oder Unstimmigkeiten bei der Ausgabe durch das Personal der Prüfzentren oder die Beauftragten des Sponsors festgestellt wurden, musste der Sponsor unverzüglich informiert werden. Die Studienmedikation wurde ausschließlich gemäß den Vorgaben des Protokolls verwendet. Die Prüferin oder Beauftragte führte ein Protokoll über alle erhaltenen, verwendeten und zurückgegebenen/entsorgten Studienmedikationen. Am Ende der Studie wurden alle ungenutzten Studienmedikationen entweder an Sage Therapeutics zur Vernichtung zurückgeschickt oder gemäß den Verfahren des Prüfzentrums vor Ort entsorgt. Sage Therapeutics hatte jederzeit mit entsprechender Vorankündigung während oder nach Abschluss der Studie Zugang zu den Studienmaterialien, um eine Überprüfung der Medikamentenverantwortlichkeit durchzuführen.
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,	Die Randomisierung wurde zentral über ein IRT-System durchgeführt. Patientinnen (a, b), Kliniker*innen (b, c) und das Studienteam (b, c) waren bis zur Datenbanksperre verblindet gegenüber der Behandlungszuweisung. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn

Item[a]	Charakteristikum	Studieninformation
	verblindet oder nicht verblindet, wie wurde die Verblindung vorgenommen?	die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing. Es fand während der Studie keine Aufhebung der Verblindung statt.
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Zuranolon-Kapseln waren als harte Gelatinekapseln (inklusive Hilfsstoffe) erhältlich, die ein weißes bis cremefarbenes Pulver enthielten. Die passenden Placebo-Kapseln waren harte Gelatinekapseln, die nur die entsprechenden Hilfsstoffe enthielten
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	Das Full Analysis Set (FAS) wurde definiert als alle randomisierten Teilnehmerinnen, die die Studienmedikation mit einem gültigen Baselinewert im Gesamtwert und mindestens einem Nachuntersuchungswert im Gesamtwert in mindestens einem der Instrumente HAMD-17, HAM-A, MADRS, CGI-S, EPDS oder PHQ-9 oder mindestens einem Nachuntersuchungswert des CGI-I erhalten haben. Die Wirksamkeitsanalysen wurden unter Verwendung des FAS durchgeführt, die primäre Wirksamkeitsanalyse wurde unter Verwendung des PP-Sets durchgeführt, das als alle Teilnehmerinnen im FAS ohne wesentliche Protokollabweichungen definiert ist, die die Wirksamkeit beeinflussen könnten. Darüber hinaus schloss das PP-Set FAS-Teilnehmerinnen aus, die weniger als 22 Kapseln (d. h. weniger als 80 % der zugewiesenen Kapselanzahl) konsumierten, Teilnehmerinnen, die zu irgendeinem Zeitpunkt während der Studie falsche Studienmedikation (d. h. Studienmedikation, das nicht dem entsprach, was ihnen zugewiesen wurde) konsumierten, sowie Teilnehmerinnen oder Studienpersonal, die vor der Datenbanksperre über die Behandlungszuweisung des Teilnehmers informiert wurden. Sicherheitsanalysen wurden unter Verwendung des Safety Sets durchgeführt, das als alle Teilnehmerinnen definiert ist, die die verblindete Studienmedikation selbst verabreicht haben. Die Sicherheitsdaten wurden basierend auf dem tatsächlich erhaltenen Studienmedikament analysiert. Die Veränderung des HAMD-17-Gesamtwerts gegenüber dem Ausgangswert wurde unter Verwendung eines MMRM analysiert, das die Behandlungsgruppe (Zuranolon oder Placebo), den Ausgangswert des HAMD-17-Gesamtwerts, den Einsatz von Antidepressiva zu Beginn (ja oder nein), den Bewertungszeitpunkt und die Interaktion zwischen Zeitpunkten und Behandlung als erklärende Variablen einbezog. Alle erklärenden Variablen wurden als feste Effekte behandelt. Alle Untersuchungszeitpunkte nach Baseline wurden in das Modell einbezogen. Der Hauptvergleich erfolgte zwischen Zuranolon und Placebo am Tag 15. Modellbasierte Punkteschätzungen (d. h. Behandlungsunterschied im LS-MW, 95 % KI und p-Werte)

Item[a]	Charakteristikum	Studieninformation
		wurden berichtet. Der p-Wert wurde auf einem zweiseitigen Signifikanzniveau von 5 % interpretiert. Wenn der Vergleich von Zuranolon versus Placebo auf dem Niveau von 0,05 signifikant war, folgte die Hypothesentestung für die wichtigsten sekundären Endpunkte mit einer Multiplikationsanpassung. Unerwünschte Ereignisse wurden gemäß dem MedDRA, Version 24.0, klassifiziert und nach SOC und PT sowie Schweregrad und Zusammenhang zusammengefasst. Vitalzeichen, klinische Laborwerte, EKG, C-SSRS, PWC-20 sowie die Verwendung von vorherigen und begleitenden Medikamenten wurden ebenfalls zusammengefasst.
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	Die Subgruppen werden für die folgenden Kategorien zusammengefasst: - Abstammung (Schwarz oder Afroamerikanisch; Weiß; Andere) - Alter (18 – 24; 25 – 45 Jahre) - Verwendung von Antidepressiva zu Beginn (Ja; Nein) - BMI (≤ 18,4; 18,5 – 24,9; 25 – 29,9; ≥ 30 kg/m²) - Ausgangswert des HAMD-Gesamtwerts (< Median des Ausgangswerts des HAMD-Gesamtwerts; ≥ Median des Ausgangswerts des HAMD-Gesamtwerts; Median wird auf dem FAS berechnet) - Beginn der postpartalen Depression (3. Trimester vs. *postpartal*) - PPD in der Anamnese (1. Episode vs. wiederkehrende Episode) - HAMD-17 zu Beginn (≥ 7 vs. < 7 oder normalisierter Score ≥ 39 vs. < 39) - Land (USA; Rest der Welt) - Depression mit erhöhter Angst, basierend auf der HAM-A-Gesamtpunktzahl
Resultate		
13	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	Von den 505 Teilnehmerinnen im Screening waren 305 Screen-Failures. 200 Teilnehmerinnen gaben ihr Einverständnis via ICF ab und wurden in die Studie randomisiert, von denen 196 Teilnehmerinnen mindestens eine Dosis der Studienmedikation erhielten (98 Placebo, 98 Zuranolon). Vier Teilnehmerinnen wurden randomisiert, aber nicht mit Studienmedikation behandelt, aufgrund von Rückzug durch den Teilnehmerinnen (2 Placebo, 1 Zuranolon) und Verlust der Nachverfolgbarkeit (1 Placebo). 180 Teilnehmerinnen (91 Placebo, 89 Zuranolon) schlossen die Behandlung mit der Studienmedikation ab. 170 Teilnehmerinnen (86 Placebo, 84 Zuranolon) beendeten die Studie. Verlust der Nachverfolgbarkeit (8 Placebo, 6 Zuranolon) und Rückzug durch

Item[a]	Charakteristikum	Studieninformation
		die Teilnehmerinnen (3 Placebo, 4 Zuranolon) waren die am häufigsten berichteten Gründe für den Abbruch der Studie.
13a	Anzahl der Studienteilnehmerinnen 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	a) in der SKYLARK-Studie wurden 101 Teilnehmerinnen der Placebo-Gruppe und 99 Teilnehmerinnen der Zuranolon-Gruppe zugewiesen. b) 196 erhielten mindestens 1 Dosis der Studienmedikation (98 Placebo und 98 Sage-217) c) In der Analyse des primären Zielkriteriums wurden alle Patientinnen, die dem FAS zugehörig waren analysiert. Dem FAS gehörten 195 Patientinnen, davon 97 Patientinnen aus dem Placebo-Arm und 98 Patientinnen aus dem Zuranolon-Arm an.
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	Die Mehrheit der Teilnehmerinnen in jeder Gruppe schloss die Behandlung mit der Studienmedikation ab (92, % Placebo, 90,8 % Zuranolon). Der Prozentsatz der Teilnehmerinnen, die die Studienmedikation vorzeitig abbrachen, war zwischen den Gruppen ähnlich (7,1 % Placebo, 9,2 % Zuranolon). Die am häufigsten berichteten Gründe für den Abbruch der Behandlung mit der Studienmedikation waren UE (2,0 % Placebo, 4,1 % Zuranolon) und Verlust der Nachverfolgbarkeit (3,1 % Placebo, 2,0 % Zuranolon). Die meisten Teilnehmerinnen in jeder Gruppe schlossen die Studie ab (87,8 % Placebo, 85,7 % Zuranolon). Die am häufigsten berichteten Gründe für den Abbruch der Studie waren Verlust der Nachverfolgbarkeit (8,2 % Placebo, 6,1 % Zuranolon) und Rückzug durch die Teilnehmerinnen (3,1 % Placebo, 4,1 % Zuranolon). Nur eine Teilnehmerin (1 %) in einer der Behandlungsgruppen brach die Studie aufgrund eines UE ab.
14	Aufnahme / Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	Die erste Patientin wurde am 08. Juni 2020 in die Studie aufgenommen. Die letzte Patientin hatte ihre letzte Visite am 12 April 2022 abgeschlossen. Die Nachbeobachtung erfolgte bis Tag 45 nach Beginn der Verabreichung des Studienmedikaments an Tag 1.
14b	Informationen, warum die Studie endete oder beendet wurde	Die Studie wurde wie geplant durchgeführt und analysiert (Studienbericht vom 22.08.2022).
a: nach CONSORT 2010.		

Stellen Sie für jede Studie den Patientenfluss in einem Flow-Chart gemäß CONSORT dar.

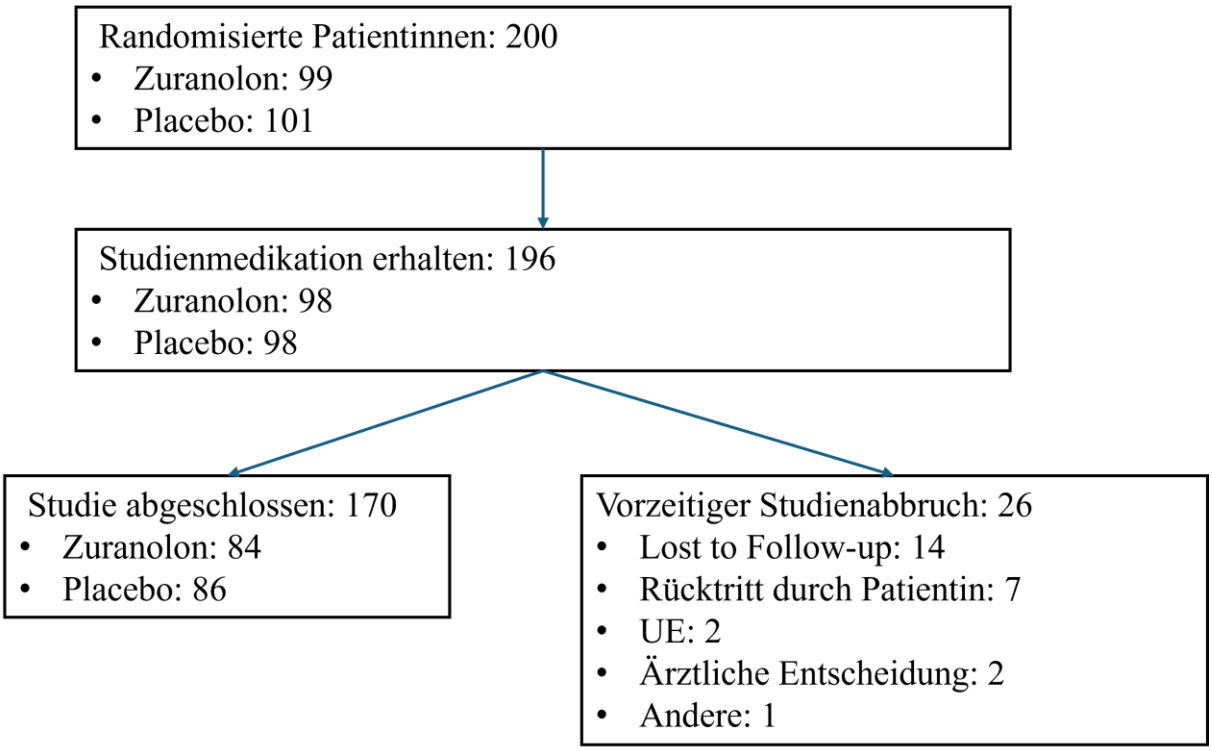


Abbildung 4-19: Patientenfluss der Studie SKYLARK
Quelle: CSR Seite 61, Tabelle 7 [14]

Anhang 4-F: Bewertungsbögen zur Einschätzung von Verzerrungsaspekten

Der nachfolgend dargestellte 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:

- Verzerrungsaspekte auf Studienebene. In diesem Teil sind die endpunktübergreifenden Kriterien aufgelistet.

- Verzerrungsaspekte auf Endpunktebene. 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 gegebenenfalls, auf sehr umfangreiche Informationen in den Quellen kann verwiesen werden).

Grundsätzlich sollen die Bögen studienbezogen ausgefüllt werden. Wenn mehrere Quellen zu einer Studie vorhanden sind, müssen die herangezogenen Quellen in der folgenden Tabelle genannt und jeweils mit Kürzeln (zum Beispiel A, B, C …) versehen werden. Quellenspezifische Angaben im weiteren Verlauf sind mit dem jeweiligen Kürzel zu kennzeichnen.

Hinweis: Der nachfolgend dargestellte Bewertungsbogen ist die Blankoversion des Bogens. Dieser Blankobogen ist für jede Studie heranzuziehen. Im Anschluss daran ist ein Bewertungsbogen inklusive Ausfüllhinweisen abgebildet, der als Ausfüllhilfe dient, aber nicht als Vorlage verwendet werden soll.

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.

Studie: ROBIN, Teil B

Tabelle 4-117 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie ROBIN (Teil B)

Studie: Robin, Teil B

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genaue Benennung der Quelle	Kürzel
Studienbericht: A Multicenter, Randomized, Double-Blind, Parallel-Group Placebo-Controlled Study Evaluating the Efficacy, Safety, and Pharmacokinetics of SAGE-217 in the Treatment of Adult Female Subjects with Severe Postpartum Depression - 217-PPD-201 (ROBIN) [15]	A
Studienprotokoll: Protocol: A Multicenter, Randomized, Double-Blind, Parallel-Group, Placebo-Controlled Study Evaluating the Efficacy, Safety, and Pharmacokinetics of SAGE-217 in the Treatment of Adult Female Subjects With Severe Postpartum Depression - Protocol Number: 217-PPD-201 (ROBIN) [17]	B
Statistischer Analyseplan: A MULTICENTER, RANDOMIZED, DOUBLE-BLIND, PARALLEL-GROUP, PLACEBO-CONTROLLED STUDY EVALUATING THE EFFICACY, SAFETY, AND PHARMACOKINETICS OF SAGE-217 IN THE TREATMENT OF ADULT FEMALE SUBJECTS WITH SEVERE POSTPARTUM DEPRESSION - Version 1.0 [18]	C

A Verzerrungsaspekte auf Studienebene:

Einstufung als randomisierte Studie

☒ **ja** → Bewertung der Punkte 1 und 2 für randomisierte Studien

☐ **nein** → Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

Bei der Studie ROBIN handelt es sich um eine multizentrische, randomisierte, kontrollierte und doppelt verblindete Phase 3 Studie im Parallelgruppendesign.

1.

<u>für randomisierte Studien:</u> Adäquate Erzeugung der Randomisierungssequenz

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Teilnehmerinnen, die die Einschlusskriterien erfüllen, wird nacheinander eine Randomisierungsnummer aus dem Randomisierungsschema zugewiesen, das dem entblindeten pharmazeutischen Personal zur Verfügung gestellt wird (vor Protokollversion 8.0) oder die Randomisierung wird zentral über eine IRT durchgeführt (Protokollversion 8.0 und danach).

für nicht randomisierte Studien: Zeitliche Parallelität der Gruppen

☐ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Nicht zutreffend.

2.

für randomisierte Studien: Verdeckung der Gruppenzuteilung („allocation concealment“)

☒ **ja** ☐ **unklar** ☐ **nein**

In der Studie ROBIN fand eine doppelte Verblindung statt, dementsprechend waren Patientinnen, Kliniker*innen und das Studienteam über die Zuteilung zur Behandlung verblindet. Vor der Protokolländerung 7, Version 8.0, waren das Apothekenpersonal, das für die Abgabe des Studienmedikaments verantwortlich war, über das Randomisierungsschema informiert. Unter Protokolländerung 7, Version 8.0 und danach, wurde die Randomisierung zentral über ein IRT-System durchgeführt, und alle Mitarbeitenden vor Ort, einschließlich des Apothekenpersonals, waren verblindet; entblindetes pharmazeutisches Personal, das vor Änderung 7 für die Abgabe des Studienmedikaments verantwortlich war, waren nach der Einführung des IRT-Systems nicht mehr an der Durchführung der Studie beteiligt.

Nur im Falle eines medizinischen Notfalls durfte das zuständige Apothekenpersonal den tatsächlichen Inhalt des Studienmedikaments dem Prüfarzt/der Prüfärztin offenlegen. Es trat kein medizinischer Notfall ein.

für nicht randomisierte Studien: Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

☐ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und behandelnden Personen

Patient:

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

In der Studie ROBIN fand eine doppelte Verblindung statt, dementsprechend waren Patientinnen, Kliniker*innen und das Studienteam über die Zuteilung zur Behandlung verblindet. Vor der Protokolländerung 7, Version 8.0, waren das Apothekenpersonal, das für die Abgabe des Studienmedikaments verantwortlich war, über das Randomisierungsschema informiert. Unter Protokolländerung 7, Version 8.0 und danach, wurde die Randomisierung zentral über ein IRT-System durchgeführt, und alle Mitarbeitenden vor Ort, einschließlich des Apothekenpersonals, waren verblindet; entblindetes pharmazeutisches Personal, das vor Änderung 7 für die Abgabe des Studienmedikaments verantwortlich war, waren nach der Einführung des IRT-Systems nicht mehr an der Durchführung der Studie beteiligt.

behandelnde bzw. weiterbehandelnde Personen:

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

In der Studie ROBIN fand eine doppelte Verblindung statt, dementsprechend waren Patientinnen, Kliniker*innen und das Studienteam über die Zuteilung zur Behandlung verblindet. Vor der Protokolländerung 7, Version 8.0, waren das Apothekenpersonal, das für die Abgabe des Studienmedikaments verantwortlich war, über das Randomisierungsschema informiert. Unter Protokolländerung 7, Version 8.0 und danach, wurde die Randomisierung zentral über ein IRT-System durchgeführt, und alle Mitarbeitenden vor Ort, einschließlich des Apothekenpersonals, waren verblindet; entblindetes pharmazeutisches Personal, das vor Änderung 7 für die Abgabe des Studienmedikaments verantwortlich war, waren nach der Einführung des IRT-Systems nicht mehr an der Durchführung der Studie beteiligt.

Nur im Falle eines medizinischen Notfalls durfte das zuständige Apothekenpersonal den tatsächlichen Inhalt des Studienmedikaments dem Prüfarzt/der Prüfärztin offenlegen. Es trat kein medizinischer Notfall ein.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C

Die in der Studie ROBIN erhobenen Endpunkte, sowie die durchzuführenden Analysen werden wie im Studienprotokoll und im statistischen Analyseplan präspezifiziert erhoben. *Post-hoc* definierte Analysen wurden ergänzt, um den Anforderungen der Nutzenbewertung zu entsprechen

5. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrungen führen können

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A,B,C
Es wurden keine sonstigen Aspekte identifiziert, die das Verzerrungspotential beeinflussen.

Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

A,B,C
Bei der zulassungsrelevanten Studie ROBIN handelt es sich um eine multizentrische, randomisierte und kontrollierte Phase 3-Studie im verblindeten Parallelgruppendesign. Das ITT-Prinzip wurde in der Studie adäquat umgesetzt. Es liegen keine sonstigen Aspekte vor, die das Verzerrungspotential beeinflussen können. Abschließend wird das Verzerrungspotential somit als niedrig eingestuft.

Studie: SKYLARK

Tabelle 4-118 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie SKYLARK

Studie: SKYLARK

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genaue Benennung der Quelle	Kürzel
Studienbericht: A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY EVALUATING THE EFFICACY AND SAFETY OF SAGE-217 IN THE TREATMENT OF ADULTS WITH SEVERE POSTPARTUM DEPRESSION - 217-PPD-301 (SKYLARK) - Version 1 [14]	A
Studienprotokoll: A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY EVALUATING THE EFFICACY AND SAFETY OF SAGE-217 IN THE TREATMENT OF ADULTS WITH SEVERE POSTPARTUM DEPRESSION - PROTOCOL NUMBER: 217-PPD-301 (SKYLARK) - EUDRACT NUMBER: 2020-001424-34 - Version 2 [19]	B
Statistischer Analyseplan: Protocol Number 217-PPD-301 (SKYLARK): A RANDOMIZED, DOUBLE-BLIND, PLACEBO-CONTROLLED STUDY EVALUATING THE EFFICACY AND SAFETY OF SAGE-217 IN THE TREATMENT OF ADULTS WITH SEVERE POSTPARTUM DEPRESSION - Version: Final Version 1.0 - Version Date of SAP: 11 April 2022 [20]	C

A Verzerrungsaspekte auf Studienebene:

Einstufung als randomisierte Studie

☒ **ja** → Bewertung der Punkte 1 und 2 für randomisierte Studien

☐ **nein** → Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

	A, B, C Bei der zulassungsrelevanten Studie SKYLARK handelt es sich um eine multizentrische, randomisierte, kontrollierte und doppelt verblindete Phase 3 Studie im Parallelgruppendesign.

1.

<u>für randomisierte Studien:</u> Adäquate Erzeugung der Randomisierungssequenz

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Die Randomisierung erfolgt zentral über ein IRT-System. Jeder Teilnehmerin wurde eine ID und eine zugehörige Medikamenten-ID vermittelt. Somit ist die adäquate Erzeugung der Randomisierungssequenz sowie eine verdeckte Gruppenzuteilung gewährleistet

für nicht randomisierte Studien: Zeitliche Parallelität der Gruppen

☐ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Nicht zutreffend.

2.

für randomisierte Studien: Verdeckung der Gruppenzuteilung („allocation concealment“)

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B, C

Bei der Studie SKYLARK handelt es sich um eine doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam bis zur Datenbanksperre verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing. Es fand während der Studie keine Aufhebung der Verblindung statt.

für nicht randomisierte Studien: Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

☐ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Nicht zutreffend.

3. Verblindung von Patienten und behandelnden Personen

Patient:

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei der Studie SKYLARK handelt es sich um eine doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam bis zur Datenbanksperre verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing. Es fand während der Studie keine Aufhebung der Verblindung statt.

behandelnde bzw. weiterbehandelnde Personen:

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei der Studie SKYLARK handelt es sich um eine doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam bis zur Datenbanksperre verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing. Es fand während der Studie keine Aufhebung der Verblindung statt.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Die in der Studie SKYLARK erhobenen Endpunkte, sowie die durchzuführenden Analysen werden wie im Studienprotokoll und im statistischen Analyseplan präspezifiziert erhoben. *Post-hoc* definierte Analysen wurden ergänzt, um den Anforderungen der Nutzenbewertung zu entsprechen.

5. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrungen führen können

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A,B,C
Es wurden keine sonstigen Aspekte identifiziert, die das Verzerrungspotential beeinflussen.

Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

A,B,C Bei der zulassungsrelevanten Studie SKYLARK handelt es sich um eine multizentrische, randomisierte und kontrollierte Phase 3-Studie im verblindeten Parallelgruppendesign. Das ITT-Prinzip wurde in der Studie adäquat umgesetzt. Es liegen keine sonstigen Aspekte vor, die das Verzerrungspotential beeinflussen können. Abschließend wird das Verzerrungspotential somit als niedrig eingestuft.

Verzerrung auf Endpunktebene für die Studien ROBIN (Teil B) und SKYLARK

B Verzerrungsaspekte auf Endpunktebene pro Endpunkt:

Endpunkt: Todesfälle

1. **Verblindung der Endpunkterheber**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing.

2. **Adäquate Umsetzung des ITT-Prinzips**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Die Analyse des Endpunkts basiert auf der Safety-Population, die randomisierte Patientinnen umfasst die mindestens eine Dosis der Studienmedikation (Zuranolon oder Placebo) erhalten haben.

3. **Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
In den Studien ROBIN und SKYLARK erfolgte die Erhebung des Endpunkts „Todesfälle“ im Rahmen der schwerwiegenden UE. Es wurden objektiv jegliche UE der Patientinnen, unabhängig davon, ob das Ereignis in kausalem Zusammenhang mit der Verabreichung der Studienmedikation stand, erhoben. Die Erhebung und Klassifikation der UE erfolgten auf Basis allgemein anerkannter internationaler Standards. Somit ist eine ergebnisunabhängige Erhebung gewährleistet.

4. **Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

	A,B,C Es liegen keine Anzeichen für sonstige Aspekte vor, die zu Verzerrung führen können.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam bis zur Datenbanksperre verblindet gegenüber der Behandlungszuweisung.
In den Studien ROBIN und SKYLARK erfolgte die Erhebung des Endpunkts „Todesfälle“ im Rahmen der schwerwiegenden UE. Es wurden objektiv jegliche UE der Patientinnen, unabhängig davon, ob das Ereignis in kausalem Zusammenhang mit der Verabreichung der Studienmedikation stand, erhoben. Die Erhebung und Klassifikation der UE erfolgten auf Basis allgemein anerkannter internationaler Standards. Somit ist eine ergebnisunabhängige Erhebung gewährleistet.
Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Endpunkt: HAMD-17

1. Verblindung der Endpunkterheber

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing.

2. Adäquate Umsetzung des ITT-Prinzips

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Die Analyse des Endpunkts basiert auf der ITT-Population, die randomisierte Patientinnen umfasst, unabhängig davon, ob sie die Studienintervention erhalten haben oder nicht. Somit ist das ITT-Prinzip adäquat erfüllt.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Der Endpunkt wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Es wird nicht von den präspezifizierten Methoden und Definitionen abgewichen.

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A,B,C
Es liegen keine Anzeichen für sonstige Aspekte vor, die zu Verzerrung führen können.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der Endpunkt wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Es wird nicht von den präspezifizierten Methoden und Definitionen abgewichen. Das ITT-Prinzip wurde adäquat umgesetzt.
Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Endpunkt: MADRS

1. **Verblindung der Endpunkterheber**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing.

2. **Adäquate Umsetzung des ITT-Prinzips**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Die Analyse des Endpunkts basiert auf der ITT-Population, die randomisierte Patientinnen umfasst, unabhängig davon, ob sie die Studienintervention erhalten haben oder nicht. Somit ist das ITT-Prinzip adäquat erfüllt.

3. **Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Der Endpunkt wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Es wird nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet.

4. **Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A,B,C
Es liegen keine Anzeichen für sonstige Aspekte vor, die zu Verzerrung führen können.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der Endpunkt wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Es wird nicht von den präspezifizierten Methoden und Definitionen abgewichen. Das ITT-Prinzip wurde adäquat umgesetzt.
Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Endpunkt: EPDS

1. **Verblindung der Endpunkterheber**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing.

2. **Adäquate Umsetzung des ITT-Prinzips**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Die Analyse des Endpunkts basiert auf der ITT-Population, die randomisierte Patientinnen umfasst, unabhängig davon, ob sie die Studienintervention erhalten haben oder nicht. Somit ist das ITT-Prinzip adäquat erfüllt.

3. **Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Der selbstberichtete Endpunkt EPDS wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Zusätzlich wurden post-hoc Analysen ergänzt, um den Anforderungen an die Nutzenbewertung zu entsprechen. Abgesehen von diesem Zusatz wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet.

4. **Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A,B,C
Es liegen keine Anzeichen für sonstige Aspekte vor, die zu Verzerrung führen können.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der selbstberichtete Endpunkt EPDS wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Zusätzlich wurden *post-hoc* Analysen ergänzt, um den Anforderungen an die Nutzenbewertung zu entsprechen. Abgesehen von diesem Zusatz wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Endpunkt: PHQ-9

1. Verblindung der Endpunkterheber

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing.

2. Adäquate Umsetzung des ITT-Prinzips

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Die Analyse des Endpunkts basiert auf der ITT-Population, die randomisierte Patientinnen umfasst, unabhängig davon, ob sie die Studienintervention erhalten haben oder nicht. Somit ist das ITT-Prinzip adäquat erfüllt.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Der selbstberichtete Endpunkt PHQ-9 wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Zusätzlich wurden *post-hoc* Analysen ergänzt, um den Anforderungen an die Nutzenbewertung zu entsprechen. Abgesehen von diesem Zusatz wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet.

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A,B,C
Es liegen keine Anzeichen für sonstige Aspekte vor, die zu Verzerrung führen können.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der selbstberichtete Endpunkt PHQ-9 wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Zusätzlich wurden *post-hoc* Analysen ergänzt, um den Anforderungen an die Nutzenbewertung zu entsprechen. Abgesehen von diesem Zusatz wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Endpunkt: HAM-A

1. **Verblindung der Endpunkterheber**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing.

2. **Adäquate Umsetzung des ITT-Prinzips**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Die Analyse des Endpunkts basiert auf der ITT-Population, die randomisierte Patientinnen umfasst, unabhängig davon, ob sie die Studienintervention erhalten haben oder nicht. Somit ist das ITT-Prinzip adäquat erfüllt.

3. **Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Der Endpunkt HAM-A wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben, um die Angstzustände der Patientinnen zu dokumentieren. Zusätzlich wurden *post-hoc* Analysen ergänzt, um den Anforderungen an die Nutzenbewertung zu entsprechen. Abgesehen von diesem Zusatz wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet.

4. **Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

	A,B,C Es liegen keine Anzeichen für sonstige Aspekte vor, die zu Verzerrung führen können.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der Endpunkt HAM-A wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben, um die Angstzustände der Patientinnen zu dokumentieren. Zusätzlich wurden *post-hoc* Analysen ergänzt, um den Anforderungen an die Nutzenbewertung zu entsprechen. Abgesehen von diesem Zusatz wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet. Das ITT-Prinzip wurde adäquat umgesetzt. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Endpunkt: SF-36

1. **Verblindung der Endpunkterheber**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei der Studie ROBIN handelt es sich um eine doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT wurden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing.

2. **Adäquate Umsetzung des ITT-Prinzips**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Die Analyse des Endpunkts basiert auf der ITT-Population, die randomisierte Patientinnen umfasst, unabhängig davon, ob sie die Studienintervention erhalten haben oder nicht. Somit ist das ITT-Prinzip adäquat erfüllt.

3. **Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Der SF-36 ist ein etabliertes und standardisiertes Instrument zur generischen Erhebung der gesundheitsbezogenen Lebensqualität bei Patientinnen. Der Endpunkt wurde wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Es wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Post-hoc definierte Analysen wurden ergänzt, um den Anforderungen der Nutzenbewertung zu entsprechen. Somit ist eine ergebnisunabhängige Berichterstattung für den Endpunkt gewährleistet.

4. **Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A,B,C
Es liegen keine Anzeichen für sonstige Aspekte vor, die zu Verzerrung führen können.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

Bei der Studie ROBIN handelt es sich um eine doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der generische Endpunkt SF-36 wird wie im Studienprotokoll und statistischen Analyseplan präspezifiziert erhoben, um die Lebensqualität der Patientinnen zu erheben. Zusätzlich wurden *post-hoc* Analysen ergänzt, um den Anforderungen an die Nutzenbewertung zu entsprechen. Abgesehen von diesem Zusatz wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. Eine ereignisunabhängige Berichterstattung ist gewährleistet. Das ITT-Prinzip wurde adäquat umgesetzt. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Endpunkt: BIMF

1. **Verblindung der Endpunkterheber**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei der Studie ROBIN handelt es sich um eine doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT wurden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing.

2. **Adäquate Umsetzung des ITT-Prinzips**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Die Analyse des Endpunkts basiert auf der ITT-Population, die randomisierte Patientinnen umfasst, unabhängig davon, ob sie die Studienintervention erhalten haben oder nicht. Somit ist das ITT-Prinzip adäquat erfüllt.

3. **Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine**

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

Der BIMF ist ein Fragebogen, den die Patientinnen eigenständig ausfüllen. Der Endpunkt wurde wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Es wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. *Post-hoc* definierte Analysen wurden ergänzt, um den Anforderungen der Nutzenbewertung zu entsprechen. Somit ist eine ergebnisunabhängige Berichterstattung für den Endpunkt gewährleistet.

4. **Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können**

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A,B,C Es liegen keine Anzeichen für sonstige Aspekte vor, die zu Verzerrung führen können.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

Bei der Studie ROBIN handelt es sich um eine doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der BIMF ist ein Fragebogen, den die Patientinnen eigenständig ausfüllen. Der Endpunkt wurde wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Es wurde nicht von den präspezifizierten Methoden und Definitionen abgewichen. *Post-hoc* definierte Analysen wurden ergänzt, um den Anforderungen der Nutzenbewertung zu entsprechen. Somit ist eine ergebnisunabhängige Berichterstattung für den Endpunkt gewährleistet. Das ITT-Prinzip wurde adäquat umgesetzt. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Endpunkt: C-SSRS

1. Verblindung der Endpunkterheber

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing.

2. Adäquate Umsetzung des ITT-Prinzips

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Die Analyse des Endpunkts basiert auf der Safety-Population, die randomisierte Patientinnen umfasst die mindestens eine Dosis der Studienmedikation (Zuranolon oder Placebo) erhalten haben.

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Der Endpunkt C-SSRS wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben.

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A,B,C
Es liegen keine Anzeichen für sonstige Aspekte vor, die zu Verzerrung führen können.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Der Endpunkt C-SSRS wird, wie im Studienprotokoll und statistischem Analyseplan präspezifiziert erhoben. Eine ereignisunabhängige Berichterstattung ist gewährleistet. Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Endpunkt: Unerwünschte Ereignisse

5. Verblindung der Endpunkterheber

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B, C
Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. Die Randomisierungspläne und IRT werden von einer unabhängigen Statistikerin/einem unabhängigen Statistiker erstellt und streng vertraulich behandelt. Während der Studie durfte die Verblindung von dem Prüfer/ der Prüferin über das IRT-System nur dann aufgehoben werden, wenn die Sicherheit einer Teilnehmerin gefährdet war und der Behandlungsplan von der erhaltenen Studienmedikation abhing.

6. Adäquate Umsetzung des ITT-Prinzips

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
Die Analyse des Endpunkts basiert auf der Safety-Population, die randomisierte Patientinnen umfasst die mindestens eine Dosis der Studienmedikation (Zuranolon oder Placebo) erhalten haben.

7. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

☒ **ja** ☐ **unklar** ☐ **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A,B,C
In den Studien ROBIN und SKYLARK erfolgte die Erhebung jeglicher Unerwünschten Ereignisse der Patientinnen objektiv, unabhängig davon, ob das Ereignis in kausalem Zusammenhang mit der Verabreichung der Studienmedikation stand. Die Erhebung und Klassifikation der UE erfolgten auf Basis allgemein anerkannter internationaler Standards. Somit ist eine ergebnisunabhängige Erhebung gewährleistet.

8. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

☒ **ja** ☐ **nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A,B,C
Es liegen keine Anzeichen für sonstige Aspekte vor, die zu Verzerrung führen können.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

☒ **niedrig** ☐ **hoch**

Begründung für die Einstufung:

Bei den Studien ROBIN und SKYLARK handelt es sich um doppelt verblindete RCTs. Entsprechend waren die Patientinnen, die Kliniker*innen, behandelnde Ärzt*innen und das Studienteam verblindet gegenüber der Behandlungszuweisung. In den Studien ROBIN und SKYLARK erfolgte die Erhebung jeglicher Unerwünschten Ereignisse der Patientinnen objektiv, unabhängig davon, ob das Ereignis in kausalem Zusammenhang mit der Verabreichung der Studienmedikation stand.
Die Erhebung und Klassifikation der UE erfolgten auf Basis allgemein anerkannter internationaler Standards. Somit ist eine ergebnisunabhängige Erhebung gewährleistet.
Es konnten keine anderen Aspekte identifiziert werden, die eine Verzerrung des Endpunkts bedingen würden. Daher wird das Verzerrungspotenzial auf Endpunktebene als niedrig eingestuft.

Hinweis: Der nachfolgend dargestellte Bewertungsbogen mit Ausfüllhinweisen dient nur als Ausfüllhilfe für den Blankobogen. Er soll nicht als Vorlage verwendet werden.

Bewertungsbogen zur Beschreibung von Verzerrungsaspekten (Ausfüllhilfe)

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A Verzerrungsaspekte 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 Anzeichen 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 kodierten 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 Anzeichen 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 behandelnden Personen

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:

behandelnde bzw. weiterbehandelnde Personen:

☐ **ja:** Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, die primär behandelnde Person (z. B. 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 behandelnden Personen 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 Anzeichen für eine ergebnisgesteuerte Darstellung eines Endpunkts zu Verzerrungen der Ergebnisse der übrigen Endpunkte führen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anzeichen dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, sind Verzerrungen für die anderen Endpunkte möglich. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Verzerrung 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 dies nicht zur Verzerrung anderer Endpunkte führt.

☐ **ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐ **unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐ **nein:** Es liegen Anzeichen 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. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrung führen können

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.*
  - *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.

Begründung für die Einstufung:

B Verzerrungsaspekte auf Endpunktebene 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**

Kommen in einer Studie Patienten vor, die die Studie entweder vorzeitig abgebrochen haben oder wegen Protokollverletzung ganz oder teilweise aus der Analyse ausgeschlossen wurden, so sind diese ausreichend genau zu beschreiben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen zu berücksichtigen (i. d. R. ITT-Analyse, siehe Äquivalenzstudien). Bei einer ITT(„intention to treat")-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen 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 Anzeichen für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

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.

Begründung für die Einstufung:

Anhang 4-G: Ergebnisse der Meta-Analyse

Die Daten werden in einem separaten Anhang vorgelegt.

Anhang 4-H: Ergebnisse der Studie ROBIN

Die Daten werden in einem separaten Anhang vorgelegt.

Anhang 4-I: Ergebnisse der Studie SKYLARK

Die Daten werden in einem separaten Anhang vorgelegt.