Dokumentvorlage, Version vom 20.01.2011

Dossier zur Nutzenbewertung gemäß § 35a SGB V

Pertuzumab (Perjeta®)

Roche Pharma AG

Modul 4 A

Brustkrebs

Medizinischer Nutzen und medizinischer Zusatznutzen, Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen

Stand: 04.03.2013

Inhaltsverzeichnis

Seite

Tabellenverzeichnis **4**
Abbildungsverzeichnis **10**
Abkürzungsverzeichnis **11**
4 Modul 4 – allgemeine Informationen **15**
4.1 Zusammenfassung der Inhalte von Modul 4 16
4.2 Methodik 23
4.2.1 Fragestellung 23
4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung 24
4.2.3 Informationsbeschaffung 28
4.2.3.1 Studien des pharmazeutischen Unternehmers 28
4.2.3.2 Bibliografische Literaturrecherche 29
4.2.3.3 Suche in Studienregistern 30
4.2.3.4 Selektion relevanter Studien 30
4.2.4 Bewertung der Aussagekraft der Nachweise 31
4.2.5 Informationssynthese und -analyse 35
4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien 35
4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien 35
4.2.5.3 Meta-Analysen 39
4.2.5.4 Sensitivitätsanalysen 40
4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren 41
4.2.5.6 Indirekte Vergleiche 47
4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen 49
4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel 49
4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel 49
4.3.1.1.1 Studien des pharmazeutischen Unternehmers 49
4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche 51
4.3.1.1.3 Studien aus der Suche in Studienregistern 53
4.3.1.1.4 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel 54
4.3.1.2 Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel 55
4.3.1.2.1 Studiendesign und Studienpopulationen 55
4.3.1.2.2 Verzerrungspotenzial auf Studienebene 66
4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien 67
4.3.1.3.1 Endpunkt – RCT 68
4.3.1.3.1.1 <Mortalität> – RCT 70
4.3.1.3.1.2 <Morbidität> – RCT 75
4.3.1.3.1.3 <Lebensqualität> – RCT 88
4.3.1.3.1.4 <Sicherheit> – RCT 95

4.3.1.3.2 Subgruppenanalysen – RCT ... 126
4.3.1.3.2.1 <Mortalität> – RCT ... 137
4.3.1.3.2.2 <Morbidität> – RCT ... 140
4.3.1.3.2.3 <Lebensqualität> – RCT ... 147
4.3.1.3.2.4 <Sicherheit> - RCT ... 152
4.3.1.3.3 Zusammenfassung der Ergebnisse aus randomisierten kontrollierten Studien ... 159
4.3.2 Weitere Unterlagen ... 164
4.3.2.1 Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien ... 164
4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche ... 164
4.3.2.1.2 Charakteristika der Studien für indirekte Vergleiche ... 164
4.3.2.1.3 Ergebnisse aus indirekten Vergleichen ... 165
4.3.2.1.3.1 <Endpunkt xxx> – indirekte Vergleiche aus RCT ... 165
4.3.2.1.3.2 Subgruppenanalysen – indirekte Vergleiche aus RCT ... 167
4.3.2.2 Nicht randomisierte vergleichende Studien ... 168
4.3.2.2.1 Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien ... 168
4.3.2.2.2 Charakteristika der nicht randomisierten vergleichenden Studien ... 168
4.3.2.2.3 Ergebnisse aus nicht randomisierten vergleichenden Studien ... 169
4.3.2.2.3.1 <Endpunkt xxx> – nicht randomisierte vergleichende Studien ... 169
4.3.2.2.3.2 Subgruppenanalysen – nicht randomisierte vergleichende Studien ... 170
4.3.2.3 Weitere Untersuchungen ... 170
4.3.2.3.1 Ergebnis der Informationsbeschaffung – weitere Untersuchungen ... 171
4.3.2.3.2 Charakteristika der weiteren Untersuchungen ... 171
4.3.2.3.3 Ergebnisse aus weiteren Untersuchungen ... 171
4.3.2.3.3.1 <Endpunkt xxx> – weitere Untersuchungen ... 171
4.3.2.3.3.2 Subgruppenanalysen – weitere Untersuchungen ... 172
4.3.2.4 Zusammenfassung der Ergebnisse aus weiteren Unterlagen ... 172
4.4 Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens ... 173
4.4.1 Beurteilung der Aussagekraft der Nachweise ... 173
4.4.2 Fazit ... 177
4.4.3 Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß ... 177
4.4.4 Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht ... 183
4.4.5 Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht – Arzneimittel zur Behandlung eines seltenen Leidens ... 184
4.5 Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte ... 185
4.5.1 Begründung für die Vorlage indirekter Vergleiche ... 185
4.5.2 Begründung für die Vorlage nichtrandomisierter vergleichender Studien und weiterer Untersuchungen ... 185
4.5.3 Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen ... 185
4.5.4 Verwendung von Surrogatendpunkten ... 185
4.6 Liste der eingeschlossenen Studien ... 187
4.7 Referenzliste ... 188

Anhang 4-A : Suchstrategien – bibliografische Literaturrecherche 196
Anhang 4-B : Suchstrategien – Suche in Studienregistern..... 198
Anhang 4-C : Liste der im Volltext gesichteten und ausgeschlossenen Studien mit Ausschlussgrund 200
Anhang 4-D : Liste der abgebrochenen Studien 201
Anhang 4-E : Liste der laufenden Studien..... 202
Anhang 4-F : Methodik der eingeschlossenen Studien – RCT..... 203
Anhang 4-G : Bewertungsbögen zur Einschätzung von Verzerrungsaspekten 225

Tabellenverzeichnis

Seite

Tabelle 4-1: Ein- und Ausschlusskriterien: RCT mit dem zu bewertenden Arzneimittel 17

Tabelle 4-2: Ein- und Ausschlusskriterien: RCT mit dem zu bewertenden Arzneimittel 25

Tabelle 4-3: Anforderungen an die Aussagen zur Beleglage:..... 34

Tabelle 4-4: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel 50

Tabelle 4-5: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel..... 51

Tabelle 4-6: Relevante Studien aus der Suche in Studienregistern – RCT mit dem zu bewertenden Arzneimittel 53

Tabelle 4-7: Studienpool – RCT mit dem zu bewertenden Arzneimittel..... 54

Tabelle 4-8: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel 56

Tabelle 4-9: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel aus CLEOPATRA (WO20698) 58

Tabelle 4-10: Charakterisierung der Studienpopulation: Baseline Charakteristika der Patienten in der Studie nach Behandlungsgruppe– RCT mit dem zu bewertenden Arzneimittel aus CLEOPATRA (WO20698) 59

Tabelle 4-11: Charakterisierung der Studienpopulation – Krankheitscharakteristika: RCT mit dem zu bewertenden Arzneimittel aus CLEOPATRA (WO20698)..... 62

Tabelle 4-12: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel..... 66

Tabelle 4-13: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel 67

Tabelle 4-14: Operationalisierung von <Endpunkt Gesamtüberleben>..... 70

Tabelle 4-15: Bewertung des Verzerrungspotenzials für <Endpunkt Gesamtüberleben> in RCT mit dem zu bewertenden Arzneimittel (finale konfirmatorische Analyse. 2. Datenschnitt Mai 2012)..... 70

Tabelle 4-16: Ergebnisse für<Endpunkt Gesamtüberleben> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (finale konfirmatorische Analyse. 2. Datenschnitt Mai 2012) 71

Tabelle 4-17: Operationalisierung von <Primärer Endpunkt Progressionsfreies Überleben bewertet durch IRF>..... 75

Tabelle 4-18: Bewertung des Verzerrungspotenzials für <Primärer Endpunkt Progressionsfreies Überleben bewertet durch IRF> in RCT mit dem zu bewertenden Arzneimittel (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011)..... 77

Tabelle 4-19: Ergebnisse für<Primärer Endpunkt Progressionsfreies Überleben bewertet durch IRF> während der Behandlungszeit aus randomisierter, kontrollierter Studie

CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011) 78

Tabelle 4-20: Ergebnisse für Sensitivitätsanalysen zum Primären Endpunkt Progressionsfreies Überleben bewertet durch IRF während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011) 81

Tabelle 4-21: Operationalisierung von <Endpunkt Gesamtansprechrate> 83

Tabelle 4-22: Bewertung des Verzerrungspotenzials für <Endpunkt Gesamtansprechrate>in RCT mit dem zu bewertenden Arzneimittel (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011) 84

Tabelle 4-23: Ergebnisse für<Endpunkt Gesamtansprechrate> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011) 86

Tabelle 4-24: Operationalisierung von <Endpunkt Zeit bis Symptomprogression> 88

Tabelle 4-25: Bewertung des Verzerrungspotenzials für <Endpunkt Zeit bis Symptomprogression>in RCT mit dem zu bewertenden Arzneimittel (1. Datenschnitt Mai 2011) 88

Tabelle 4-26: Ergebnisse für <Endpunkt Zeit bis Symptomprogression> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (1. Datenschnitt Mai 2011) 91

Tabelle 4-27: Operationalisierung von <Endpunkt Zeit bis zur Verschlechterung des BCS-Score> 92

Tabelle 4-28: Bewertung des Verzerrungspotenzials für <Endpunkt Zeit bis zur Verschlechterung des BCS-Score> in RCT mit dem zu bewertenden Arzneimittel (Erster Datenschnitt Mai 2011) 93

Tabelle 4-29: Ergebnisse für <Endpunkt Zeit bis zur Verschlechterung des BCS-Score> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (1. Datenschnitt Mai 2011) 94

Tabelle 4-30: Operationalisierung von <Endpunkt Patienten mit unerwünschtem Ereignis> 95

Tabelle 4-31: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit unerwünschtem Ereignis> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012) 95

Tabelle 4-32: Ergebnisse für <Endpunkt Patienten mit unerwünschtem Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 97

Tabelle 4-33: Operationalisierung von <Endpunkt Patienten mit schwerwiegendem unerwünschten Ereignis 98

Tabelle 4-34: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit schwerwiegendem unerwünschten Ereignis> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012) 98

Tabelle 4-35: Ergebnisse für <Endpunkt Patienten mit schwerwiegendem unerwünschten Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 100

Tabelle 4-36: Operationalisierung von <Endpunkt Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis> 102

Tabelle 4-37: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012) 102

Tabelle 4-38: Ergebnisse für <Endpunkt Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 103

Tabelle 4-39: Operationalisierung von <Endpunkt Patienten mit unerwünschtem Ereignis ≥ Grad 3> 104

Tabelle 4-40: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit unerwünschtem Ereignis ≥ Grad 3> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012) 104

Tabelle 4-41: Ergebnisse für<Endpunkt Patienten mit unerwünschtem Ereignis ≥ Grad 3> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 106

Tabelle 4-42: Operationalisierung von <Endpunkt Patienten mit UE Grad 3> 107

Tabelle 4-43: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit UE Grad 3> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012) 107

Tabelle 4-44: Ergebnisse für<Endpunkt Patienten mit UE Grad 3> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 109

Tabelle 4-45: Operationalisierung von <Endpunkt Patienten mit UE Grad 4> 110

Tabelle 4-46: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit UE Grad 4> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012) 110

Tabelle 4-47: Ergebnisse für<Endpunkt Patienten mit UE Grad 4> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 112

Tabelle 4-48: Operationalisierung von <Endpunkt Patienten mit UE Grad 5> 113

Tabelle 4-49: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit UE Grad 5> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012) 113

Tabelle 4-50: Ergebnisse für<Endpunkt Patienten mit UE Grad 5> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 115

Tabelle 4-51: Operationalisierung von <Endpunkt Patienten mit UE von besonderem Interesse> 116

Tabelle 4-52: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit UE von besonderem Interesse> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012) 118

Tabelle 4-53: Zusammenfassung für Patienten mit UE von besonderem Interesse während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 119

Tabelle 4-54: Ergebnisse für<Patienten mit UE von besonderem Interesse: linksventrikuläre Dysfunktion> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region 124

Tabelle 4-55: Anzahl der post-hoc durchgeführten Interaktionstests 127

Tabelle 4-56: Interaktionstest: zu erwartende Zufallsbefunde bei verschiedenen, gegebenen Signifikanzniveaus 128

Tabelle 4-57: Subgruppen der Patienten mit viszeralen Metastasen *Viszeral* versus Patienten mit nicht-viszeralen Metastasen *Nicht Viszeral* für das Gesamtüberleben 129

Tabelle 4-58: Ergebnis der Simulation des OS: Einfluss zusätzlicher Ereignisse auf die geschätzten HR, Ergebnissicherheit 129

Tabelle 4-59: Demografische Charakteristika und Krankheitsmerkmale zu Studienbeginn bei Patienten mit nicht-viszeraler Erkrankung vs. Patienten mit viszeraler Erkrankung 132

Tabelle 4-60: Wichtige demografische Charakteristika und Krankheitsmerkmale zu Studienbeginn in den Behandlungsarmen (nach viszeraler / nicht-viszeraler Erkrankung) .. 134

Tabelle 4-61: Subgruppenergebnisse für<Endpunkt Gesamtüberleben> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (finale konfirmatorische Analyse. 2. Datenschnitt Mai 2012) 137

Tabelle 4-62: Subgruppenergebnisse für<Primärer Endpunkt Progressionsfreie Überleben bewertet durch IRF> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (finale konfirmatorische Analyse, 1. Datenschnitt Mai 2011) 140

Tabelle 4-63: Subgruppenergebnisse für<Endpunkt Gesamtansprechrate> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (finale konfirmatorische Analyse, 1. Datenschnitt Mai 2011) 144

Tabelle 4-64: Subgruppenergebnisse für<Endpunkt Zeit bis Symptomprogression (Gesundheitsbezogene Lebensqualität)> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (1. Datenschnitt Mai 2011) 147

Tabelle 4-65: Subgruppenergebnisse für<Endpunkt Zeit bis zur Verschlechterung (Gesundheitsbezogene Lebensqualität)> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (1. Datenschnitt Mai 2011) 150

Tabelle 4-66: Subgruppenergebnisse für<Endpunkt Patienten mit unerwünschtem Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 152

Tabelle 4-67: Subgruppenergebnisse für<Endpunkt Patienten mit schwerwiegendem unerwünschten Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 153

Tabelle 4-68: Subgruppenergebnisse für<Endpunkt Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 154

Tabelle 4-69: Subgruppenergebnisse für<Endpunkt Patienten mit UE ≥Grad 3> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 155

Tabelle 4-70: Subgruppenergebnisse für<Endpunkt Patienten mit UE Grad 3> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 156

Tabelle 4-71: Subgruppenergebnisse für<Endpunkt Patienten mit UE Grad 4> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 157

Tabelle 4-72: Subgruppenergebnisse für<Endpunkt Patienten mit UE Grad 5> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012) 158

Tabelle 4-73: Zusammenfassung der Ergebnisse aus aktiv kontrollierter RCT mit dem zu bewertenden Arzneimittel aus CLEOPATRA (WO20698) 162

Tabelle 4-74: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche 165

Tabelle 4-75: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden 165

Tabelle 4-76: Operationalisierung von <Endpunkt xxx> 166

Tabelle 4-77: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche 166

Tabelle 4-78: Ergebnisse für <Endpunkt xxx> aus RCT für indirekte Vergleiche..... 167

Tabelle 4-79: Verzerrungsaspekte auf Studienebene – nichtrandomisierte vergleichende Interventionsstudien 169

Tabelle 4-80: Operationalisierung von <Endpunkt xxx>..... 169

Tabelle 4-81: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien 170

Tabelle 4-82: Operationalisierung von <Endpunkt xxx>..... 171

Tabelle 4-83: Ausmaß des Zusatznutzens auf Endpunktebene aus der RCT CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel auf Ebene der Studienpopulation 178

Tabelle 4-84: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens..... 183

Tabelle 4-85: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens –Arzneimittel zur Behandlung eines seltenen Leidens 184

Tabelle 4-86 (Anhang): Studienpool, abgebrochene Studien, zu bewertendes Arzneimittel 201

Tabelle 4-87 (Anhang): Studienpool, laufende Studien, zu bewertendes Arzneimittel..... 202

Tabelle 4-88 (Anhang): Studiendesign und -methodik für die CLEOPATRA-Studie (WO20698C/TOC4129g) (Datenquellen: CLEOPATRA (WO20698). Clinical Study Report –WO20698C/TOC4129g - A Phase III, Randomized, Double-Blind, Placebo-Controlled Clinical Trial to Evaluate the Efficacy and Safety of Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel in Previously Untreated HER2-Positive Metastatic Breast Cancer – Report No. 1046288, Roche, October 2011. **Und** CLEOPATRA (WO20698). Update Clinical Study Report - WO20698C/TOC4129g - A Phase III, Randomized, Double-Blind, Placebo-Controlled Clinical Trial to Evaluate the Efficacy and Safety of Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel in Previously Untreated HER2-Positive Metastatic Breast Cancer - Report No. 1053649, Roche, Dezember 2012.)..... 203

Tabelle 4-89 (Anhang): Erhebungszeitpunkte für die einzelnen Befunde während des Verlaufs der Studie CLEOPATRA (Studienbericht der Studie CLEOPATRA, 2011) 222

Tabelle 4-90 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie <CLEOPATRA (WO20698)> 225

Abbildungsverzeichnis

Seite

Abbildung 1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten, kontrollierten Studien mit dem zu bewertenden Arzneimittel 52

Abbildung 2: Studiendesign der Studie CLEOPATRA: Behandlungsarme und Bewertungszeitpunkte [46] 64

Abbildung 3: Kaplan-Meier Kurve zum Gesamtüberleben (ITT Population; CSR 12/2012 S.45) 74

Abbildung 4: Patientenfluss in der Studie, zweiter Datenschnitt..... 221

Abbildung 5: Angabe von Gründen für Patienten, die die Studie beendet hatten bis Mai 2012 221

Abkürzungsverzeichnis

Abkürzung	Bedeutung
AEGT	Adverse Events Group Terms
ALT (GPT)	Alanin-Aminotransferase (Glutamat-Pyruvat-Transaminase)
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
aPTT	activated Partial Thromboplastin Time
AST (GOT)	Aspartat-Aminotransferase (Glutamat-Oxalacetat-Transaminase)
BC	Breast Cancer
BCS	Breast Cancer Scale
CCTR	Cochrane Central Register of Controlled Trials
CHF	Congestive Heart Failure
CHMP	Committee for Medicinal Products for Human Use
CMH	Cochran-Mantel-Haenszel
CONSORT	Consolidated Standards of Reporting Trials
CPMP	Committee for Proprietary Medicinal Products
CR	Complete Response (Komplettes Ansprechen)
CRC	Cardiac Review Committee
CRF	Case Report Form
CSR	Clinical Study Report
CT	Computer Tomography
CTCAE	Common Terminology Criteria for Adverse Events
D	Docetaxel
DGHO	Deutsche Gesellschaft für Hämatologie und Onkologie e. V.
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information
DKG	Deutsche Krebsgesellschaft
DMC	Data Management Committee
E	Einschlusskriterium
ECHO	Echokardiogramm
ECOG	Eastern Cooperative Oncology Group
eCRF	Electronic Case Report Form
EGFR	Epidermal Growth Factor Receptor
EMA	European Medicines Agency

ER	Estrogen (Östrogen) Rezeptor
EU	Europäische Union
FACT	Functional Assessment of Cancer Therapy
FACT-B	Functional Assessment of Cancer Therapy-Breast Cancer
FDA	Food and Drug Administration
FISH	Fluoreszenz-InSitu-Hybridisierung
G-BA	Gemeinsamer Bundesausschuss
GCP	Good Clinical Practice
HER2	Human Epidermal Growth Factor Receptor 2
HR	Hazard Rate
HRQoL	Health Related Quality of Life
ICH	International Conference on Harmonisation of technical requirements for registration of pharmaceuticals for human use
IHC (2+ / 3+)	ImmunoHistoChemistry (2+ / 3+)
ICTRP	International Clinical Trials Registry Platform
INR	Internationale normalisierte Ratio
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IRF	Independent Review Facility
ITT	Intent-to-Treat
IVRS	Interactive Voice Response System
kg	Kilogramm
KI	Konfidenzintervall
KM	Kaplan-Meier
KOF	Körperoberfläche
LD	Longest Diameter
LVEF	Left Ventricular Ejection Fraction (Linksventrikuläre Auswurffraktion)
LVSD	Left Ventricular Systolic Dysfunction (Linksventrikuläre systolische Fehlfunktion)
m^2	Quadratmeter
MedDRA	Medical Dictionary for Regulatory Activities
mg	Milligramm
MTC	Mixed Treatment Comparison
MUGA	Multiple Gated Acquisition Scan

N.A.	Not Applicable
NACT	Next-Line Anti-Cancer Therapy
NCI	National Cancer Institute
NMR	Nuclear Magnetic Resonance
NYHA	New York Heart Association
ORR	Overall Response Rate (Gesamtansprechrate)
OS	Overall Survival (Gesamtüberleben)
pCR	Pathologic Complete Response
PD	Progressive Disease (Krankheitsprogression)
PFS	Progression-Free Survival (Progressionsfreies Überleben)
PgR	Progesteron
Pla	Placebo
PR	Partial Response (partielles Ansprechen)
PT	Preferred Term
PTT	Partielle Thromboplastinzeit
Ptz	Pertuzumab
q3w	every three weeks
QoL	Quality of Life
RCT	Randomized Controlled Trial
RECIST	Response Evaluation Criteria in Solid Tumors
RMP	Risk Management Plan
RR	Relatives Risiko
SAE	Serious Adverse Event
SAP	Safety-Analysis-Population
SAS	Statistical Analysis System
SD	Stable Disease (stabile Erkrankung)
SGB	Sozialgesetzbuch
SMQ	Standard MedDRA Queries
SOC	System Organ Class
STE	Surrogate Threshold Effects
STROBE	Strengthening The Reporting of Observational Studies in Epidemiology
SUE	Schwerwiegendes unerwünschtes Ereignis
T	Trastuzumab

T-DM1	Trastuzumab Emtansine
TOI-PFB	Trial Outcome Index—Physical / Functional / Breast
TREND	Transparent Reporting of Evaluations with Non-Randomized Design
UE	Unerwünschtes Ereignis
ULN	Upper Limit Normal
USA	United States of America
VerfO	Verfahrensordnung
vs.	versus
WHO	World Health Organization

4 Modul 4 – allgemeine Informationen

Modul 4 enthält folgende Angaben:

– Zusammenfassung (Abschnitt 4.1)

– Angaben zur Methodik der im Dossier präsentierten Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens (Abschnitt 4.2)

– Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen (Abschnitt 4.3)

– eine abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens, einschließlich der Angabe von Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht (Abschnitt 4.4)

– ergänzende Informationen zur Begründung der vorgelegten Unterlagen (Abschnitt 4.5)

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

Für Arzneimittel, die zur Behandlung eines seltenen Leidens nach der Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden zugelassen sind, müssen keine Nachweise zum medizinischen Nutzen und medizinischen Zusatznutzen vorgelegt werden (Abschnitt 4.3), solange der Umsatz des Arzneimittels mit der gesetzlichen Krankenversicherung zu Apothekenverkaufspreisen einschließlich Umsatzsteuer in den letzten 12 Monaten einen Betrag von 50 Millionen Euro nicht übersteigt. Angaben zu Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, müssen jedoch vorgelegt werden. Zu diesem Zweck enthält die Dokumentvorlage den Abschnitt 4.4.5, der ausschließlich in Dossiers für die oben genannten Arzneimittel zur Behandlung eines seltenen Leidens bearbeitet werden soll. In diesem Abschnitt sind Angaben zu Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich der begründenden Nachweise vorzulegen. Weitere Abschnitte von Modul 4 können dabei nach Bedarf bearbeitet werden.

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

4.1 Zusammenfassung der Inhalte von Modul 4

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

Fragestellung

Ist für Pertuzumab (Intervention) im Anwendungsgebiet der Behandlung erwachsener Patienten mit HER2-positivem metastasiertem oder lokal rezidivierendem, inoperablem Brustkrebs, die zuvor noch keine anti-HER2-Therapie oder Chemotherapie zur Behandlung ihrer metastasierten Erkrankung erhalten hatten (Patientenpopulation), basierend auf den Daten aus randomisierten kontrollierten klinischen Studien (Studientypen) im Vergleich zu Trastuzumab in Kombination mit einem Taxan (Paclitaxel, Docetaxel) (zweckmäßige Vergleichstherapie) ein Zusatznutzen unter Berücksichtigung des Gesamtüberlebens (Mortalität), Progressionsfreie Überlebens bewertet durch IRF, Gesamtansprechrate (Morbidität), Lebensqualität (Zeit bis Symptomprogression, Zeit bis zur Verschlechterung des BCS-Score), Patienten mit unerwünschen Ereignissen (UE), Patienten mit schwerwiegenden unerwünschten Ereignissen (SUE), Patienten mit Behandlungsabbruch wegen UE, Patienten mit UE ≥ Grad 3, Patienten mit UE Grad 3, Patienten mit UE Grad 4, Patienten mit UE Grad 5 sowie Patienten mit UE von besonderem Interesse (patientenrelevante Endpunkte) nachweisbar?

Datenquellen

Die Bewertung wurde auf Grundlage einer randomisierten, kontrollierten klinischen Studie vorgenommen, für welche das Studienprotokoll, die Studienberichte für zwei Auswertungszeitpunkte (Mai 2011 und Mai 2012) sowie weitere Dokumente für zusätzliche Auswertungen, die entsprechend der Vorgaben aus dem G-BA Beratungsgespräch durchgeführt werden sollten, vorlagen. Zur Identifikation relevanter Studien wurden alle Studien zu Pertuzumab im vorliegenden Anwendungsgebiet gelistet, die an die Zulassungsbehörden übermittelt wurden bzw. für die Roche regulatorischer Sponsor beziehungsweise auf andere Weise finanziell beteiligt ist oder war. Ferner wurde eine Suche in den Studienregistern clinicaltrials.gov (http://www.clinicaltrials.gov/) der International Clinical Trials Registry Platform Search Portal (ICTRP Search Portal, Suchportal der WHO, http://apps.who.int/trialsearch/) sowie firmenintern in der Roche- Datenbank „Clinical Trial Protocol Registry“ durchgeführt. Die Selektion der für diese Untersuchung relevanten Studien entsprechend der aufgeführten Ein- und Ausschlusskriterien wurde von zwei Reviewern unabhängig voneinander vorgenommen.

Ein- / Ausschlusskriterien für Studien

Tabelle 4-1: Ein- und Ausschlusskriterien: RCT mit dem zu bewertenden Arzneimittel

E1	Patientenpopulation (Indikation)	Erwachsene Patienten mit HER2-positivem metastasiertem oder lokal rezidivierendem, inoperablem Brustkrebs, die zuvor noch keine anti-HER2-Therapie oder Chemotherapie zur Behandlung ihrer metastasierten Erkrankung erhalten haben.
E2	Intervention	Pertuzumab lt. Zulassung in Kombination mit Trastuzumab und Docetaxel.
E3	Vergleichstherapie	Trastuzumab in Kombination mit einem Taxan (Paclitaxel, **Docetaxel**) (zweckmäßige Vergleichstherapie entsprechend Modul 3.1 und Ergebnis der Beratung mit dem G-BA am 14.6.2012) – entspricht therapeutischem Standard (*Standard of care)*
E4	Endpunkte	Mortalität: · Gesamtüberleben (OS) Morbidität: · Progressionsfreies Überleben (PFS), bewertet durch IRF · Gesamtansprechrate (ORR) Lebensqualität: · Zeit bis Symptomprogression · Zeit bis zur Verschlechterung des BCS-Score Sicherheit: · Alle unerwünschten Ereignisse (UE) · Schwerwiegende unerwünschte Ereignisse (SUE) · Behandlungsabbruch wegen UE · UE ≥ Grad 3 · UE Grad 3 · UE Grad 4 · UE Grad 5 · UE von besonderem Interesse
E5	Studientyp (Design)	Randomisierte, kontrollierte Studien.
E6	Studiendauer	Keine Einschränkungen, da diese bei onkologischen Indikationen in Abhängigkeit des Auftretens der Überlebensereignisse bestimmt wird.
E7	Sprache	Keine Einschränkung.
E8	Vollpublikation verfügbar	Ja (Studienbericht).

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

Für die in die Nutzenbewertung eingeschlossenen Studien wurde das Verzerrungspotential bewertet. Entsprechend der Vorgaben in der Verfahrensordnung des Gemeinsamen Bundesausschusses (G-BA) und dem Vorgehen des IQWiG erfolgte dies auf Studien- und Endpunktebene.

Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen

Mortalität

Die finale konfirmatorische Analyse **des Gesamtüberlebens (OS)** im Mai 2012 ergab ein HR von 0,66 (95 % KI [0,52; 0,84]) zugunsten von Ptz. Das mediane OS lag bei Pla+T+D bei 37,6 Monaten (154 Todesfälle; 37,93 %). Demgegenüber wurde die mediane Überlebenszeit unter Ptz+T+D (113 Fälle; 28,11 %) in der finalen Analyse nicht erreicht. Durch die Kombination aus Ptz+T+D wurde das Überleben im Vergleich zu Pla+T+D signifikant verlängert (p = 0,0008).

Morbidität

Die finale konfirmatorische Analyse des **progressionsfreies Überlebens (PFS bewertet durch IRF)** im Mai 2011 ergab ein HR von 0,62 (95 % KI [0,51; 0,75]), entsprechend einer signifikanten Verlängerung des PFS (Median) um 6,1 Monate unter Ptz+T+D (p < 0,0001). Insgesamt traten 433 PFS Ereignisse auf: 242 (59,6 %) unter Pla+T+D und 191 (47,5 %) unter Ptz+T+D.

Die **Gesamtansprechrate (bewertet durch das IRF)** wurde bei den Patienten der ITT-Population bewertet, die eine durch das IRF bestimmte, messbare Erkrankung zu Studienbeginn aufwiesen. Über 80 % der Patienten in jedem Arm waren in diese Analyse eingeschlossen (336 im Pla+T+D Behandlungsarm und 343 im Ptz+T+D Behandlungsarm). Die Gesamtansprechrate konnte von 69,3 % unter Pla+T+D auf 80,2 % unter Ptz+T+D signifikant gesteigert werden: RR = 1,16; 95 % KI [1,06; 1,26]; p = 0,0012.

Die Endpunkte IRF-bewertetes PFS, OS und Gesamtansprechrate stellen auch aus statistischer Sicht insofern wesentliche Endpunkte dar, als für diese drei Endpunkte im Protokoll eine Testhierarchie zur Adjustierung bzgl. multiplen Testens festgelegt worden war. Die statistische Auswertung zeigt nun für jeden dieser 3 wesentlichen Endpunkte eine signifikante Überlegenheit der Behandlung mit Ptz+T+D gegenüber Pla+T+D.

Lebensqualität

Hinsichtlich Symptomprogression – gemessen als Abnahme des **FACT-B TOI-PFB** von mindestens 5 Punkten – zeigte sich kein signifikanter Unterschied zwischen den Studienarmen.

Die Zeit bis zur Verschlechterung des **FACT-B BCS-Score** war unter Ptz+T+D signifikant verlängert: HR = 0,77; 95 % KI [0,64; 0,93]. Die mediane Zeit bis zur Verschlechterung lag bei 18,3 Wochen unter Pla+T+D vs. 26,7 Wochen unter Ptz+T+D (p = 0,0058).

Unerwünschte Ereignisse

Die Inzidenz **aller UE** lag bei Pla+T+D bei 98,74 % vs. 100,0 % der Patienten unter Ptz+T+D (RR: 1,01 [95 % KI: 1,00; 1,02]). Das Ergebnis der Kategorie alle UE bedeutet, dass ein 1 %-Unterschied zwischen den Gruppen besteht, der lediglich aufgrund der hohen Fallzahl signifikant wurde. Der Punktschätzer für das Relative Risiko liegt bei 1,012, die exakte untere Grenze des zweiseitigen 95%-Konfidenzintervalls liegt bei 1,0016, bereits 1 Patient mehr oder weniger in einem der Studienarme hätte dieses Ergebnis nicht signifikant werden lassen. Aus diesem Grund ist das Ergebnis als formal signifikant aber nicht medizinisch relevant zu betrachten.

Die Inzidenz der **SUE** lag bei Pla+T+D bei 29,0 % vs. 36,3 % der Patienten unter Ptz+T+D (RR 1,25; 95 % KI [1,02; 1,53]). Der Unterschied zwischen beiden Armen war hauptsächlich durch die höhere Inzidenz febriler Neutropenien unter Ptz+T+D bedingt, die eine Verlängerung des stationären Aufenthalts erforderten.

Die Inzidenz der **UE, die zum Behandlungsabbruch** führten, lag unter Pla+T+D bei 28,8 % (n = 114) gegenüber 30,6 % (n = 125) unter Ptz+T+D (RR 1,06; 95 % KI [0,86; 1,32]). Bei Ausschluss von Ereignissen, die zum Abbruch der Docetaxel-Therapie führten, lauten die Ergebnisse 6,1 % der Patienten unter Pla+T+D vs. 7,8 % unter Ptz+T+D.

Obwohl die meisten UE von Schweregrad 1 oder 2 waren, trat bei der Mehrzahl der Patienten mindestens ein **UE ≥ Grad 3** auf (bei n = 291, 73,5 % der Patienten unter Pla+T+D und bei n = 311, 76,2 % der Patienten im Pertuzumab-Arm). Somit war die Inzidenz der UE von Grad ≥ 3 zwischen den Behandlungsarmen ausgewogen (RR: 1,04 [95% KI: 0,96; 1,12].

Die Betrachtung der Inzidenzen der UE spezifisch in den einzelnen Schweregradgruppen (**Grad 3, Grad 4 und Grad 5**) zeigte keine statistisch signifikanten Unterschiede zwischen den Behandlungsarmen. Die Inzidenz der **UE vom Grad 5** lag unter Pla+T+D bei 3,03 % (n = 12) vs. 1,96 % (n = 8) unter Ptz+T+D (RR 0,65; 95 % KI [0,27; 1,57]).

Die Inzidenz der **Diarrhö** lag unter Pla+T+D bei 48,2 % vs. 68,1 % unter Ptz+T+D (RR 1,41; 95 % KI [1,25; 1,60]). Diarrhö vom Grad ≥ 3 trat bei 5,1 % (n = 20) der Patienten unter Pla+T+D und bei 9,1 % (n = 37) unter Ptz+T+D auf. Weder im Kontrollarm, noch im Pertuzumab-Therapiearm wurden fatale Verläufe beobachtet.

Die Inzidenz der **Exantheme** lag unter Pla+T+D bei 36,4 % vs. 47,5 % unter Ptz+T+D (RR 1,31; 95 % KI [1,11; 1,54]). Exantheme mit Schweregrad ≥ 3 waren sehr selten (n = 5 vs. n = 12). Das Exanthem wird derzeit nicht als bedeutsames identifiziertes Risiko klassifiziert und gilt nicht als medizinisch signifikant (RMP 2.1). Exantheme lassen sich mit Hydrocortison, Bethamethason, Clindamycin oder Hydroxyzin sehr gut behandeln.

Hinsichtlich **febriler Neutropenie** lag die Inzidenz bei 7,6 % (n = 30) unter Pla+T+D vs. 13,8 % (n = 56) unter Ptz+T+D: RR 1,81; 95 % KI [1,19; 2,76]. **Febrile neutropenische Infektionen** traten unter Pla+T+D mit einer Häufigkeit von n = 3 und n = 14 unter Ptz+T+D auf (RR = 4,53; 95 % KI [1,31; 15,64]).

Die Häufigkeit einer **Mukositis** lag unter Pla+T+D bei 37,9 % vs. 49,8 % unter Ptz+T+D: RR 1,31; 95 % KI [1,12, 1,54]. Der Anteil mit Mukositis Grad ≥ 3 war gering (n=8 vs. n=13). Das Auftreten einer Mukositis unter einer Brustkrebstherapie mit Pertuzumab in Kombination mit Trastuzumab und Docetaxel beeinflusste die Therapie nur unwesentlich und ließ sich durch Standardmaßnahmen im Rahmen einer Chemotherapie sehr gut beherrschen.

Hinsichtlich der weiteren unerwünschten Ereignisse von besonderem Interesse ergab sich zwischen den Studienarmen kein statistisch signifikanter Unterschied.

Durch das Hinzufügen von Pertuzumab zum im klinischen Alltag etablierten Trastuzumab in Kombination mit Docetaxel Therapieschemata veränderte sich das Sicherheitsprofil der Brustkrebstherapie kaum und ist somit in beiden Therapiearmen vergleichbar.

Subgruppen-Analysen

Insgesamt treten weniger signifikante Interaktionstests auf, als per Zufall erwartbar wären und die scheinbar auf eine Interaktion hindeuten. Für diese Ergebnisse, die formal in eine der beiden IQWiG-Kategorien für Interaktionstests fallen, sind auf Einzelvariablenebene in Kapitel 4.3.1.3.2 genannte Kriterien nicht erfüllt. Somit resultiert keine Differenzierung des Zusatznutzens über einzelne Subgruppen.

Schlussfolgerungen zum Zusatznutzen und zum therapeutisch bedeutsamen Zusatznutzen

Die Nachweise für den Nutzen und Zusatznutzen basieren auf der randomisierten, doppelblinden, klinischen Studie WO20698 (CLEOPATRA). Die Studie erfüllt alle Anforderungen, um bereits mit einer Studie die Ergebnissicherheit für einen Beleg zu erreichen (s.a. Kriterienkatalog CHMP Guideline). **Daher sind die Anforderungen an die Ergebnissicherheitskategorie „Beleg“ für einen Zusatznutzen erfüllt.**

Mortalität

Hinsichtlich des **Gesamtüberlebens** wurde durch die Zugabe von Ptz zu T+D das Sterberisiko signifikant um 34 % reduziert (HR 0,66; 95 % KI [0,53; 0,84]). Dabei wurde mit der aktuellen Standardtherapie T+D ein medianes OS von 37,6 Monaten erreicht. Der zu erwartende zusätzliche mediane Überlebensvorteil mit Ptz+T+D liegt zwischen 8 und 19,4 Monaten. Damit liegt bereits die erwartete Untergrenze des Überlebensvorteils in der Größenordnung, die durch die Gabe von T+D erzielt wurde. Diese Ergebnisse stellen einen erneuten Therapiedurchbruch bei der Behandlung von HER2-positiven Patienten dar. Somit wird eine erhebliche und bisher nicht erreichte therapierelevante Verlängerung der Überlebensdauer erzielt. Dies entspricht einem **erheblichen Zusatznutzen beim Gesamtüberleben.**

Morbidität

Das Risiko für Progression oder Tod (IRF-bewertetes PFS) verringerte sich mit Pertuzumab signifikant um 38 % (HR 0,62; 95 % KI [0,51; 0,75]). Das IRF-bewertete PFS betrug unter Pla+T+D 12,4 Monate. Unter zusätzlicher Gabe von Pertuzumab wurde dies signifikant um 6,1 Monate auf 18,5 Monate verlängert. Eine entsprechende Verlängerung des PFS auf 18,5 Monate durch die zusätzliche Gabe einer Therapie zu T+D wurde bisher für diese Patientenpopulation nicht erreicht. Dies entspricht einem **erheblichen Zusatznutzen beim PFS.**

Die Gesamtansprechrate (ORR) unter Pla+T+D lag bei 69,3 %. Mit der zusätzlichen Gabe von Pertuzumab konnte sie auf 80,2 % gesteigert werden. Die bereits sehr hohe ORR unter Pla+T+D konnte weiter verbessert werden. **Es besteht ein beträchtlicher Zusatznutzen hinsichtlich der Gesamtansprechrate.**

Lebensqualität

Für die Zeit bis zur Verschlechterung der in der BCS Subskala erfassten brustkrebsspezifischen Symptome zeigte sich, dass die Kombination Ptz+T+D in einem verzögerten Auftreten dieser Symptome resultierte: HR = 0,77; 95 % KI [0,64; 0,93]. Die mediane Zeit bis zur Verschlechterung lag bei 18,3 Wochen unter Pla+T+D vs. 26,7 Wochen unter Ptz+T+D. Durch die zusätzliche Gabe von Ptz zu T+D wird eine beträchtliche Verbesserung der Lebensqualität in Bezug auf brustkrebsspezifische Symptome erreicht. **Damit liegt ein beträchtlicher Zusatznutzen in der Dimension Lebensqualität vor.**

Unerwünschte Ereignisse

Eine ausschließlich auf der statistischen Analyse einzelner UE-Kategorien beruhende Beurteilung der UE führt aus medizinischer Sicht zu einem unvollständigen, eventuell sogar verzerrten Bild. Von daher erfolgt die Bewertung der UE aus statistischer als auch medizinischer Sicht.

Zwischen den beiden Therapiearmen zeigt sich insgesamt keine unterschiedliche Inzidenz **unerwünschter Ereignisse (UE)**. Auf die Therapiedauer bezogen, traten unter Pla+T+D 18,72 UE gegenüber 16,88 UE pro Patient und Jahr unter Ptz+T+D auf. **Das Sicherheitsprofil war auch aus medizinischer Sicht vergleichbar**.

Die Gesamtbetrachtung der SUE zeigt, dass unter Ptz+T+D zwar statistisch mehr SUE beobachtet wurden, diese in der Regel aber gut behandelbar waren. Somit war das Sicherheitsprofil von **Ptz+T+D hinsichtlich SUE aus medizinischer Sicht vergleichbar zu Pla+T+D**.

Die Werte für die Kategorie **„Behandlungsabbruch wegen UE“** zwischen beiden Armen waren statistisch nicht unterschiedlich. **Das Sicherheitsprofil war auch aus medizinischer Sicht vergleichbar**.

Die Inzidenz **UE ≥ Grad 3** war zwischen den Behandlungsarmen ausgewogen. Das Sicherheitsprofil von Ptz+T+D hinsichtlich der UE Grad ≥ 3 war **auch aus medizinischer Sicht vergleichbar zu Pla+T+D**.

Die Betrachtung der Inzidenzen der UE spezifisch in den einzelnen Schweregradgruppen (**Grad 3, Grad 4 und Grad 5**) zeigte keine statistisch signifikanten Unterschiede zwischen den Behandlungsarmen.

Todesfälle (UE Grad 5) waren in beiden Armen selten und **das Sicherheitsprofil war aus medizinischer Sicht vergleichbar**.

Aus medizinischer Sicht liegt **kein Unterschied des Sicherheitsprofils** von Ptz hinsichtlich **Diarrhö** vor, auch wenn deren Inzidenz statistisch signifikant erhöht ist. Gleiches gilt für die Exantheme und Mukositis.

Aufgrund der statistisch signifikant häufiger beobachteten **febrilen Neutropenien** und febrilen neutropenischen Infektionen liegt ein Signal vor. In der klinischen Gesamtbetrachtung im Vergleich zu Pla+T+D hat dieses Signal jedoch **auf das Nutzen-Risikoprofil von Ptz+T+D keinen Einfluss**.

Insgesamt traten nicht mehr und keine schwereren Nebenwirkungen auf, wenn Patienten mit HER2-positivem metastasierten oder lokal rezidivierenden inoperablen Brustkrebs zusätzlich mit Ptz behandelt wurden. In der **Gesamtbetrachtung ist das Sicherheitsprofil von Ptz+T+D vergleichbar mit dem Sicherheitsprofil von Pla+T+D.**

Zusammengefasst fordert die Definition gemäß AM-NutzenV für einen erheblichen Zusatznutzen eine nachhaltige und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte große Verbesserung des therapierelevanten Nutzens im Sinne einer erheblichen Verbesserung der Mortalität, Morbidität oder Lebensqualität.

Die zusätzliche Gabe von Pertuzumab zur Standardtherapie resultierte in einer signifikanten und klinisch relevanten erheblichen Verlängerung des **Gesamt- und des progressionsfreien Überlebens,** sowie in einer beträchtlichen Verbesserung der **Gesamtansprechrate**. Diese gehen einher mit einer beträchtlichen Verbesserung der brustkrebsspezifischen **Lebensqualität**. Damit liegt eine Verbesserung durch Pertuzumab in den Dimensionen Mortalität, Morbidität und Lebensqualität vor.

In der Gesamtbetrachtung zeigt sich unter der Therapie mit Pertuzumab in Kombination mit Trastuzumab und Docetaxel gegenüber der Standardtherapie ein **erheblicher 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.

4.2.1 Fragestellung

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

– Patientenpopulation
– Intervention
– Vergleichstherapie
– Endpunkte
– Studientypen

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

Ist für Pertuzumab (Intervention) im Anwendungsgebiet der Behandlung erwachsener Patienten mit HER2-positivem metastasiertem oder lokal rezidivierendem, inoperablem Brustkrebs, die zuvor noch keine anti-HER2-Therapie oder Chemotherapie zur Behandlung ihrer metastasierten Erkrankung erhalten hatten (Patientenpopulation), basierend auf den Daten aus randomisierten kontrollierten klinischen Studien (Studientypen) im Vergleich zu Trastuzumab in Kombination mit einem Taxan (Paclitaxel, Docetaxel) (zweckmäßige Vergleichstherapie) ein Zusatznutzen unter Berücksichtigung des Gesamtüberlebens (Mortalität), Progressionsfreies Überleben bewertet durch IRF, Gesamtansprechrate (Morbidität), Lebensqualität (Zeit bis zur Symptomprogression, Zeit bis zur Verschlechterung des BCS-Score), Patienten mit unerwünschten Ereignissen (UE), Patienten mit schwerwiegenden unerwünschten Ereignissen (SUE), Patienten mit Behandlungsabbruch aufgrund UE, Patienten mit UE Grad 3 und höher, Patienten mit UE Grad 3, Patienten mit UE Grad 4, Patienten mit UE Grad 5 sowie Patienten mit unerwünschten Ereignissen von besonderem Interesse (patientenrelevante Endpunkte) nachweisbar?

4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung

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

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

Die folgende Tabelle gibt die Ein- und Ausschlusskriterien für die Studienauswahl wieder. Nähere Erläuterungen finden sich unterhalb der Tabelle zu den einzelnen Punkten.

Tabelle 4-2: Ein- und Ausschlusskriterien: RCT mit dem zu bewertenden Arzneimittel

E1	Patientenpopulation (Indikation)	Erwachsene Patienten mit HER2-positivem metastasiertem oder lokal rezidivierendem, inoperablem Brustkrebs, die zuvor noch keine anti-HER2-Therapie oder Chemotherapie zur Behandlung ihrer metastasierten Erkrankung erhalten haben.
E2	Intervention	Pertuzumab lt. Zulassung in Kombination mit Trastuzumab und Docetaxel
E3	Vergleichstherapie	Trastuzumab in Kombination mit einem Taxan (Paclitaxel, **Docetaxel**) (zweckmäßige Vergleichstherapie entsprechend Modul 3.1 und Ergebnis der Beratung mit dem G-BA am 14.6.2012) – entspricht therapeutischem Standard (*Standard of care)*
E4	Endpunkte	Mortalität: · Gesamtüberleben (OS) Morbidität: · Progressionsfreies Überleben (PFS), bewertet durch IRF · Gesamtansprechrate (ORR) Lebensqualität: · Zeit bis Symptomprogression · Zeit bis zur Verschlechterung des BCS-Score Sicherheit: · Alle Unerwünschten Ereignisse (UE) · Schwerwiegende unerwünschte Ereignisse (SUE) · Behandlungsabbruch wegen UE · UE ≥ Grad 3 · UE Grad 3 · UE Grad 4 · UE Grad 5 · UE von besonderem Interesse
E5	Studientyp (Design)	Randomisierte, kontrollierte Studien.
E6	Studiendauer	Keine Einschränkungen, da diese bei onkologischen Indikationen in Abhängigkeit des Auftretens der Überlebensereignisse bestimmt wird.
E7	Sprache	Keine Einschränkung.
E8	Vollpublikation verfügbar	Ja (Studienbericht)

Population

Gemäß IQWiG ist das Einschlusskriterium bezüglich der Population („*Erwachsene Patienten mit HER2-positivem metastasiertem oder lokal rezidivierendem, inoperablem Brustkrebs, die zuvor noch keine anti-HER2-Therapie oder Chemotherapie zur Behandlung ihrer metastasierten Erkrankung erhalten haben*“) erfüllt, „wenn bei mindestens 80 % der in der Studie eingeschlossenen Patienten dieses Indikationsgebiet vorliegt.“ [1] Die Auswirkungen auf das Verzerrungspotential wurden ggf. im Rahmen der zugehörigen Bewertung auf Studienebene berücksichtigt.

Intervention und Vergleichstherapie

Es wurden alle Studien eingeschlossen, bei denen mindestens 80 % der Patienten das Einschlusskriterium bezüglich Prüfintervention (Interventionsgruppe der Studie) bzw. zu mindestens 80 % das Einschlusskriterium bezüglich Vergleichsintervention (Vergleichsgruppe der Studie) erfüllten. War in einer Studie eines der beiden Kriterien verletzt, wurde die Studie aus der Nutzenbewertung ausgeschlossen. [1]

Die zu prüfende Intervention ist Pertuzumab in Kombination mit Trastuzumab und Docetaxel bei erwachsenen Patienten mit HER2-positivem metastasiertem oder lokal rezidivierendem, inoperablem Brustkrebs, die zuvor noch keine anti-HER2-Therapie oder Chemotherapie zur Behandlung ihrer metastasierten Erkrankung erhalten haben. [2]

Ein Beratungsgespräch nach § 8 AM-NutzenV zum Thema „zweckmäßige Vergleichstherapie“ fand am 14. Juni 2012 in der Geschäftsstelle des Gemeinsamen Bundesausschusses (G-BA) in Berlin statt. [3] Vom Gemeinsamen Bundesausschuss wurde die zweckmäßige Vergleichstherapie wie folgt mitgeteilt:

„Als zweckmäßige Vergleichstherapie für Pertuzumab in Kombination mit Trastuzumab und Docetaxel zur Behandlung des HER2-positiven, metastasierten Brustkrebs hat der G-BA bestimmt: Trastuzumab in Kombination mit einem Taxan (Paclitaxel, Docetaxel).“

Als zweckmäßige Vergleichstherapie für Pertuzumab in Kombination mit Trastuzumab und Docetaxel zur Behandlung des HER2-positiven, lokal rezidivierenden, inoperablen Brustkrebses wurde seitens des G-BA „Strahlentherapie“ bestimmt. [4]

Für die Patientengruppe mit HER2-positivem, lokal rezidivierendem, inoperablem Brustkrebs ist aktuell keine weitere medikamentöse Therapie explizit zugelassen. Von daher kann dieses Kriterium bei der Bestimmung der ZVT für diese kleine Patientengruppe nicht zur Anwendung gebracht werden.

Die Strahlentherapie sowie auch die medikamentöse Therapie kann im Rahmen der GKV erbracht werden. Laut Anwendungsgebiet von Pertuzumab ist die Strahlentherapie bei Patienten mit HER2-positivem lokal rezidivierendem, inoperablem Brustkrebs als Vortherapie auch nicht ausgeschlossen, jedoch eine anti-HER2-Therapie oder Chemotherapie zur Behandlung ihrer metastasierten Erkrankung.

Bisher liegen keine Beschlüsse zum patientenrelevanten Nutzen der Strahlentherapie und auch medikamentösen Therapie bei Patienten mit HER2-positivem, lokal rezidivierendem inoperablem Brustkrebs vor.

Hinsichtlich der Evidenz stellte der G-BA fest: „Bei insgesamt limitierter Evidenz für Therapien in dieser Indikation ist die Strahlentherapie nach nationalen wie internationalen Leitlinien die Therapie der Wahl bei Inoperabilität eines lokalen Rezidivs…“. [4]

Roche stimmt dem G-BA zu, dass die verfügbare Evidenz für Therapien bei lokal rezidivierendem, inoperablem Brustkrebs sehr limitiert ist. Insbesondere für die die Therapie des HER2-positiven lokal rezidivierenden, inoperablen Brustkrebses liegt keine belastbare Evidenz für Strahlentherapie vor.

Die nationalen und internationalen Leitlinien empfehlen den Einsatz von palliativen medikamentösen Therapien. [5-9] Die Leitlinie der AGO und die NCCN Guideline empfehlen auch eine systemische Therapie mit Trastuzumab beim Rezidiv eines HER2 überexprimierenden nicht resezierbaren Tumors.

Eine Therapie mit Pertuzumab in Kombination mit Trastuzumab und Docetaxel beim HER2-positiven lokal rezidivierende, inoperablen Brustkrebs hat zum Ziel, nicht nur die lokale sondern auch die für das Überleben entscheidende systemische Erkrankung zu kontrollieren. Somit ist für diese Patientengruppe mit HER2-positivem lokal rezidivierendem, inoperablem Brustkrebs eine palliative Behandlung indiziert, die sich von der systemischen Therapie der Patienten mit HER2-positiver Metastasierung nicht unterscheidet. [10]

Aus Sicht der Roche ergibt sich anhand der in der VerfO definierten Kriterien zur Bestimmung der zweckmäßigen Vergleichstherapie Trastuzumab in Kombination mit Docetaxel als zweckmäßige Vergleichstherapie für die Therapie des HER2-positiven, metastasierten und auch des lokal rezidivierenden, inoperablen Brustkrebs, bei Patienten, die zuvor noch keine anti-HER2-Therapie oder Chemotherapie zur Behandlung ihrer metastasierten Erkrankung erhalten haben.

Endpunkte

Der Nachweis des Zusatznutzens soll auf Basis patientenrelevanter Endpunkte erfolgen, die laut Verfahrensordnung des Gemeinsamen Bundesausschuss im 5. Kapitel § 3 Abs. 1 VerfO der Morbidität, Mortalität bzw. der Lebensqualität und den unerwünschten Wirkungen zugeordnet werden können. [11] Studien, die für die Bewertung im Rahmen des vorliegenden Dossiers relevant waren, mussten mindestens einen der folgenden patientenrelevanten Endpunkte enthalten:

- Mortalität:
  - Gesamtüberleben (OS)
- Morbidität:
  - Progressionsfreies Überleben (PFS), bewertet durch IRF
  - Gesamtansprechrate (ORR)
- Lebensqualität:
  - Zeit bis Symptomprogression
  - Zeit bis zur Verschlechterung des BCS-Score
- Sicherheit:

- Alle Unerwünschten Ereignisse (UE)
- Schwerwiegende unerwünschte Ereignisse (SUE)
- Behandlungsabbruch wegen UE
- UE ≥ Grad 3
- UE Grad 3
- UE Grad 4
- UE Grad 5
- UE von besonderem Interesse

Studientyp

Im Rahmen von Therapiestudien wird randomisierten, kontrollierten Studien (RCTs) und systematischen Übersichten von RCTs der höchste Evidenzgrad zugeordnet (Evidenzklassifizierung siehe 5. Kapitel, § 5 Abs. 6 VerfO. [1;11]

Es wurden daher ausschließlich Studien mit diesem Design in die Nutzenbewertung eingeschlossen.

Studiendauer

Bezüglich der Studiendauer erfolgten keine Einschränkungen, da diese bei onkologischen Indikationen keine unabhängige Größe darstellt, sondern in Abhängigkeit des Auftretens der Überlebensereignisse bestimmt wird.

4.2.3 Informationsbeschaffung

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

4.2.3.1 Studien des pharmazeutischen Unternehmers

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

Die Darstellung soll auf Studien mit Patienten in dem Anwendungsgebiet, für das das vorliegende Dokument erstellt wird, beschränkt werden.

4.2.3.2 Bibliografische Literaturrecherche

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

– Für das zu bewertende Arzneimittel wurde eine Zulassung in dem Anwendungsgebiet, für das das vorliegende Dokument erstellt wurde, mindestens 12 Monate vor dem Zeitpunkt der Dossiereinreichung erteilt. Dabei sind auch Zulassungen außerhalb Deutschlands bzw. Europas relevant.

– Es werden indirekte Vergleiche zur Bewertung des medizinischen Zusatznutzens herangezogen.

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

Die Suche soll in jeder Datenbank einzeln und mit einer für die jeweilige Datenbank adaptierten Suchstrategie durchgeführt werden. Die Suchstrategien sollen jeweils in Blöcken, insbesondere getrennt nach Indikation, Intervention und ggf. Studientypen, aufgebaut werden. Wird eine Einschränkung der Strategien auf bestimmte Studientypen vorgenommen (z. B. randomisierte kontrollierte Studien), sollen aktuelle validierte Filter hierfür verwendet werden. Eine gemeinsame Suche nach Studien zu mehreren Fragestellungen (z. B. direkt vergleichende Studien sowie Studien für einen indirekten Vergleich) ist möglich. Alle Suchstrategien sind in Anhang 4-A zu dokumentieren.

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

Die Durchführung einer bibliografischen Literaturrecherche ist für den Wirkstoff Pertuzumab nicht erforderlich:

1. Die weltweit erste Zulassung für Pertuzumab wurde weniger als 12 Monate (FDA Zulassung am 08. Juni 2012) vor dem Zeitpunkt der Dossiereinreichung erteilt.

2. Zur Darstellung des (Zusatz-)Nutzens wurden keine indirekten Vergleiche herangezogen.

Der entsprechende Abschnitt 4.3.1.1.2 enthält daher keine Angaben.

4.2.3.3 Suche in Studienregistern

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

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

Die Suche soll in jedem Studienregister einzeln und mit einer für das jeweilige Studienregister adaptierten Suchstrategie durchgeführt werden. Die Suche soll abgeschlossene, abgebrochene und laufende Studien erfassen. Eine gemeinsame Suche nach Studien zu mehreren Fragestellungen (z. B. direkt vergleichende Studien sowie Studien für einen indirekten Vergleich) ist möglich. Alle Suchstrategien sind in Anhang 4-Bzu dokumentieren.

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

Durchsucht wurden die Studienregister clinicaltrials.gov sowie International Clinical Trials Registry Platform. Die Suchstrategie richtete sich nach der Syntax und den Optionen der jeweiligen Suchoberfläche. Pro Register wurde eine Suche zu klinischen Studien mit den Suchbegriffen “Pertuzumab OR Omnitarg OR Perjeta OR RhuMab 2C4 OR RO4368451” durchgeführt. Die Suche in den Studienregistern erfolgte am 10. und 11.12.2012. Es wurde keine zeitliche Einschränkung der Suche vorgenommen. Firmenintern wurden in der Roche-Datenbank „Roche Trial Database: Clinical Trial Protocol Registry“ am 09.01.2013 in der Rubrik „Pharmaceuticals only“ aus den verfügbaren Listeneinträgen die relevanten Einträge „Perjeta“ und „Perjeta [Pertuzumab] durchsucht.

Die detaillierte Suchstrategie findet sich in Anhang 4-B.

4.2.3.4 Selektion relevanter Studien

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

Die Auswahl der Treffer erfolgte, analog dem üblichen Vorgehen bei systematischen Recherchen, in fogenden Schritten: Nach Entfernung der Duplikate wurden zunächst die angezeigten Treffer anhand der elektronisch verfügbaren Informationen auf offensichtliche Verletzungen der vorab definierten Einschlusskriterien (siehe Tabelle 4-1) untersucht. Treffer, die bereits an dieser Stelle eine eindeutige Verletzung eines oder mehrerer der Einschlusskriterien erkennen liessen, wurden im ersten Selektionsschritt ohne weitere Dokumentation der Gründe für ihre Exklusion ausgeschlossen (Vorselektion). Danach wurden die Volltexte der verbleibenden Treffer erneut anhand der Einschlusskriterien hinsichtlich ihrer Relevanz bewertet. Soweit sich in diesem zweiten Selektionsschritt die Notwendigkeit einer Exklusion ergab, wurde die entsprechende Quelle mit Ausschlussgrund dokumentiert.

Die Auswahl wurde von zwei Reviewern unabhängig voneinander durchgeführt. Diskrepanzen zwischen den beiden Reviewern wurden im Konsens gelöst.

4.2.4 Bewertung der Aussagekraft der Nachweise

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

<u>A: Verzerrungsaspekte der Ergebnisse auf Studienebene</u>

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

<u>B: Verzerrungsaspekte der Ergebnisse auf Endpunktebene</u>

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

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

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

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

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

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

Die Bewertung der eingeschlossenen Studien erfolgt anhand der zur Verfügung stehenden Studienberichte und Vollpublikationen der Studien. Die Bewertung erfolgt in zwei Schritten: Erstens der Extraktion bewertungsrelevanter Studiencharakteristika und zweitens der darauf basierenden Bewertung des Verzerrungspotenzials (studienbezogen und endpunktspezifisch). Die Bewertung des Verzerrungspotenzials wurde durch einen zweiten Reviewer überprüft. Diskrepanzen in der Bewertung durch die Reviewer wurden durch Diskussion aufgelöst.

Die Kriterien zur Bewertung entsprechen den Vorgaben der Verfahrensordnung (siehe Anhang 4-G). [11]

Dazu werden endpunktübergreifende (A) und endpunktspezifische (B) Aspekte, die das Verzerrungspotenzial beeinflussen, systematisch extrahiert und bewertet:

A. Aspekte des Verzerrungspotenzials der Ergebnisse auf Studienebene:

- Adäquate Erzeugung der Randomisierungssequenz
- Verdeckung der Gruppenzuteilung
- Verblindung des Patienten sowie des Behandlers
- Hinweis auf ergebnisgesteuerte Berichterstattung
- Sonstige, das Verzerrungspotenzial beeinflussende Punkte

Dabei hatte die Bewertung des Verzerrungspotentials keinen Einfluss auf den Einschluss einer Studie in die Nutzenbewertung. Potentiell „hoch“ verzerrte Studien wurden nicht ausgeschlossen.

B. Aspekte des Verzerrungspotentials der Ergebnisse von Endpunkten:

- Adäquate Umsetzung des ITT-Prinzips
- Verblindung der Endpunkterheber
- Hinweise auf ergebnisgesteuerte Berichterstattung
- Sonstige, das Verzerrungspotenzial beeinflussende Punkte

Das Verzerrungspotenzial wurde als entweder „niedrig“ oder „hoch“ eingestuft.

Bei niedrigem Verzerrungspotenzial kann davon ausgegangen werden, dass die Ergebnisse der Studie nicht relevant verzerrt sind. Ein hohes Verzerrungspotenzial wurde zugewiesen, wenn eine relevante Verzerrung nicht ausgeschlossen werden konnte. Bei einer relevanten Verzerrung auf Studienebene ist die Gesamtaussage der Studie in Frage zu stellen, da eine Behebung der Mängel möglicherweise zu anderen Schlussfolgerungen führen würde.

In einem ersten Schritt wurde das Verzerrungspotenzial gesamthaft pro Studie an Hand der unter A aufgeführten Aspekte als „hoch“ oder „niedrig“ bewertet. Anschließend wurde das Verzerrungspotenzial auf Endpunktebene für jeden Endpunkt separat ebenfalls in den Kategorien „hoch“ oder „niedrig“ bestimmt. Wurde das Verzerrungspotenzial bereits auf Studienebene mit „hoch“ beurteilt, wurden die Endpunkte jeweils trotzdem neu bewertet.

Zusammenfassung der Bewertung der Aussagekraft der Nachweise

Die Ergebnisse aus dem Vergleich der zu untersuchenden Intervention und Placebo („Nutzen“) bzw. mit der zweckmäßigen Vergleichstherapie („Zusatznutzen“) wurden in die von der Verfahrensordung vorgesehenen Kategorien eingeordnet: [11]

„Für Arzneimittel nach Absatz 3 sind das Ausmaß des Zusatznutzens und die therapeutische Bedeutung des Zusatznutzens unter Berücksichtigung des Schweregrades der Erkrankung gegenüber der zweckmäßigen Vergleichstherapie wie folgt zu quantifizieren:

1. Ein erheblicher Zusatznutzen liegt vor, wenn eine nachhaltige und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte große Verbesserung des therapierelevanten Nutzens im Sinne von § 3 Absatz 1 erreicht wird, insbesondere eine Heilung der Erkrankung, eine erhebliche Verlängerung der Überlebensdauer, eine langfristige Freiheit von schwerwiegenden Symptomen oder die weitgehende Vermeidung schwerwiegender Nebenwirkungen nach § 4 Absatz 13 AMG und 4. Kapitel § 23 Absatz 1 VerfO des Gemeinsamen Bundesausschusses.

2. Ein beträchtlicher Zusatznutzen liegt vor, wenn eine gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte deutliche Verbesserung des therapierelevanten Nutzens im Sinne von § 3 Absatz 1 erreicht wird, insbesondere eine Abschwächung schwerwiegender Symptome, eine moderate Verlängerung der Lebensdauer, eine für die Patientinnen und Patienten spürbare Linderung der Erkrankung, eine relevante Vermeidung schwerwiegender Nebenwirkungen oder eine bedeutsame Vermeidung anderer Nebenwirkungen.

3. Ein geringer Zusatznutzen liegt vor, wenn eine gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte moderate und nicht nur geringfügige Verbesserung des therapierelevanten Nutzens im Sinne von § 3 Absatz 1 erreicht wird, insbesondere eine Verringerung von nicht schwerwiegenden Symptomen der Erkrankung oder eine relevante Vermeidung von Nebenwirkungen.

4. Ein Zusatznutzen liegt vor, ist aber nicht quantifizierbar, weil die wissenschaftliche Datengrundlage dies nicht zulässt.

5. Es ist kein Zusatznutzen belegt.

6. Der Nutzen des zu bewertenden Arzneimittels ist geringer als der Nutzen der zweckmäßigen Vergleichstherapie.

Bei der Bewertung der Beleglage stellt das IQWiG an die Ableitung des Belegs für einen Zusatznutzen folgende Anforderungen:

Tabelle 4-3: Anforderungen an die Aussagen zur Beleglage:

Anforderung			
Aussage	**Anzahl der Studien**	**Ergebnissicherheit**	**Effekt**
Beleg	≥ 2	mehrheitlich hoch	„gleichgerichtet“
Hinweis	≥ 2	mehrheitlich mäßig	„gleichgerichtet“
	1	hoch	statistisch signifikant
Anhaltspunkt	≥ 2	mehrheitlich gering	„gleichgerichtet“
	1	mäßig	statistisch signifikant

Zudem kann auch eine einzige Studie für einen Zusatznutzenbeleg genügen, wenn sie besondere Qualitätskriterien gemäß der „CPMP. Points to consider on application with: 1. meta-analyses; 2. one pivotal study” erfüllt. [1;12]

Zur Ableitung der Beleglage für Pertuzumab wird diese in den Allgemeinen Methoden Version 4.0 beschriebene Verfahrensweise (inklusive der Ergänzung zum Nutzenbeleg anhand einer Studie) angewandt.

4.2.5 Informationssynthese und -analyse

4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien

Das Design und die Methodik der eingeschlossenen Studien soll in den Abschnitten 4.3.1 und 4.3.2, jeweils in den Unterabschnitten „Charakteristika der in die Bewertung eingeschlossenen Studien“ und den dazugehörigen Anhängen, dargestellt werden. Die Darstellung der Studien soll für randomisierte kontrollierte Studien mindestens die Anforderungen des CONSORT-Statements erfüllen (Items 2b bis 14, Informationen aus dem CONSORT-Flow-Chart)[1]. Die Darstellung nicht randomisierter Interventionsstudien und epidemiologischer Beobachtungsstudien soll mindestens den Anforderungen des TREND-[2] bzw. STROBE-Statements[3] 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.

Auf Basis der Angaben im Studienbericht wurden alle Detailinformationen zu Studiendesign und Ergebnissen extrahiert und in den vorgesehenen Tabellen und Modulen des Dossiers dokumentiert. Die vorgegebene Tabelle zu Design und Durchführung der Studie wurde nach den Vorgaben von CONSORT für jede Studie separat ausgefüllt (siehe Tabelle 4-88 im Anhang 4-F des Dossiers). Alle Ergebnisdaten werden im Modul 4.3 dargestellt.

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 Endpunkte, zu denen Ergebnisse dargestellt werden. Begründen Sie, wenn Sie von den oben benannten Vorgaben abgewichen sind.

---

[1] Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials.BMJ 2010; 340: c332.
[2] Des Jarlais DC, Lyles C, Crepaz N. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. Am J Publ Health 2004; 94(3): 361-366.
[3] Von Elm E, Altman DG, Egger M, Pocock SJ, Gøtsche PC, Vandenbroucke JP. The strengthening the reporting of observational studies in epidemiology (STROBE) statement: guidelines for reporting observational studies. Ann Intern Med 2007; 147(8): 573-577.

Beschreiben Sie für jeden Endpunkt, warum Sie ihn als patientenrelevant einstufen, und machen Sie Angaben zur Validität des Endpunkts (z. B. zur Validierung der eingesetzten Fragebögen). Beachten Sie bei der Berücksichtigung von Surrogatendpunkten Abschnitt 4.5.4.

Im vorliegenden Bericht wurden die Studien anhand ihrer Designcharakteristika (Studiendauer, Zahl randomisierter Patienten, Ort und Zeitraum der Durchführung, primäre und sekundäre Zielkriterien) beschrieben. Darüber hinaus wurde die krankheitsspezifische Therapie in den eingeschlossenen Studien dargestellt.

Die Studienpopulation wurde durch die wesentlichen Ein- und Ausschlusskriterien, demografische Daten und weitere Charakteristika beschrieben. Daneben wurden Details zu Intervention und Vergleichstherapie erfasst.

Die Ergebnisse zu den in den Studien berichteten patientenrelevanten Endpunkten wurden für die jeweiligen Behandlungsarme beschrieben und verglichen. Zunächst wurden die in den Studien verwendeten Endpunkte den verwendeten Zielgrößen zugeordnet (Operationalisierung des Endpunkts). Die detaillierten Auswertungen pro Endpunkt wurden pro Endpunkt in gesonderten Kapiteln dargestellt. Folgende Endpunkte wurden ausgewertet:

- Mortalität: Gesamtüberleben (OS): finale konfirmatorische Analyse, Mai 2012

Gesamtüberleben (overall survival, OS) wird in randomisierten klinischen Studien als Zeitpunkt der Randomisierung bis zum Tod aus jeglicher Ursache operationalisiert. Somit stellt das Gesamtüberleben eine direkte Übertragung der Mortalität auf den Zeitraum klinischer Studien dar. Dies entspricht der Definition dieses Endpunkts in den Leitlinien für onkologische Studien der Zulassungsbehörden (Definition EMA / FDA: „time from randomization to death from any cause“).

- Morbidität:
    - Progressionsfreies Überleben: finale konfirmatorische Analyse, Mai 2011

Das progressionsfreie Überleben (bewertet durch IRF) ist definiert als Zeit zwischen Randomisierung und dem Zeitpunkt einer Krankheitsprogression oder des Todes durch jegliche Ursache. Die Endpunkterheber und die Behandler waren hinsichtlich dieses Endpunktes verblindet. In der G-BA Beratung wurde das progressionsfreie Überleben diskutiert.

Roche sieht das progressionsfreie Überleben als patientenrelevanten Endpunkt, da auch die Therapieentscheidung in der Onkologie von der Krankheitsprogression abhängig gemacht wird und das progressionsfreie Überleben häufig das Kriterium für die Trennung der Therapielinien in der Onkologie darstellt. Die optimale Behandlung der Erkrankung zu jedem Zeitpunkt ist als unmittelbar patientenrelevant anzusehen und ist nach Auffassung von Roche im Zusammenhang mit der Morbidität zu diskutieren.

Eine vertiefende Diskussion dieses Themas geht über den Inhalt dieses Dossiers hinaus. Roche bittet den G-BA zu klären, wie Morbidität (sowohl symptomatische als auch asysmptomatische Morbidität) bei onkologischen Krankheiten innerhalb kleiner Studien operationalisiert werden kann (vgl. dazu das Kurzgutachten von Prof. Schwartz) [13].

- Gesamtansprechrate (ORR): finale konfirmatorische Analyse, Mai 2011

Das Tumoransprechen (bewertet durch IRF) wird als nach den RECIST Kriterien ermitteltes, vollständiges oder teilweises Ansprechen des Tumors definiert.

Abweichend von der IQWiG Position sieht Roche den Endpunkt „Tumoransprechen“ als patientenrelevant im Sinne eines Morbiditätskriteriums an. [14] Das Tumoransprechen könnte beispielsweise im Zusammenhang mit dem in den Vorgaben zu diesem Abschnitt genannten Endpunkt „Verbesserung des Gesundheitszustandes“ bei onkologischen Erkrankungen diskutiert werden.

- Lebensqualität: 1. Datenschnitt Mai 2011

Die gesundheitsbezogene Lebensqualität von Patienten sollte unter einer Behandlung nicht verschlechtert werden, wobei dieses Ziel insbesondere bei onkologischen Therapien durch die damit verbundene Toxizität eine spezielle Bedeutung besitzt.

Entsprechend der Aussagen im IQWiG Rapid Report „Aussagekraft von Surrogatendpunkten in der Onkologie“ ist „(…) die gesundheitsbezogene Lebensqualität (…) ein direkt patientenrelevantes Maß für die Progression der Erkrankung.“ [14]

Die Lebensqualität sollte in klinischen Studien mit anerkannten und validierten Messinstrumenten erfasst und operationalisiert werden.

Ein bibliographischer Review der Literatur von Montazeri aus dem Jahr 2008 zeigt, dass EORTC QLQ-BR23 und FACT-B die am häufigsten benutzten Fragebögen zur Erfassung der Lebensqualität bei Brustkrebspatienten sind. Die Tabelle 5 dieses Reviews zeigt außerdem die Validierung in verschiedenen Sprachen. [15] Die aktuelle deutsche S3-Leitlinie der DKG benennt ebenfalls die EORTC QLQ C30 und FACT-G (Hauptbestandteile der spezifischen Brustkrebs-Fragebögen EORTC QLQ-BR23 und FACT-B) als standardisierte Fragebögen, die hinsichtlich Validität und Reliabilität überprüft worden sind. [8]

Algorithmen zur Auswertung der Subskalen sowie des Gesamtscores sind vorgegeben. Die Imputation fehlender Werte unter Beibehaltung der Validität definiert das Guideline Manual von FACIT.org (http://www.facit.org). [16]

- Sicherheit: 2. Datenschnitt Mai 2012
  - Unerwünschte Ereignisse (UE)
  - Schwerwiegende unerwünschte Ereignisse (SUE)
  - Behandlungsabbruch wegen UE
  - UE ≥ Grad 3

- UE Grad 3
- UE Grad 4
- UE Grad 5
- UE von besonderem Interesse

Unerwünschte Ereignisse können eine direkte Übertragung der Morbidität (Beschwerden und Komplikationen) auf in klinischen Studien messbare Endpunkte sein. Leichte Abweichungen eines Laborparameters, die vom Patienten nicht wahrgenommen werden, stellen nicht unbedingt ein patientenrelevantes Ereignis dar. Demgegenüber ist bei unerwünschten Ereignissen höherer Schweregrade (z.B. nach National Cancer Institute Common Terminology Criteria for Adverse Events), schwerwiegenden unerwünschten Ereignissen und unerwünschten Ereignissen, die zum Therapieabbruch führen, unmittelbar von einer Patientenrelevanz auszugehen.

4.2.5.3 Meta-Analysen

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

Für die statistische Auswertung sollen primär die Ergebnisse aus Intention-to-treat-Analysen, so wie sie in den vorliegenden Dokumenten beschrieben sind, verwendet werden. Die Meta-Analysen sollen in der Regel auf Basis von Modellen mit zufälligen Effekten[4] erfolgen. In begründeten Ausnahmefällen sollen zusätzlich Modelle mit festen Effekten eingesetzt 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 anhand der Odds Ratio durchgeführt werden. In begründeten Ausnahmefällen können auch andere Effektmaße zum Einsatz kommen. Bei kategorialen Variablen soll ein geeignetes Effektmaß in Abhängigkeit vom konkreten Endpunkt und den verfügbaren Daten verwendet[5] werden.

Die Effektschätzer und Konfidenzintervalle aus den Studien sollen mittels Forest Plots zusammenfassend dargestellt werden. Anschließend soll die Einschätzung einer möglichen Heterogenität der Studienergebnisse anhand des Maßes I^2 und des statistischen Tests auf Vorliegen von Heterogenität[6] erfolgen. Ist die Heterogenität der Studienergebnisse nicht bedeutsam, soll der gemeinsame (gepoolte) Effekt inklusive Konfidenzintervall dargestellt werden. Bei bedeutsamer Heterogenität sollen die Ergebnisse nur in begründeten Ausnahmefällen gepoolt werden. Außerdem soll untersucht werden, welche Faktoren diese Heterogenität möglicherweise erklären könnten. Dazu zählen methodische Faktoren (siehe Abschnitt4.2.5.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt4.2.5.5).

Beschreiben Sie die für Meta-Analysen eingesetzte Methodik. Begründen Sie, wenn Sie von der oben beschriebenen Methodik abweichen.

Sofern methodisch möglich, werden zur Informationssynthese und -analyse Meta-Analysen verwendet. Gemäß IQWiG Vorgaben in den Allgemeinen Methoden Version 4.0 kommen vorrangig Modelle mit zufälligen Effekten zur Anwendung. [1] Nur in begründeten Ausnahmefällen auf Modelle mit festen Effekten ausgewichen.

---

[4] *DerSimonian R, Laird N. Meta-analysis in clinical trials. Control Clin Trials 1986; 7(3):177-188.*
[5] *Deeks JJ, Higgins JPT, Altman DG. Analysing data and undertaking meta-analyses. In: Higgins JPT, Green S (Ed). Cochrane handbook for systematic reviews of interventions. Chichester: Wiley; 2008. s. 243-296.*
[6] *Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. BMJ 2003; 327(7414):557-560.*

Auf den Einsatz von Meta-Analysen wird verzichtet, wenn,

1) weniger als zwei randomisierte kontrollierte Studien vorliegen. In diesem Fall ist der Einsatz von Meta-Analysen nicht sinnvoll bzw. nicht möglich. In die entsprechenden Abschnitte im Dossier wird in diesem Fall „Nicht zutreffend." eingetragen.

2) eine große Heterogenität vorliegt. Ergibt sich für die zu untersuchenden Studien gemäß der Vorgaben für diesen Abschnitt (s.o.) eine hohe Heterogenität, so wird im Einzelfall begründet entschieden, dass keine Meta-Analyse durchgeführt wird oder warum eine Meta-Analyse der Studien dennoch sinnvoll ist. Mittels Sensitivitätsanalysen (siehe folgender Abschnitt) wird die Ursache für die Heterogenität näher untersucht.

Da bei der Suche nach Studien für diese Nutzenbewertung nur eine Studie identifiziert wurde, entfällt die Durchführung von Meta-Analysen.

4.2.5.4 Sensitivitätsanalysen

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

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

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

Auf die Durchführung studienübergreifender Sensitivitätsanalysen wird verzichtet, wenn die Anzahl der randomisierten kontrollierten Studien, die in diese Nutzenbewertung eingehen, kleiner als zwei ist. In diesem Fall werden verschiedene Auswertungen der Ergebnisse (nicht parametrisch: log rank-Test, parametrisch: Hazard ratio etc.) dargestellt. Die Konsistenz der Ergebnisse wird zusätzlich mittels Subgruppenanalysen auf Einzelstudienebene untersucht (siehe folgender Abschnitt).

4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren

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

– Geschlecht
– Alter
– Krankheitsschwere bzw. -stadium

Sollten sich aus den verfügbaren Informationen Anhaltspunkte für weitere mögliche Effektmodifikatoren ergeben, können diese ebenfalls begründet einbezogen werden. Die Ergebnisse von in Studien a priori geplanten und im Studienprotokoll festgelegten Subgruppenanalysen für patientenrelevante Endpunkte sind immer darzustellen.

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

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

Um Anhaltspunkte für die Konsistenz des Therapieeffektes hinsichtlich Wirksamkeit und Sicherheit von Pertuzumab zu erhalten, wurden die folgenden präspezifizierten Subgruppen bzw. Effektmodifikatoren für den primären Endpunkt „IRF-bewertetes PFS“ sowie für das „Gesamtüberleben“ analysiert:

- Vorausgehender Behandlungsstatus: de novo; adjuv. od. neoadjuvante Behandlung
- Region: Europa, Nordamerika, Südamerika, Asien
- Altersgruppe: <65, ≥65
- Altersgruppe: <75, ≥75
- Rasse: kaukasisch, afro-amerikanisch, asiatisch, weitere
- Art der Erkrankung: Viszeral, nicht-viszeral
- IHC-HER2-Status: 3+
- FISH-Status: FISH positiv
- ER / PgR-Status; ER und / oder PgR positiv, ER und PgR negativ, Unbekannt
- (Biomarker, aufgrund der fehlenden Klassifikation als patientenrelevant nicht im vorliegenden Bericht berücksichtigt, zur Wahrung der Vollständigkeit aber hier erwähnt)

Neben diesen Analysen wurden zusätzlich *post hoc* folgende Analysen für weitere Subgruppen über diese beiden Endpunkte durchgeführt:

- IHC-HER2-Status: 2+
- Baseline ECOG: 0; 1+
- Dosiseskalation mit Docetaxel auf 100 mg / m^2: ja, nein

Post-hoc Analysen für das Gesamtüberleben zum zweiten Datenschnitt im Mai 2012 wurden nur für G-BA vorgeschlagene Subgruppen zusätzlich durchgeführt.

Im Rahmen des Beratungsgesprächs wurden vom G-BA weitere Subgruppenauswertungen vorgeschlagen:

- Vorherige (neo-)adjuvante Taxanbehandlung: ja, nein
- Lokalrezidiv: ja, nein

Zusätzlich wurden für das vorliegende Dossier zu allen übrigen Wirksamkeits- und Lebensqualitätsendpunkten *post hoc Analysen* für alle oben genannten Subgruppenmerkmale durchgeführt, um zu überprüfen, ob der Therapieeffekt über die Subgruppen hinweg konstant ist oder ob Hinweise für eine Effektmodifikation vorliegen.

- Vorausgehender Behandlungsstatus: de novo; adjuvante. od. neoadjuvante Behandlung
- Region: Europa, Nordamerika, Südamerika, Asien
- Altersgruppe: <65, ≥65
- Altersgruppe: <75, ≥75
- Rasse: kaukasisch, afro-amerikanisch, asiatisch, weitere
- Art der Erkrankung: Viszeral, nicht-viszeral
- IHC-HER2-Status: 3+

- FISH-Status: FISH positiv
- ER / PgR-Status; ER und / oder PgR positiv, ER und PgR negativ, Unbekannt
- IHC-HER2-Status: 2+
- Baseline ECOG: 0, 1+,
- Vorherige (neoadjuvante oder adjuvante) Behandlung mit Trastuzumab: ja, nein
- Dosiseskalation mit Docetaxel auf 100 mg/m^2: ja, nein
- Vorherige (neo-)adjuvante Taxanbehandlung: ja, nein
- Lokalrezidiv: ja, nein

Bei der Auswertung zur Sicherheit von Pertuzumab wurden für alle Sicherheitsendpunkte folgende Standardsubgruppen ausgewertet:

- Altersgruppe: <65, ≥65
- Rasse: kaukasisch, afro-amerikanisch, asiatisch, weitere

Die Begründungen für die Wahl der einzelnen Kriterien sind nachfolgend kurz ausgeführt.

- Adjuvante Chemotherapie: Eine vorangegangene adjuvante Chemotherapie wird als ungünstiger Prognosefaktor für Patientinnen mit metastasiertem Mammakarzinom angesehen. [17;18]

- Geographische Regionen: In der CLEOPATRA Studie erfolgte eine Stratifikation nach den geographischen Regionen Europa, Nordamerika, Südamerika und Asien. Die CLEOPATRA Studie wurde weltweit durchgeführt. Generell versucht ein Studienprotokoll über die Definition von diagnostischen Standards und klaren Regeln für die Begleittherapie sowie mit der genauen Definition von Dosierungen und Verabreichung von Standardtherapie und experimenteller Therapie, möglichst homogene Voraussetzungen für die Bewertung der Effektivität einer neuen experimentellen Therapie zu gewährleisten. Trotz allem können sich medizinische Standards in unterschiedlichen Ländern und Regionen der Welt unterscheiden und Wirksamkeit und Verträglichkeit von Therapien beeinflussen.

- Alter: Das Alter wird nicht zu den Prognosefaktoren für das metastasierte Mammakarzinom gezählt. [8;17;18] Allerdings ist das biologische Alter der Patientien durchaus ein Kriterium, das die Auswahl der systemischen Therapie beim metastasierten Mammakarzinom beeinflusst. [19] Dabei stimmen biologisches Alter und kalendarisches Alter nicht notwendigerweise überein. Dennoch ist es aus regulatorischen Gründen üblich, den Effekt einer Therapie bei jüngeren und älteren Patienten zu betrachten. Sowohl die europäische als auch die amerikanische Zulassungsbehörde fordern ausdrücklich eine gesonderte Betrachtung von älteren und sehr alten Patienten. Die Altersgrenzen sind von der EMA bei 65 und 75 Jahren festgelegt. [20-23]

- Ethnische Zugehörigkeit: Die Betrachtung nach ethnischer Zugehörigkeit sollte nach den Richtlinien der Zulassungsbehörden, insbesondere bei globalen klinischen Studien erfolgen. Rationale hierfür sind Beobachtungen über Unterschiede in der Wirksamkeit von Arzneimitteln bei ethnischen Subgruppen. Als Gründe dafür werden intrinsische Faktoren wie z.B. Unterschiede beim Metabolismus oder bei der Elimination von Arzneimitteln, äußere Faktoren wie Diät oder Umweltfaktoren, oder eine Kombination aus beidem angenommen. Man kennt z.B. Unterschiede in der Ausstattung mit dem für die Metabolisierung von vielen Medikamenten wichtigen Enzym CYP2D6 im Ansprechen auf antihypertensive Medikamente oder auf Interferon-alpha bei Hepatitis C. [24-29]

- Metastasenlokalisation: Abhängig von der Lokalisation der metastatischen Läsionen unterscheidet man beim metastasierten Mammakarzinom den viszeralen Befall von inneren Organen wie Lunge und Leber von einer nicht-viszeralen Erkrankung mit einem Befall von Knochen, Lymphknoten und Haut. Wenn sich die Metastasierung auf Knochen und / oder Haut beschränkt, wird von einem günstigeren Krankheitsverlauf ausgegangen. [8;17;18] Dazu liegen Untersuchungen aus Kollektiven von Patientinnen mit metastasiertem Mammakarzinom vor, in der noch keine Bestimmung des HER2-Status erfolgte. [30-33]

- HER2-Status: Der Nachweis von HER2 gilt als prädiktiver Faktor für eine Therapie mit Trastuzumab. Patienten, deren Tumore eine starke Überexpression von HER2 zeigen (IHC 3+) und / oder bei denen etwa mit FISH eine HER2-Genamplifikation nachgewiesen werden konnte (FISH-positiv), profitieren am meisten von der Anti-HER2-Therapie. Das zeigte eine Analyse von Gewebeproben aus der randomisierten Phase III Studie, die Trastuzumab + Chemotherapie vs. Chemotherapie bei Patientinnen mit metastasiertem HER2-positivem Mammakarzinom prüfte. So hatten Patientinnen mit einer First-Line Therapie mit Trastuzumab und Chemotherapie, deren Tumore IHC 3+ oder FISH-positiv waren, ein niedrigeres relatives Risiko für die Zeit bis zur Progression gegenüber Patientinnen mit einem HER2-Status von IHC 2+ (RR für IHC 3+ 0,42, für FISH-positiv 0,44 und für ICH 2+ 0,76). [34;35]

- Hormonrezeptorstatus: Ein negativer Hormonrezeptorstatus (Kein Nachweis des Östrogen (ER)- und des Progesteronrezeptors (PR)) gilt als ungünstiger, ein positiver Hormonrezeptorstatus (Nachweis von ER und / oder PR) als günstiger prognostischer Faktor beim metastasierten Mammakarzinom. [8;17;18]

- Allgemeinzustand: Der Allgemeinzustand gilt als prognostischer Faktor für das metastasierte Mammakarzinom. [8;17;18] Dabei wird bei einem guten Allgemeinzustand von einer besseren Prognose ausgegangen. Eine gängige Einteilung des Allgemeinzustands erfolgt nach der Klassifikation der Eastern Cooperative Group (ECOG), die die Stufen 0-5 für den ECOG Performance Status vorsieht. ECOG 0 bedeutet, dass der Patient ohne oder mit nur sehr geringen Krankheitszeichen ohne Einschränkungen normal aktiv ist. ECOG 1 bedeutet, dass der Patient in seiner

körperlichen Aktivität leicht eingeschränkt ist, jedoch noch leichtere Arbeiten durchführen kann und noch eine ambulante Behandlung möglich ist. [36] Die ECOG Stadien 0 und 1 werden in der Regel als guter Allgemeinzustand zusammengefasst, die Stadien 2 und mehr stehen für einen reduzierten Allgemeinzustand.

- Vorherige Behandlung mit Trastuzumab: Wenn Patienten nach einer (neo-) adjuvanten Trastuzumab-basierten Therapie einen Rückfall erleiden, könnte das ein Hinweis darauf sein, dass sie eine umfassendere HER2-Blockade benötigen, welche mit Pertuzumab in Kombination mit Trastuzumab erreicht wird. Es gibt derzeit keine differenzierten Empfehlungen für die Anti-HER2-Therapie für Patienten, die nach einer (neo)-adjuvanten Therapie mit Trastuzumab progredient sind. Bei der ABC Konsensus-konferenz 2011 wurde betont, dass diesen Patienten eine Anti-HER2-Therapie angeboten werden sollte, wobei sich die Auswahl nach der Länge des rezidivfreien Intervalls, der vorangegangenen Anti-HER2-Therapie und nach der Verfügbarkeit von zielgerichteten Therapien richten sollte. [37]

- Dosiseskalation von Docetaxel auf 100 mg / m^2: In einigen Ländern wird Docetaxel nicht in der Dosierung 100 mg / m^2 eingesetzt. Sowohl die Docetaxel als auch die Pertuzumab Fachinformation, kann die Docetaxel-Dosis auf bis zu 100 mg / m^2 gesteigert werden, wenn die Initialdosis gut vertragen wird.

- Vorherige Therapie mit Taxanen: Taxane sind effektive Substanzen in der Therapie des metastasierten Mammakarzinoms. [5;8] Mehrere Patienten erhalten auch eine (neo-)adjuvante Taxantherapie. Um eine Aussage über die Konsistenz des Therapieeffekts treffen zu können, wurde eine Subgruppenanalyse für Patienten mit einer vorangegangenen (neo-)adjuvanten Taxantherapie durchgeführt. [3]

- Der Resektionsstatus des Lokalrezidivs gilt als prognostischer Faktor für den weiteren Krankheitsverlauf. [8] Genauere Daten zur Prognose des inoperablen lokoregionären Rezidivs sind spärlich, weisen jedoch auf eine sehr ungünstige Prognose hin. Um eine Aussage über die Konsistenz des Therapieeffekts treffen zu können, wurde für die Subgruppe „Lokalrezidiv“ jeweils nur der Test auf Subgruppeninteraktion durchgeführt, da das Verhältnis Ja / Nein bei 19 / 787 (entsprechend 2,36 %) der Patienten lag.

Für alle Subgruppen wurden in einem ersten Schritt Tests auf Interaktionen durchgeführt. Diese Tests wurden für die Time-to-event Analysen in SAS in ‚PROC PHREG‘ durchgeführt. Dafür wurde ein Interaktionsterm (treatment*subgroup) ins Modell eingefügt und ein likelihood ratio Test mit dem Modell ohne Interaktionsterm berechnet. Es wurden Typ 3 Tests spezifiziert für die default Wald-Teststatistik. Die Interaktionstests für das Relative Risiko wurden in SAS mit ‚PROC LOGISTIC‘ durchgeführt. Die p-Werte wurden auf Basis des Wald-Tests berechnet.

Eingangs des Kapitels 4.2.5.5 wird darauf verwiesen, dass „soweit sinnvoll“ die Faktoren Geschlecht, Alter und Krankheitsschwere bzw. –stadium hinsichtlich einer potentiellen Effektmodifikation berücksichtigt werden sollen. Dies wurde für die letztgenannten Kriterien

„Alter und Krankheitsschwere bzw. –stadium“ berücksichtigt. Angesichts der Seltenheit des Brustkrebs bei Männern (weniger als 1 % aller Brustkrebs-Neudiagnosen) und des entsprechend geringen Anteils männlicher Patienten an der für die Nutzenbewertung von Pertuzumab herangezogenen Studienpopulation der CLEOPATRA-Studie (n = 2 / 808 (0,2%) männliche Studienteilnehmer) wurde auf die Durchführung einer Subgruppenanalyse über das Merkmal „Geschlecht“ verzichtet. [38]

4.2.5.6 Indirekte Vergleiche

Zurzeit sind international Methoden in der Entwicklung, um indirekte Vergleiche zu ermöglichen. Nicht adjustierte indirekte Vergleiche (d. h. Vergleiche einzelner Behandlungsgruppen aus verschiedenen Studien ohne Bezug zu einem gemeinsamen Komparator) stellen dabei keine valide Analysemethode dar, der Einsatz einfacher adjustierter indirekter Vergleiche ist möglich[7]. Komplexe Verfahren für den simultanen Vergleich von mehr als zwei Therapien unter Berücksichtigung sowohl direkter als auch indirekter Vergleiche werden in der Literatur unterschiedlich bezeichnet, z. B. als „Mixed-Treatment-Comparison(MTC)-Meta-Analysen“[8], „Multiple-Treatment-Meta-Analysen“[9] oder auch „Netzwerk-Meta-Analysen“[10], sie gehen aber im Prinzip von denselben wesentlichen Annahmen aus.

Grundannahme für solche komplexen Analysen ist die Annahme der Konsistenz innerhalb des zu analysierenden Netzwerkes. Als Inkonsistenz wird dabei die Diskrepanz zwischen dem Ergebnis eines direkten und eines oder mehreren indirekten Vergleichen verstanden, die nicht mehr nur durch Zufallsfehler oder Heterogenität erklärbar ist[11].

Insgesamt ist es notwendig, die zugrunde liegende Methodik genau und reproduzierbar zu beschreiben und die Annahme der Konsistenz zu untersuchen[12].

[7] *Glenny AM, Altman DG, Song F, Sakarovitch C, Deeks JJ, D‘Amico R et al. Indirect comparisons of competing interventions. Health Technol Assess 2005; 9(26): 1-148.*
[8] *Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004; 23(20): 3105-3124.*
[9] *Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005; 331(7521): 897-900.*
[10] *Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. Stat Methods Med Res 2008;17(3): 279-301.*
[11] *B. Schöttker, D. Lühmann, D. Boulkhemair, and H. Raspe. Indirekte Vergleiche von Therapieverfahren. Schriftenreihe Health Technology Assessment Band 88, DIMDI, Köln, 2009.*
[12] *Song F, Loke YK, Walsh T, Glenny AM, Eastwood AJ, Altman DJ. Methodological problems in the use of indirect comparisons for evaluating healthcare interventions: survey of published systematic reviews. BMJ 2009; 338: b1147.*

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

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

Liegt mindestens eine direkt vergleichende Studie vor, die Pertuzumab direkt mit der zweckmäßigen Vergleichstherapie im relevanten Anwendungsgebiet vergleicht, wird auf die Durchführung eines indirekten Vergleichs verzichtet. In die entsprechenden Abschnitte wird „Nicht zutreffend“ eingetragen.

Seitens der Roche Pharma AG wurde entsprechend den genannten Vorgaben keine Notwendigkeit für die Durchführung eines indirekten Vergleichs gemäß der oben genannten Vorgaben gesehen.

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)
– Nichtrandomisierte vergleichende Studien (Abschnitt4.3.2.2)
– Weitere Untersuchungen (Abschnitt 4.3.2.3)

4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel

4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel

4.3.1.1.1 Studien des pharmazeutischen Unternehmers

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

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

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

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

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

Studienbezeichnung	**Zulassungsstudie (ja/nein)**	**Status (abgeschlossen/abgebrochen/ laufend)**	**Studiendauer**	**Therapiearme**
CLEOPATRA (WO20698)	ja	Laufend: nach Rekrutierung, Studie in Follow-up-Phase	Beginn Februar 2008 Datenschnitte von 05 / 2011 und 05 / 2012	Pertuzumab+Trastuzumab+Docetaxel vs. Placebo+Trastuzumab+Docetaxel
MARIANNE (BO22589) NCT01120184 EUCTR2009-017905-13-AT	Nein	Laufend	Beginn Juli 2010	T-DM1+Pertuzumab T-DM1+Placebo Trastuzumab+Taxane
PERTAIN (MO27775) NCT01491737 EUCTR2011-002132-10-ES	Nein	Laufend	Beginn Februar 2012	Trastuzumab+Pertuzumab+Aromatase Inhibitor Trastuzumab+Aromatase Inhibitor
NCT01597414 EORTC-75111-10114	Nein	Rekrutierung noch nicht begonnen	Beginn geplant	Pertuzumab+Trastuzumab Pertuzumab+Trastuzumab+metronomic chemotherapy

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

Aufgelistet sind alle randomisierten, kontrollierten Studien des pharmazeutischen Unternehmers für das zu bewertende Arzneimittel, die im Anwendungsgebiet (HER2-positiver metastasierter oder lokal rezidivierender, inoperabler Brustkrebs, der zuvor noch nicht mit einer anti-HER2-Therapie oder Chemotherapie aufgrund der metastasierten Erkrankung behandelt wurde – entsprechend Erstlinientherapie) geplant, laufend oder durchgeführt worden sind. Stand: 10.Dezember 2012

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

Tabelle 4-5: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	**Begründung für die Nichtberücksichtigung der Studie**
MARIANNE (BO22589) NCT01120184 EUCTR2009-017905-13-AT [39]	Verletzt Einschlusskriterium E2 / E3
PERTAIN (MO27775) NCT1491737 EUCTR2011-002132-10-ES [40]	Verletzt Einschlusskriterium E2 / E3
NCT01597414 EORTC-75111-10114 [41]	Verletzt Einschlusskriterium E2 / E3
Einschlusskriterien befinden sich in Tabelle 4-2.	

Ausgeschlossen wurde weiterhin die VELVET-Studie (NCT01565083; EUCTR2011-003308-18-ES). Diese wird unter dem ICTRP-Register fälschlicherweise als „randomized“ gelistet, unter www.clinicaltrials.gov hingegen korrekt als „non-randomized“ beschrieben.

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

4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche

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

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

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

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

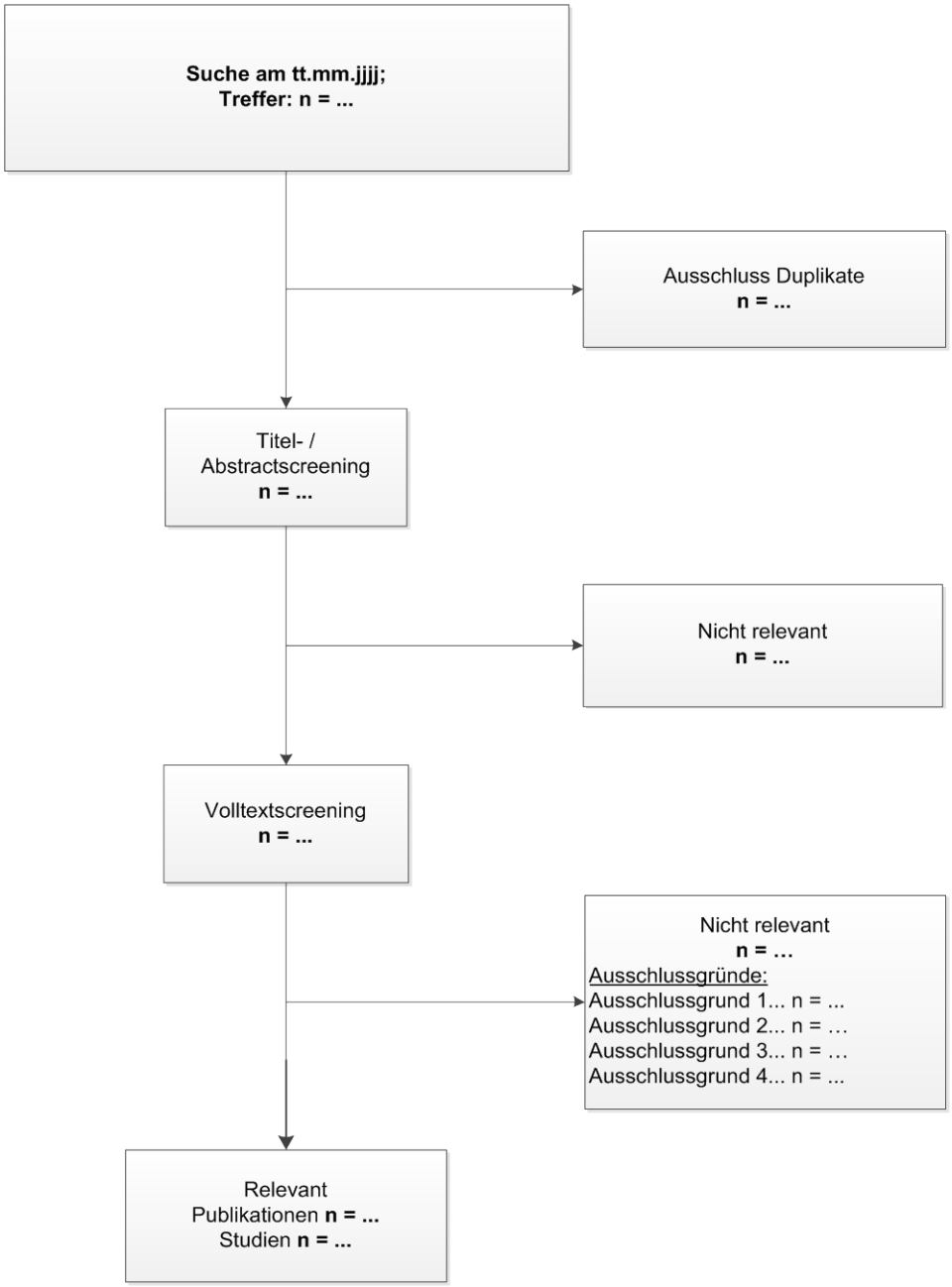


Abbildung 1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten, kontrollierten Studien mit dem zu bewertenden Arzneimittel

Das zu bewertende Arzneimittel Pertuzumab ist ein neuer Wirkstoff. Die weltweit erste Zulassung wurde am 8. Juni 2012 von der FDA erteilt. Gemäß den Vorgaben zur Erstellung des Dossiers zum Punkt 4.2.3.2 entfällt die bibliografische Literaturrecherche.

4.3.1.1.3 Studien aus der Suche in Studienregistern

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

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

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

Tabelle 4-6: Relevante Studien aus der Suche in Studienregistern – RCT mit dem zu bewertenden Arzneimittel

Studienbezeichnung	**Identifikationsorte (Name der Studienregister und Angabe der Zitate[a])**	**Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja / nein)**	**Studie durch bibliografische Literaturrecherche identifiziert (ja / nein / n. d.[b])**
<CLEOPATRA (WO20698)>	ClinicalTrials.gov:[42] ICTRP: [43;44] Roche Trials Database: [45]	ja	n.d
a: Zitat des Studienregistereintrags sowie, falls vorhanden, der im Studienregister aufgelisteten Berichte über Studiendesign und / oder -ergebnisse. b: n. d. = bibliografische Literaturrecherche nicht durchgeführt.			

4.3.1.1.4 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel

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

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

Tabelle 4-7: Studienpool – RCT mit dem zu bewertenden Arzneimittel

Studie	**Studienkategorie**			**Verfügbare Datenquellen[a]**		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	**gesponserte Studie[b] (ja/nein)**	**Studie Dritter (ja/nein)**	**Studienbericht (ja/nein [Zitat])**	**Registereintrag[c] (ja/nein [Zitat])**	**Publikation (ja/nein [Zitat])**
ggf. Zwischenüberschrift zur Strukturierung des Studienpools						
CLEOPATRA (WO20698)	ja	ja	nein	ja [46;47]	ja ClinicalTrials.gov:[42] ICTRP: [43;44] Roche Trials Database: [45]	Ja [48;49]

a: Bei Angabe „ja“ sind jeweils die Zitate der Datenquelle(n) mit anzugeben.
b: Studie, für die der Unternehmer Sponsor war oder an der der Unternehmer anderweitig finanziell beteiligt war.
c: Zitat der Studienregistereinträge sowie, falls vorhanden, der in den Studienregistern aufgelisteten Berichte über Studiendesign und / oder -ergebnisse.

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

4.3.1.2.1 Studiendesign und Studienpopulationen

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

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

Tabelle 4-8: Charakterisierung der eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel

Studie	**Studiendesign** <RCT, doppelblind/einfach verblindet/offen, parallel/cross-over etc.>	**Population** <relevante Charakteristika, z. B. Schweregrad>	**Interventionen (Zahl der randomisierten Patienten)**	**Studiendauer** <ggf. Run-in, Behandlung, Nachbeobachtung>	**Ort und Zeitraum der Durchführung**	**Primärer Endpunkt; patientenrelevante sekundäre Endpunkte**
<CLEOPATRA (WO20698)>	RCT Doppelblind für Studienperiode 1 und Studienperiode 2 Parallel	Erwachsene Patienten mit HER2-positivem metastatiertem oder lokal rezidivierendem, inoperablem Brustkrebs, die zuvor noch keine anti-HER2-Therapie oder Chemotherapie zur Behandlung ihrer metastasierten Erkrankung erhalten haben.	Placebo + Trastuzumab + Docetaxel (Pla+T+D) (n = 406) Pertuzumab + Trastuzumab + Docetaxel (Ptz+T+D) (n = 402)	Laufend, keine Patientenrekrutierung Studienperiode 1: 2 / 2008 – 5 / 2011 *First Patient in* bis klinischer *cut-off*: Konfirmatorische Analyse des PFS nach Auftreten von ungefähr 381 IRF-bewerteten PFS-Ereignissen geplant. Erster Datenschnitt war 05 / 2011 erreicht. Zu diesem Zeitpunkt erfolgte die erste Interimanalyse für OS. Studienperiode 2: 06 / 2011-05 / 2012 Finale konfirmatorische Analyse für OS. Finaler Datenschnitt mit konfirmatorischen Analysen. Studienperiode 3: 6 / 2012 – Q3 / 4 / 2013	Multizentrisch: Brasilien, Kanada, China, Costa Rica, Kroatien, Ecuador, Frankreich, Finnland, Deutschland, United Kingdom, Guatemala, Italien, Japan, Lettland, Mazedonien, Mexiko, Polen, Argentinien, Korea, Philippinen, Russland, Singapur, Spanien, Thailand, USA Studienphase. 1: 2 / 2008 – 5 / 2011 Studienphase. 2: 6 / 2011 – 5 / 2012 Studienphase. 3: 06 / 2012 – Q3 / 4 / 2013	**Primärer Endpunkt:** Progressionsfreies Überleben bewertet durch IRF (PFS) **Sekundäre Endpunkte:** Gesamtüberleben (OS) Gesamtansprechrate (ORR) Für diese Endpunkte war im Protokoll für den Zeitpunkt der primären Analyse des PFS folgende Testhierarchie zur Adjustierung bzgl. multiplen Testens festgelegt: 1. Zweiseitiger Test des PFS (IRF) zum Signifikanzniveau von 5%. Bei Signifikanz weiter mit Stufe 2, sonst Ende des Testprocedere 2. Zweiseitiger Test des OS zum Gesamtsignifikanzniveau von 5%. Bei Signifikanz weiter mit Stufe 3, sonst Ende des Testprocedere 3. Zweiseitiger Test der ORR zum Signifikanzniveau von 5%. **Weitere sekundäre Endpunkte:** Progressionsfreies Überleben (festgestellt durch

	Analyse des OS nach 385 Todes-fällen.	Investigator) Zeit bis Symptomprogression **Sicherheit:** · Alle Unerwünschten Ereignisse (UE) · Schwerwiegende unerwünschte Ereignisse (SUE) · Behandlungsabbruch wegen UE · UE ≥ Grad 3 · UE Grad 3 · UE Grad 4 · UE Grad 5 · UE von besonderem Interesse **Patientenrelevante exploratorische Endpunkte** Zeit bis zur Verschlechterung des BCS-Score

Alle Patienten erhielten den Behandlungsstandard als Basistherapie (hier: Trastuzumab und Docetaxel).

Tabelle 4-9: Charakterisierung der Interventionen – RCT mit dem zu bewertenden Arzneimittel aus CLEOPATRA (WO20698)

Studie	**< Placebo + Trastuzumab + Docetaxel>**	**< Pertuzumab + Trastuzumab + Docetaxel>**	**Behandlungscharakteristika**
<CLEOPATRA (WO20698)>	Pertuzumab Placebo: i.v. Infusion alle 3 Wochen (q3w) Trastuzumab: Start-dosis von 8 mg / kg IV, gefolgt von 6 mg / kg i.v. q3w Docetaxel 75 mg / m^2 i.v. q3w für mind. 6 Wochen	Pertuzumab: Startdosis von 840 mg / kg i.v., gefolgt von 420 mg / kg i.v. q3w Trastuzumab: Startdosis von 8 mg / kg i.v., gefolgt von 6 mg / kg i.v. q3w Docetaxel 75 mg / m^2 i.v. q3w für mind. 6 Zyklen	Nach Einschätzung des Investigators, konnte die Dosis für Docetaxel auf 100 mg / m^2 für Patienten erhöht werden, die diese Dosis ohne signifikante Toxizität tolerierten. Die Behandlung wurde bis zum Auf-treten einer durch den Investigator bewerteten radiografisch oder klinisch progressiven Krankheit, einer unakzeptablen Toxizität oder einem Widerruf der Einverständniserklärung der Patienten gegeben. Im Fall eines permanenten Therapie-abbruchs oder einer Unterbrechung von mehr als zwei Zyklen, wurde der Patient von der Studie ausgeschlossen. Unterbrach ein Patient nur die Behandlung mit Docetaxel aufgrund einer Toxizität, konnte die Therapie mit Pertuzumab / Placebo und Trastuzumab fortgesetzt werden.

Es war geplant, insgesamt 800 Patienten in der Studie aufzunehmen, randomisiert im Verhältnis 1:1 auf einen der Behandlungsarme (A oder B):

<u>Arm A (Pla+T+D):</u>

- Pertuzumab Placebo: i.v. Infusion alle 3 Wochen (q3w)
- Trastuzumab: Startdosis von 8 mg / kg i.v., gefolgt von 6 mg / kg i.v. q3w
- Docetaxel 75 mg / m^2 KOF i.v. q3w für mindestens sechs Zyklen

<u>Arm B (Ptz+T+D):</u>

- Pertuzumab: Startdosis von 840 mg / kg i.v., gefolgt von 420 mg / kg i.v. q3w
- Trastuzumab: Startdosis von 8 mg / kg i.v., gefolgt von 6 mg / kg i.v. q3w
- Docetaxel 75 mg / m^2 KOF i.v. q3w für mindestens sechs Zyklen.

Die Studienpopulation umfasste erwachsene Patienten mit HER2-positivem metastatiertem oder lokal rezidivierendem, inoperablem Brustkrebs, die zuvor noch keine anti-HER2-

Therapie oder Chemotherapie zur Behandlung ihrer metastasierten Erkrankung erhalten haben. Damit entspricht die Population dem Zulassungstext für Pertuzumab. Die CLEOPATRA-Studie erfüllt somit das Kriterium des IQWiG für den Einschluss von Studien in die Nutzenbewertung, nachdem bei mindestens 80 % der in eine Studie eingeschlossenen Patienten das entsprechende Indikationsgebiet vorliegen muss.

Tabelle 4-10: Charakterisierung der Studienpopulation: Baseline Charakteristika der Patienten in der Studie nach Behandlungsgruppe– RCT mit dem zu bewertenden Arzneimittel aus CLEOPATRA (WO20698)

	Total N = 808	**Placebo +Trastuzumab +Docetaxel** N = 406	**Pertuzumab+Trastuzumab +Docetaxel** N = 402
Geschlecht (Anzahl, %)			
Weiblich	806 (99,8 %)	404 (99,5 %)	402 (100 %)
Männlich	2 (0,2 %)	2 (0,5 %)	0
N	808	406	402
Alter (Jahre)			
Mittelwert	53,5	53,5	53,4
SD	11,14	11,35	11,94
SEM	0,39	0,56	0,55
Median	54,0	54,0	54,0
Min-Max	22-89	27-89	22-82
N	808	406	402
Altersgruppen (Jahre) (Anzahl, %)			
< 65	681 (84,3 %)	339 (83,5 %)	342 (85,1 %)
≥ 65	127 (15,7 %)	67 (16,5 %)	60 (14,9 %)
N	808	406	402
Altersgruppen (Jahre) (Anzahl, %)			
< 75	789 (97,6 %)	392 (96,6 %)	397 (98,8 %)
≥ 75	19 (2,4 %)	14 (3,4 %)	5 (1,2 %)
N	808	406	402
Gewicht (in kg)			
Mittelwert	66,6	66,2	67,0
SD	15,05	14,70	15,40
SEM	0,53	0,73	0,77
Median	64,9	64,7	65,0
Min-Max	39-142	39-142	39-129
N	808	406	402
Größe (in cm)			
Mittelwert	158,8	158,8	158,8
SD	6,84	6,95	6,75
SEM	0,24	0,35	0,34
Median	159,0	159,0	158,0
Min-Max	139-181	139-180	139-181
N	807	405	402

	Total N = 808	**Placebo +Trastuzumab +Docetaxel** N = 406	**Pertuzumab+Trastuzumab +Docetaxel** N = 402
Region (Anzahl, %)			
Asien	253 (31,3 %)	128 (31,5 %)	125 (31,1 %)
Europa	306 (37,9 %)	152 (37,4 %)	154 (38,3 %)
Nordamerika	135 (16,7 %)	68 (16,7 %)	67 (16,7 %)
Südamerika	114 (14,1 %)	58 (14,3 %)	56 (13,9 %)
N	808	406	402
Rasse (Anzahl, %)			
Asiatisch	261 (32,3 %)	133 (32,8 %)	128 (31,8 %)
Afro-amerikanisch	30 (3,7 %)	20 (4,9 %)	10 (2,5 %)
Weitere	37 (4,6 %)	18 (4,4 %)	19 (4,7 %)
Kaukasisch	480 (59,4 %)	235 (57,9 %)	245 (60,9 %)
N	808	406	402
Frauen: Reproduktionsstatus (Anzahl, %)			
Postmenopausal	512 (63,5 %)	251 (62,1 %)	261 (64,9 %)
Chirurgisch steril	86 (10,7 %)	46 (11,4 %)	40 (10,0 %)
Mit kont. Verhütung	206 (25,6 %)	107 (26,5 %)	99 (24,6 %)
Ohne kont. Verhütung	2 (0,2 %)	-	2 (0,5 %)
N	806	404	402
Männer: Reproduktionsstatus (Anzahl, %)[a]			
ja	2	2	-
N	2	2	0
Raucher-Status (Anzahl, %)			
Aktuell Raucher	75 (9,3 %)	41 (10,1 %)	34 (8,5 %)
Niemals geraucht	629 (77,8 %)	315 (77,6 %)	314 (78,1 %)
Ehemaliger Raucher	104 (12,9 %)	50 (12,3 %)	54 (13,4 %)
N	808	406	402
Baseline ECOG Status (Anzahl, %)			
0	522 (64,6 %)	248 (61,1 %)	274 (68,2 %)
1	282 (34,9 %)	157 (38,7 %)	125 (31,1 %)
2	3 (0,4 %)	-	3 (0,7 %)
3	1 (0,1 %)	1 (0,2 %)	-
N	808	406	402
Baseline LVEF(%)			
Mittelwert	65,2	65,5	64,9
SD	6,62	6,53	6,71
SEM	0,23	0,33	0,34
Median	65,0	65,0	65,0
Min-Max	50-88	50-88	50-88
N	803	403	400
Baseline Thorax-Röntgen abnormal (Anzahl, %)			
Nein	476 (62,8 %)	237 (62,4 %)	239 (63,2 %)
Ja	282 (37,2 %)	143 (37,6 %)	139 (36,8 %)
N	758	380	378

	Total N = 808	**Placebo +Trastuzumab +Docetaxel** N = 406	**Pertuzumab+Trastuzumab +Docetaxel** N = 402
Vorheriger Behandlungsstatus			
Adjuvante oder neo-adjuvante Therapie	376 (46,5 %)	192 (47,3 %)	184 (45,8 %)
De novo	432 (53,5 %)	214 (52,7 %)	218 (54,2 %)
N	808	406	402
Vorherige neo-adjuvante Chemotherapie oder biologische Therapie* (Anzahl, %)			
Ja	107 (13,2 %)	57 (14,0 %)	50 (12,4 %)
N	808	406	402
Vorherige adjuvante Chemotherapie oder biologische Therapie* (Anzahl, %)			
Ja	337 (41,7 %)	172 (42,4 %)	165 (41,0 %)
N	808	406	402
Vorherige Behandlung mit Trastuzumab* (Anzahl, %)			
Ja	88 (10,9 %)	41 (10,1 %)	47 (11,7 %)
N	808	406	402
Vorherige Behandlung mit Taxanen* (Anzahl, %)			
Docetaxel	72 (8,9 %)	38 (9,4 %)	34 (8,5 %)
Paclitaxel	111 (13,7 %)	57 (14,0 %)	54 (13,4 %)
Taxane (nicht anderweitig spezifiziert)	1 (0,1 %)	0 (0,0 %)	1 (0,2 %)
Taxane gesamt	184 (22,8 %)	95 (23,4 %)	89 (22,1 %)
N	808	406	402
Krankheitstyp bei Screening (Anzahl, %)			
Nicht-viszerale Erkrankung	178 (22,0 %)	90 (22,2 %)	88 (21,9 %)
Viszerale Erkrankung	630 (78,0 %)	316 (77,8 %)	314 (78,1 %)
N	808	406	402
IRF-determinierte Erkrankung bei Screening (Anzahl, %)			
Messbare Erkrankung	679 (88,6 %)	336 (88,7 %)	343 (88,6 %)
Nicht-messbare Erkrankung	87 (11,4 %)	43 (11,3 %)	44 (11,4 %)
N	766	379	387
Investigator-determinierte Erkrankung bei Screening (Anzahl, %)			
Messbare Erkrankung	738 (91,3 %)	371 (91,4 %)	367 (91,3 %)
Nicht-messbare Erkrankung	70 (8,7 %)	35 (8,6 %)	35 (8,7 %)
N	808	406	402

n repräsentiert die Anzahl Patienten, die in die Summenstatistik eingehen
Prozentangaben basieren auf n (Anzahl valider Werte); Prozentzahlen wurden nicht kalkuliert, wenn n < 10
a: Stimmt der Patient der Verwendung einer effektiven Verhütungsmethode zu? (nur auf Männer zutreffend, die zu Protokoll C gehörten)
ITT-Population (Gruppen nach randomisierter Behandlung), alle Zentren
Snapshot Date: 06 Juli 2011; Clinical Cut-Off Date: 13 Mai 2011

Tabelle 4-11: Charakterisierung der Studienpopulation – Krankheitscharakteristika: RCT mit dem zu bewertenden Arzneimittel aus CLEOPATRA (WO20698)

	Placebo+Trastuzumab+Docetaxel N = 406	**Pertuzumab+Trastuzumab+Docetaxel N = 402**
Brustkrebs Subtyp (Anzahl, %)		
Duktal	368 (90,6 %)	368 (91,5 %)
Lobulär	18 (4,4 %)	20 (5,0 %)
Medullär	3 (0,7 %)	1 (0,2 %)
Tubulär	4 (1,0 %)	2 (0,5 %)
Muzinös	3 (0,7 %)	10 (2,5 %)
Komedo	16 (3,9 %)	17 (4,2 %)
Inflammatorisch	10 (2,5 %)	6 (1,5 %)
Unbekannt	13 (3,2 %)	8 (2,0 %)
Andere	21 (5,2 %)	14 (3,5 %)
Histologischer Tumorgrad (Anzahl, %)		
Anaplastisch	1 (0,2 %)	2 (0,5 %)
Moderat differenziert	132 (32,6 %)	132 (32,8 %)
Schlecht differenziert	125 (30,9 %)	130 (32,3 %)
Unbekannt	131 (32,3 %)	124 (30,8 %)
Gut differenziert	16 (4,0 %)	14 (3,5 %)
N	405	402
Lokalrezidiv (Anzahl, %)		
Lokalrezidiv	8 (2,0 %)	11 (2,7 %)
Metastasierte Erkrankung	397 (98,0 %)	390 (97,3 %)
N	405	401
ER / PgR Status (Anzahl, %)		
Negativ	196 (48,3 %)	212 (52,7 %)
Positiv	199 (49,0 %)	189 (47,0 %)
Unbekannt	11 (2,7 %)	1 (0,2 %)
n	406	402
HER2-Status IHC (Anzahl, %)		
0+	-	2 (0,5 %)
1+	2 (0,5 %)	2 (0,5 %)
2+	32 (7,9 %)	47 (11,7 %)
3+	371 (91,6 %)	350 (87,3 %)
N	405	401
HER2-Status FISH (Anzahl, %)		
Negativ	4 (1,0 %)	1 (0,3 %)
Positiv	383 (99,0 %)	384 (99,7 %)
N	387	385
HER2-Status IHC / FISH kombiniert (Anzahl, %)		
- / FISH positiv	1 (0,2 %)	1 (0,2 %)
IHC 0+ / FISH positiv	-	2 (0,5 %)
IHC 1+ / FISH positiv	2 (0,5 %)	2 (0,5 %)
IHC 2+ / FISH negativ	1 (0,2 %)	-

	Placebo+Trastuzumab+Docetaxel N = 406	**Pertuzumab+Trastuzumab+Docetaxel N = 402**
IHC 2+ / FISH positive	31 (7,6 %)	47 (11,7 %)
IHC 3+ / -	19 (4,7 %)	17 (4,2 %)
IHC 3+ / FISH negativ	3 (0,7 %)	1 (0,2 %)
IHC 3+ / FISH positiv	349 (86,0 %)	332 (82,6 %)
N	406	402
ITT-Analyse (Gruppen nach randomisierter Behandlung) Snapshot Date: 06 Juli 2011; Clinical Cut-Off Date: 13 Mai 2011 CSR 10 / 2011 Seite 952-953		

Für die Auswertung der Studie wurden zwei Analysepopulationen definiert: Intent-to-Treat Population (ITT) und Safety-Analysis-Population (SAP).

Die ITT Population beinhaltet alle 808 randomisierten Patienten (Pla+T+D: n = 406; Ptz+T+D: n = 402). Die SAP bezieht sich auf Patienten, welche mindestens eine Dosis von einer der Komponenten der Studienbehandlung erhalten hatten. Die SAP beinhaltet 804 Patienten, von denen zwei im jeweiligen Behandlungsarm keine Studienmedikation nach der Randomisierung erhalten hatten. Dies trifft für den ersten wie auch für den zweiten Datenschnitt zu.

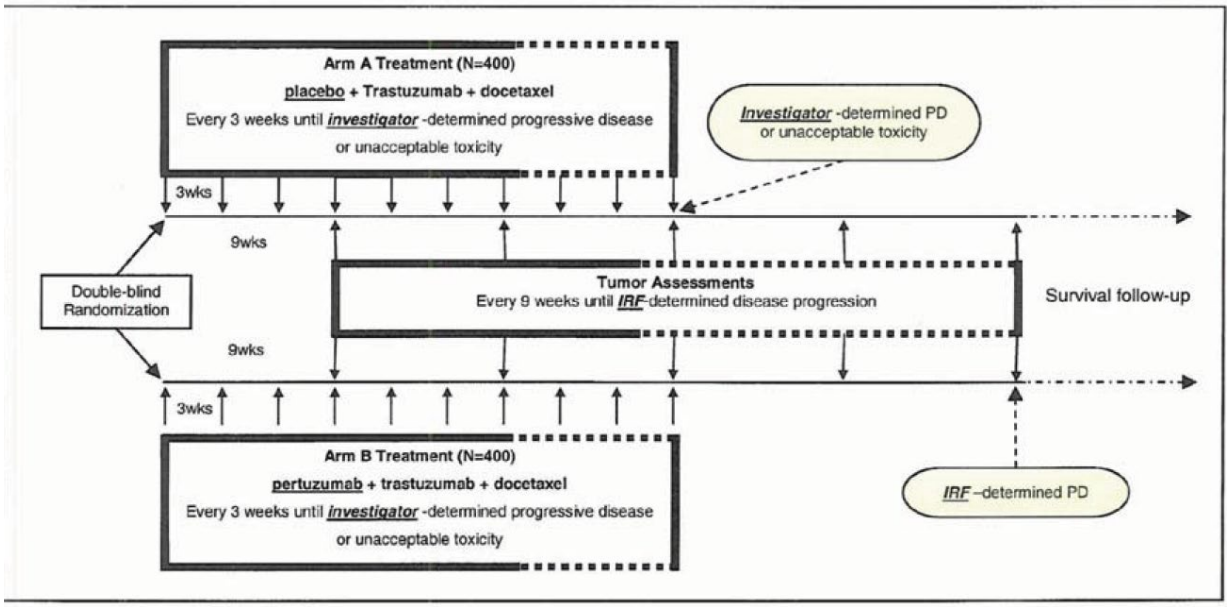
Für die Safety Analyse wurden alle Patienten eingeschlossen, die mindestens eine Komponente der Studienmedikation erhalten hatten. Erster Datenschnitt (13.5.2011): Acht der randomisierten Patienten, welche Placebo erhalten sollten, hatten tatsächlich mindestens eine Dosis Pertuzumab während der Studie erhalten. Für die Safety Analyse wurden diese Patienten in den Ptz+T+D Arm mit einbezogen. Zusätzlich erhielt ein Patient, der auf den Behandlungsarm mit Pertuzumab randomisiert war, in jedem Zyklus Placebo. Dieser Patient wurde in den Pla+T+D Arm einbezogen. Folglich beinhaltet die SAP 397 Patienten im Pla+T+D Arm und 407 Patienten im Ptz+T+D Arm. Seit der Durchführung der 1. Analyse (13.5.2011), erhielt ein weiterer Patient aus der Placebo-Gruppe versehentlich Pertuzumab, und wurde gemäß. Analyseplan der Safety-Population zugeordnet, so dass die Anzahl Patienten in der Safety Analyse sich änderte: Der betroffene Patient erhielt im Zyklus 33 (April 2012) Pertuzumab, im folgenden Zyklus 34 (kurz vor dem Datenschnitt) wurde wieder Placebo verabreicht. Daher wurde die Safety Analyse für den zweiten Datenschnitt mit 396 Patienten im Pla+T+D Arm und mit 408 Patienten im Ptz+T+D Arm gerechnet.

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

Studiendesign

Die CLEOPATRA-Studie ist eine randomisierte, doppelblindete, kontrollierte, internationale, multizentrische klinische Studie der Phase III im Parallelgruppen-Design.

Insgesamt ist die Studie in drei Perioden unterteilt: Am Ende der Studienperiode 1 (Februar 2008 bis Mai 2011) erfolgte der erste Datenschnitt mit (i) der finalen konfirmatorischen Analyse des primären Endpunktes IRF-bewertetes PFS und (ii) der 1. Interimsanalyse des Gesamtüberlebens. Mit dem zweiten Datenschnitt am Ende der Studienperiode 2 (Juni 2011 bis Mai 2012) erfolgte die finale konfirmatorische Analyse des Gesamtüberlebens. In dieser Analyse war der Effekt des Gesamtüberlebens so erheblich, dass die konfirmatorischen Phasen der Studie beendet wurden. Die Studie wurde entblindet und Cross-Over ermöglicht. Derzeit wird die Studie in Studienperiode 3 weitergeführt (Juni 2012 – Q3 / 4 2013), bis insgesamt 385 Todesfälle aufgetreten sein werden.



IRF: Independent Review Facility; PD: progressive disease.

Abbildung 2: Studiendesign der CLEOPATRA-Studie: Behandlungsarme und Bewertungszeitpunkte [46]

Interimanalysen und Follow-up Datenschnitte

Studien mit Interimanalysen und Follow-up-Analysen (im Folgenden Datenschnitte genannt) sind für die Bewertung der Studienergebnisse eine Herausforderung. Die unterschiedlichen Datenschnitte sind keineswegs als gleichwertig zu betrachten.

Mit einer adäquaten Planung einer Interimanalyse sind die formalen statistischen Kriterien zur vorzeitigen Beendigung einer klinischen Studie erfüllt. Der erhöhten Ergebnisunsicherheit,

die sich zu diesem Zeitpunkt durch die geringere Patientenzahl ergibt, wird in diesem Fall durch den sehr großen Behandlungseffekt entgegengewirkt. Diese Sicht wird durch die behördliche Zulassung bestätigt.

Weitere spätere Datenschnitte sind als explorative ‚follow-up'-Analysen zu betrachten, die keine inferenzstatistische Bedeutung haben. Dies gilt insbesondere in Fällen, in denen die Studien verblindete RCTs sind, da die vorzeitige Beendigung der Studie zumeist auch Änderungen am Studiendesign (z.B. cross-over) mit sich bringt. Damit ändert sich für die späteren Datenschnitte nicht nur der inferenzstatistische Status der Ergebnisse, sondern auch noch die Aussagesicherheit, da sich das Verzerrungspotential für einige Endpunkte durch die vorgenommene Entblindung erhöht. Dies gilt für alle zu bewertenden Endpunkte (QoL, PFS, ORR), aber weniger für Mortalitäts-und Safety-Endpunkte.

Nach Erreichen der finalen konfirmatorischen Analyse des Gesamtüberlebens im Mai 2012 wurde die CLEOPATRA-Studie entblindet, sowie ein Cross-Over erlaubt. Der spätere Datenschnitt nach insgesamt 385 Todesfälle (Q3 / 4 2013) ist somit als explorative ‚follow-up'-Analyse zu betrachten.

Endpunkte

Primäres Ziel der Studie war der Vergleich des progressionsfreien Überlebens (*progression-free survival* (PFS)) zwischen den beiden Behandlungsgruppen, basierend auf der Bewertung durch ein unabhängiges Review-Kommittee (*independent review facility* (IRF)). Sekundäre Ziele waren der Vergleich zwischen den Studienarmen für folgende Endpunkte:

- Gesamtüberleben (OS)
- Gesamtansprechrate (ORR)
- Zeit bis Symptomprogression
- Zeit bis zur Verschlechterung des BCS-Score
- Alle Unerwünschten Ereignisse (UE)
- Schwerwiegende unerwünschte Ereignisse (SUE)
- Behandlungsabbruch wegen UE
- UE ≥ Grad 3
- UE Grad 3
- UE Grad 4
- UE Grad 5
- UE von besonderem Interesse

Ein unabhängiges Data Monitoring Committee (DMC) war für das Monitoring der Sicherheit zuständig. Zusätzlich dazu wurde ein unabhängiges *Cardiac Review Committee* (CRC) etabliert, um die verblindeten kardialen Daten während der Studie zu reviewen und alle sechs Monate die Ergebnisse an das DMC für einen Review zu berichten.

Die Befunde bezüglich der Krankheitsprogression (PD) und des Gesamtansprechens (*ORR)* wurden durch das IRF geprüft, das in regelmäßigen Abständen die radiografischen, zytologischen und photografischen Daten aller Patienten einem Review unterzog.

Die Stratifikationsfaktoren „Region“ und „vorhergehender Behandlungsstatus“ waren zwischen den Studienarmen gut balanciert. Circa ein Drittel aller Patienten kam aus Europa (37,9 %), ein weiteres Drittel aus Asien (31,3 %). Das übrige Drittel stammte gleichermaßen aus Nord- und Südamerika. Circa die Hälfte aller Patienten in jeder Behandlungsgruppe erhielt vorhergehend entweder eine adjuvante oder neoadjuvante Therapie (192 Patienten [47,3 %] im Arm Pla+T+D vs. 184 Patienten [45,8 %] im Arm Ptz+T+D). Die übrigen Patienten hatten eine de novo Diagnose mit Metastasen oder lokal rezidivierenden, inoperablen Brustkrebs.

Von den Patienten, deren Tumoruntersuchungen zu Studienbeginn durch das IRF ausgewertet wurden (379 im Arm Pla+T+D und 387 im Arm Ptz+T+D), zeigte sich bei 89 % der Patienten der jeweiligen Arme (336 bzw. 343) ein messbarer Tumor. Etwas weniger Patienten unter Pla+T+D hatten einen ECOG Status von 0 (248 Patienten [61,1 %] vs 274 Patienten [68,2 %]), und zeigten einen leicht schlechteren Leistungsstatus nach ECOG, verglichen mit Patienten unter Ptz+T+D. Nur vier Patienten in der gesamten Studie hatten einen schlechteren ECOG Leistungsstatus als 0 oder 1 zu Studienbeginn.

4.3.1.2.2 Verzerrungspotenzial auf Studienebene

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

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

Tabelle 4-12: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel

Studie	Adäquate Erzeugung der Randomisierungs-sequenz	Verdeckung der Gruppenzuteilung	Verblindung		Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Ver-zerrungspotenzial beeinflussende Punkte	Verzerrungs-potenzial auf Studienebene
			Patient	Behandler			
<CLEOPATRA (WO20698) >	ja	ja	ja	ja	nein	nein	niedrig

Begründen Sie für jede Studie die abschließende Einschätzung.

In der randomisierten, kontrollierten, doppelt verblindeten parallelen CLEOPATRA-Studie wurden die Patienten im Verhältnis 1:1 auf die Studienarme randomisiert. als Strata fanden der vorherige Behandlungsstatus (de novo vs. vorherige adjuvante oder neoadjuvante Therapie) und die Region (Europa, Nordamerika, Süd-Amerika und Asien) Eingang in die Studie. Patienten und Ärzte waren in dieser Studie verblindet. Um die Ausgeglichenheit zwischen den Behandlungsarmen zu gewährleisten, wurde innerhalb jeden Stratums ein komplettes Block-Randomisierungsschema angewendet. Die Screening-Informationen wurden mittels eines Interactive Voice Response Systems (IVRS) erfasst und die Patienten entsprechend auf die Behandlungsarme randomisiert. Die Verblindung der Ärzte und Patienten in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Eine Absicherung durch den medizinischen Monitor des Sponsors war dafür erforderlich. Die Studie läuft derzeit unverblindet in der Studienperiode 3, bis insgesamt 385 Todesfälle aufgetreten sein werden.

Der Abgleich der Publikation mit dem Studienbericht ergab keine Hinweise auf ergebnisgesteuerte Berichterstattung. Baselga, 2012 [49] Sonstige, das Verzerrungspotential beeinflussende Punkte wurden ebenfalls nicht identifiziert. Insgesamt wurde daher das Verzerrungspotential auf Studienebene mit „niedrig“ beurteilt.

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.

Tabelle 4-13: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel

Studie	**Gesamtüberleben**	**Progressionsfreies Überleben bewertet durch IRF**	**Gesamtansprechrate**
CLEOPATRA (WO20698)	ja	ja	ja

Studie	**Zeit bis Symptomprogression**	**Zeit bis zur Verschlechterung des BCS-Score**	**Patienten mit UE**	**Patienten mit SUE**	**Patienten mit Behandlungsabbruch wegen UE**
CLEOPATRA (WO20698)	ja	ja	ja	ja	ja

Studie	**Patienten mit UE ≥ Grad 3**	**Patienten mit UE Grad 3**	**Patienten mit UE Grad 4**	**Patienten mit UE Grad 5**	**Patienten mit UE von besonderem Interesse**
CLEOPATRA (WO20698)	ja	ja	ja	ja	ja

4.3.1.3.1 Endpunkt – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer 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
– dem Endpunkt entsprechende Kennzahlen pro Behandlungsgruppe
– bei Verlaufsbeobachtungen Werte zu Studienbeginn und Studienende inklusive Standardabweichung
– bei dichotomen Endpunkten die Anzahlen und Anteile pro Gruppe
– entsprechende Maße bei weiteren Messniveaus
– Effektschätzer mit zugehörigem Standardfehler

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

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

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

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

Begründen Sie für jede Studie die abschließende Einschätzung.

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

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.

Da bei der Suche nach Studien für diese Nutzenbewertung nur eine Studie identifiziert wurde, entfällt die Durchführung von Meta-Analysen.

Exkurs: Medianes Gesamtüberleben vs. Hazard Ratio

Belastbarkeit von Endpunkten

Die Hazardfunktionen in den jeweiligen Behandlungsarmen sind Bausteine des **Hazard Ratios**. Letzteres ist eine eindimensionale Maßzahl (analog dem relativen Risiko) unter Berücksichtigung der Zeit als stetiger Größe. Zur Berechnung werden alle verfügbaren Daten verwendet und die Annahme getroffen, dass die Hazardfunktionen proportional zueinander sind; dies ist äquivalent zu der Annahme, dass das Hazard Ratio konstant ist. Damit ist das Hazard Ratio ein stabiler Schätzer, auch wenn die Follow-up-Zeiten nicht allzu lang sind, gegeben die korrekte Annahme proportionaler Hazardfunktionen.

Das Hazard Ratio hat ferner gegenüber dem medianen Überleben den Vorteil, nicht von der Größe der Hazard Rate abzuhängen. [50]

Survival ist eine dichotome Variable, insofern ist eine Responder-Analyse implizit angelegt. Betrachtet man nun das Hazard Ratio als eine auf einem Quotienten basierende Responder-Analyse, so genügt gemäß Lange et al. (2010) der Test auf einfache statistische Signifikanz zum Nachweis der Relevanz. [51] Somit ist der Test einer verschobenen Nullhypothese nach Victor (1987) nicht erforderlich. [52]

Die **mediane Überlebenszeit** (Median Survival) ist das 50 %-Quantil der Ereigniszeitverteilung, der früheste Zeitpunkt an dem die geschätzte Survivalfunktion kleiner oder gleich 0,5 ist, unter der Voraussetzung, dass die Survivalfunktion unter 0,5 gesunken ist. Dieser Wert wird auf die Zeitachse projiziert und der Wert des entsprechenden Patienten bildet die mediane Überlebenszeit (analog für die Konfidenzintervallgrenzen). Hierbei sind einige methodische Aspekte erheblich:

- Zensierte Patienten, also solche, die zum Zeitpunkt der Betrachtung noch ohne Ereignis sind, werden bei der Berechnung nicht berücksichtigt bzw. sind ohne Belang (individueller KM-Schätzer = 1). Sind die Follow-up Zeiten kurz und damit der Zensierungsanteil hoch, so ist das Median Survival bezogen auf die Gesamtpopulation ein nicht belastbarer Schätzer; er schätzt eine zu kurze mediane Überlebenszeit. Einem Vorschlag von Pocock et al. (2002) folgend, sind die Kaplan-Meier Schätzer für bestimmte Zeitpunkte dann nicht belastbar, wenn die Anzahl der Patienten, die über diesen Zeitpunkt hinaus im Follow-up ist, unangemessen klein ist (z.B. Vorschlag von <10 %). [53]
- Anders als das Hazard Ratio ist der mediane Überlebenszeitunterschied zwischen zwei Behandlungen abhängig von der Größe der zugrundeliegenden Hazard-Rate: Je höher die Hazard Rate, desto geringer ist der mediane Überlebensvorteil. [50]

4.3.1.3.1.1 <Mortalität> – RCT

4.3.1.3.1.1.1 <Endpunkt Gesamtüberleben> – RCT

Tabelle 4-14: Operationalisierung von <Endpunkt Gesamtüberleben>

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	Gesamtüberleben (*overall survival* (OS)) war definiert als Zeit zwischen dem Tag der Randomisierung bis zum Tod durch jegliche Ursache. Zensierung: Patienten, die noch leben oder als *lost to follow-up* verzeichnet sind oder die Einverständniserklärung zurückgezogen hatten, werden zum Zeitpunkt der Analyse mit dem letzten bekannten Datum, an dem sie lebend waren, zensiert. Patienten, für die kein Fragebogen ausgefüllt wurde (z.B. wenn der Patient zum Zeitpunkt der Analyse noch unter Behandlung ist und noch lebt) wird zum letzten bekannten Datum zensiert. Patienten, für die keine Informationen nach der Baseline vorliegen, werden mit dem Tag nach der Randomisierung + einen Tag zensiert.

Tabelle 4-15: Bewertung des Verzerrungspotenzials für <Endpunkt Gesamtüberleben> in RCT mit dem zu bewertenden Arzneimittel (finale konfirmatorische Analyse. 2. Datenschnitt Mai 2012)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Das Verzerrungspotential für den Endpunkt <Gesamtüberleben> wurde bei einem niedrigen Verzerrungspotential auf Studienebene beurteilt: Die Patienten in den Studienarmen zeigten eine Homogenität hinsichtlich der Baseline-Charakteristika. Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Eine Absicherung durch den medizinischen Monitor des Sponsors war dafür erforderlich. Das ITT-Prinzip wurde adäquat umgesetzt, da alle randomisierten Patienten in die Auswertung eingingen. Hinweise auf eine ergebnisgesteuerte Berichterstattung ergaben sich beim Abgleich mit der Publikation der Studie nicht. [49] Auch sonstige, das Verzerrungspotential beeinflussende Punkte, wurden nicht beobachtet. Daher wurde das Verzerrungspotential für den Endpunkt <Gesamtüberleben > mit „niedrig“ beurteilt.

Tabelle 4-16: Ergebnisse für<Endpunkt Gesamtüberleben> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (finale konfirmatorische Analyse. 2. Datenschnitt Mai 2012)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T+D Hazard Ratio [95% KI] p-Wert**	**Ptz+T+D vs Pla+T+D Zeit bis zum Ereignis in Monaten Median p-Wert[b]**
	Patienten mit Ereignis [a] n/N (%)	**Zeit bis zum Ereignis in Monaten Median [95% KI]**	**Patienten mit Ereignis [a] n/N (%)**	**Zeit bis zum Ereignis in Monaten Median [95% KI]**		
Gesamtüberleben [c]	154 / 406 (37,9 %)	37,6 [34; n.e.]	113 / 402 (28,1 %)	[42; n.e.]	0,66 [0,52; 0,84] 0,0008	0,0008[b]

a: Ereignis bezieht sich hier auf den Tod
b: Log-Rank Test, stratifiziert nach vorherigem Behandlungsstatus und Region
c: Die Ergebnisse sind nach vorherigem Behandlungsstatus und Region stratifiziert; ITT Population
Datenquelle: Tabelle 7 aus CSR 12/2012
n.e. nicht erreicht

Analyse der ITT-Population:

Bis zum Zeitpunkt der finalen konfirmatorischen Analyse (Mai 2012) traten 267 Todesfälle auf: 154 Todesfällen (37,9 %) unter Pla+T+D im Vergleich zu 113 Fällen (28,1 %) unter Ptz+T+D (ITT-Population). Die Ergebnisse für das Gesamtüberleben (OS) in dieser Analyse (HR = 0,66, 95% KI: 0,.52; 0,84; p = 0,0008; stratifiziert nach vorheriger Behandlung und Region) überschritten die prädefinierte Grenze nach O'Brien-Fleming (*O'Brien−Fleming stopping boundary*) für die α-Funktion nach Lan−DeMets (*α-spending function*) (HR ≤ 0,739; p ≤ 0,0138) und zeigten daher einen statistisch signifikanten Überlebensvorteil für die Patienten, die auf den Behandlungsarm Ptz+T+D randomisiert waren. Das mediane Überleben wurde im Pla+T+D Arm mit 37,6 Monaten erreicht, jedoch nicht im Ptz+T+D Arm. (CSR12/2012 S.44 Tabelle 7)

Eine unstratifizierte Analyse bestätigte die Ergebnisse wie oben beschrieben: HR = 0,66 95 %KI [0,52; 0,84]. (CSR 12/2012 S. 302)

Berechnung der medianen Gesamtüberlebenszeit für Pertuzumab

Szenario 1: Für die analytische Bestimmung der medianen Gesamtüberlebenszeit der mit Ptz+T+D behandelten Patienten, die aus den Daten der Interimanalyse des OS (1. Datenschnitt vom Mai 2011) noch nicht belastbar geschätzt werden konnte, werden folgende Annahmen getroffen: (i) exponentiell verteilte Hazards, (ii) eine gerechtfertigte proportional Hazards-Annahme auch für den zukünftigen Verlauf der Überlebenskurven und (iii) ausreichend Beobachtungen für eine valide Hazard-Schätzung im Kontrollarm.

Das Hazard Ratio (HR) von Ptz+T+D vs. Pla+T+D ist unverzerrt und mit 0,66 belastbar geschätzt. Eine ebensolche Schätzung liegt für das mediane Überleben im Arm Pla+T+D vor (37,6 Monate). Die Bestimmung des medianen Überlebens im Arm Ptz+T+D ergibt sich unter Verwendung der Eigenschaften der Exponentialverteilung zu:

$\tilde{\mu}_2 = \frac{\tilde{\mu}_1}{HR} = \frac{37.6}{0.66}$ = 56,97 Monate.

Zwischenfazit: Für das mediane Überleben im Arm Pertuzumab+Trastuzumab+Docetaxel kann ein Wert um 57 Monate erwartet werden, was eine absolute Differenz der Mediane von 19,37 Monaten zugunsten von Pertuzumab+Trastuzumab+Docetaxel ergäbe.

Szenario 2: Eine weitaus konservativere Abschätzung für die Differenz der medianen Überlebenszeiten ergibt sich aus der Differenz der unteren Grenze der zweiseitigen 95 %-Konfidenzintervalle für das mediane Überleben, dass in beiden Armen zum 2. Datenschnitt bestimmt werden konnte. Die Abschätzung ist konservativ, da sich aufgrund der geringen Zahl der Ereignisse im Verum-Arm ein sehr weites Konfidenzintervall ergibt. Diese Werte liegen bei 34 Monaten unter Placebo+Trastuzumab+Docetaxel und 42 Monaten unter Pertuzumab+Trastuzumab+Docetaxel, was eine Differenz der Mediane von 8 Monaten ergibt.

Gesamtfazit: Eine beobachtete Differenz der medianen Überlebenszeiten zwischen den Behandlungsarmen lässt sich noch nicht angeben, da diese Ereigniszahlen im Verum-Arm noch nicht eingetreten sind. Basierend auf (i) einer sehr konservativen Abschätzung (Differenz der unteren Grenze der zweiseitigen 95%-Konfidenzintervalle) und (ii) einer Extrapolation der proportional hazards – Annahme für zukünftige Ereignisse liegt der zu erwartende mediane Überlebensvorteil zwischen 8 und 19,4 Monaten.

Die Kaplan-Meier-Kurve für das Gesamtüberleben zeigt eine frühe Separierung nach ca. 4 Monaten und einen Vorteil für die Patienten im Behandlungsarm unter Ptz+T+D, einen auch in der anschließenden Zeitperiode erhalten bleibt.

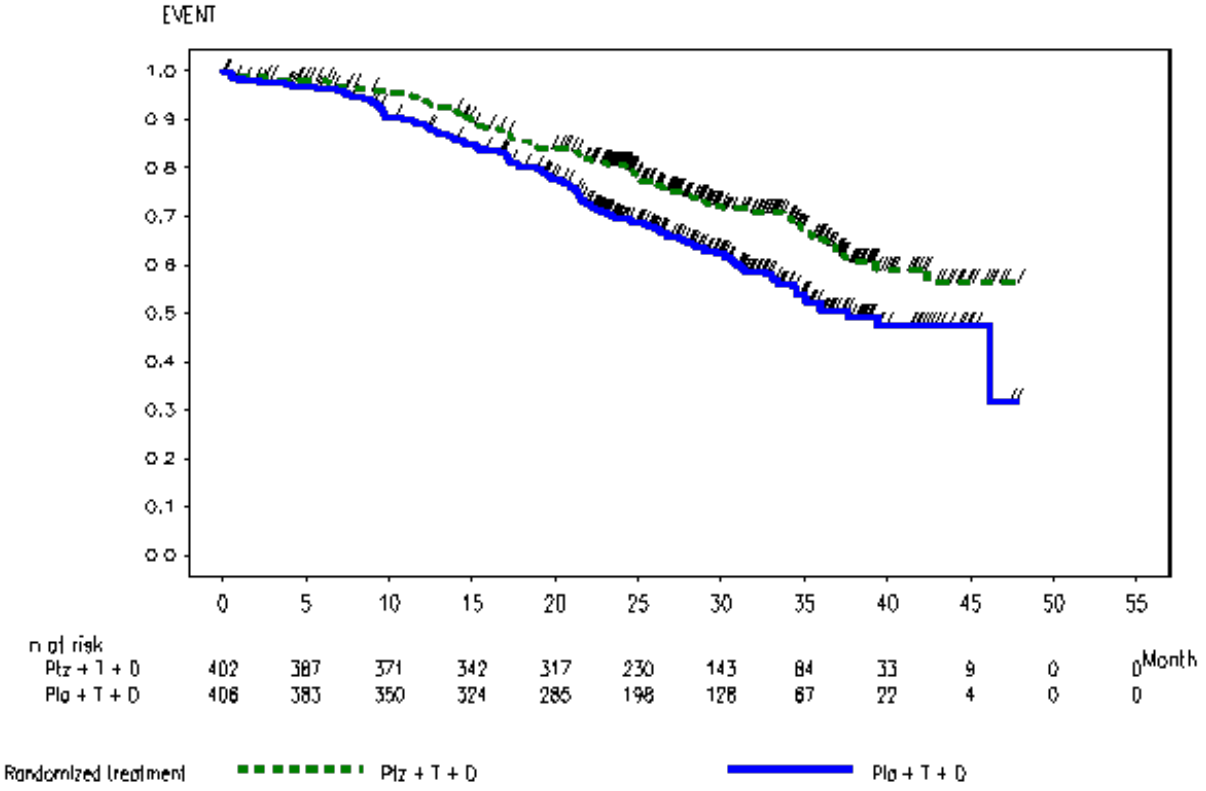


Abbildung 3: Kaplan-Meier Kurve zum Gesamtüberleben (ITT Population; CSR 12/2012 S.45)

Analyse der Sicherheits-Population:

Trotz der Zugrundelegung einer anderen Studienpopulation sind die Ergebnisse aus der Analyse der Safety-Population vergleichbar: Bis zum Zeitpunkt des klinischen Datenschnittes traten unter Pla+T+D Behandlungsarm traten mehr Todesfälle auf (152 Patienten, 38,4 %). Unter Ptz+T+D waren es 113 Patienten, 27,7 %). Die häufigste Todesursache war ein Fortschreiten der Erkrankung, welche häufiger unter Pla+T+D (136 Patienten, 34,3 %) als unter Ptz+T+D (100 Patienten, 24,5 %) auftrat. Für weitere Ergebnisse wird auf den Abschnitt 4.3.1.3.1.4 verwiesen.

4.3.1.3.1.2 <Morbidität> – RCT

4.3.1.3.1.2.1 <Primärer Endpunkt Progressionsfreies Überleben (PFS) bewertet durch IRF> – RCT

Tabelle 4-17: Operationalisierung von <Primärer Endpunkt Progressionsfreies Überleben bewertet durch IRF>

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	Der primäre Endpunkt war PFS, basierend auf der Tumorbewertung durch ein IRF. PFS war als Zeit zwischen der Randomisierung und der ersten Dokumentation einer Krankheitsprogression (PD) definiert, bestimmt durch das IRF auf der Basis der RECIST-Kriterien, Version 1.0, oder Tod jeglicher Ursache (innerhalb von 18 Wochen nach der letzten Tumorbewertung), je nachdem, welches Ereignis zuerst auftrat. Die Bewertung der PD basierte auf dem Review von radiografischen (NMR, CT, Knochenscan, Brust-Röntgen etc.) wie auch zytologischen (z.B. relevante zytologische Berichte zur Dokumentation von malignen pleuralen Effusionen, Knochenmark-Aspirationen, Liquor, etc.) Berichten und photographischer Dokumentation, wenn verfügbar. Bei Patienten mit multiplen, messbaren Läsionen, wurde ein Maximum von 5 Läsionen pro Organ und insgesamt 10 Läsionen, die als repräsentativ für alle betroffene Organe galten, als Ziel-Läsionen (*target lesions*) bestimmt und beim Screening ausgemessen. Weitere Läsionen sollten als nicht-Ziel-Läsionen eingeordnet werden. Das Vorhandensein bzw. das nicht-Vorhandensein (nicht jedoch die Ausmaße) jeder Läsion sollte während der Behandlungsperiode und der Nachbeobachtungszeit bis zum Auftreten einer Krankheits-progression notiert werden. Tumorbewertungen wurden alle 9 Wochen nach dem Datum der Randomisierung durchgeführt. Für Patienten, die aus anderen Gründen, als einem Todesfall oder IRF-bewerteter Progression die Studie beendeten, sollten möglichst auch weiterhin die geplanten Beurteilungstermine alle 9 Wochen für den Tumor wahrnehmen, bis ein Todesfall oder eine Krankheitsprogression eingetreten war. Zensierung: Für Patienten bei denen keine dokumentierte Krankheitsprogression vorlag, oder die nicht innerhalb 18 Wochen nach der letzten Tumorbewertung gestorben waren, wurde das PFS zum Zeitpunkt der letzter IRF basierten Tumorbewertung zensiert. Falls keine post-baseline Beurteilungen vorhanden waren, wurde das PFS zum Zeitpunkt der Randomisierung plus einem Tag zensiert. **RECIST [54]** **Messbare Erkrankung**: Vorhandensein mindestens einer messbaren Läsion. Liegt nur eine messbare Läsion vor, sollte die neoplastische Eigenschaft durch Zytologie / Histo-logie bestätigt worden sein. Messbare Läsionen, die nicht in mindestens einer Dimension mit dem längsten Durchmesser ≥ 20 mm mittels konventioneller Techniken ausgemessen werden konnten oder ≥ 10 mm, wurden mittels Spiral-CT, z.B. Knochenläsionen, oder CT Scan dokumentiert. **Nicht-Messbare Erkrankung**: Alle anderen Läsionen, inklusive kleiner Läsionen (längster Durchmesser < 20 mm mit konventionellen Methoden oder < 10 mm mit Spiral CT Scan), d.h. Knochenläsionen, Hirnhaut-Erkrankungen, Aszites, pleurale / perikardiale Effusion, entzündliche Brusterkrankungen, Lymphangitis, zystische Läsionen und auch abdominelle Massen, die nicht bestätigt waren, mussten durch bildgebende Verfahren abgeklärt werden. **Ansprechkriterien der Tumorbewertung:** **Evaluation der Zielläsionen** Komplettes Ansprechen (*Complete Response – CR*) — Verschwinden aller Zielläsionen

Partielles Ansprechen (*Partial Response – PR*)	Verminderung um mindestens 30 % für die Summe der längsten Durchmesser der Zielläsionen (als Referenz die Summe der längsten Durchmesser zu Studienbeginn (*baseline*))
Fortschreitende Erkrankung (*Progressive disease – PD*)	Erhöhung um mindestens 20 % für die Summe der längsten Durchmesser der Zielläsionen (als Referenz die Summe der kürzesten Durchmesser, die seit Behandlungsbeginn gemessen wurde)
Stabile Erkrankung (*Stable disease* – SD)	Weder ausreichende Verminderung, um als PR eingestuft zu werden, noch ausreichende Erhöhung, um als PD eingestuft zu werden (als Referenz die Summe der kürzesten Durchmesser, die seit Behandlungsbeginn gemessen wurde).
	Evaluation der Nicht-Zielläsionen
Komplettes Ansprechen (*Complete Response – CR*)	Verschwinden aller Nicht-Zielläsionen und Normalisierung der Tumormarker-Level
Partielles Ansprechen (*Partial Response – PR*)	Bestehenbleiben von einer oder mehreren Nicht-Zielläsionen und / oder Bestehenbleiben eines erhöhten Tumormarkerlevels über den Normalwerten
Fortschreitende Erkrankung (*Progressive disease – PD*)	Eindeutige Zunahme von bestehenden Nicht-Zielläsionen.

Kategorien der Tumorbewertung :

Zielläsion	**Nicht-Zielläsion**	**Neue Läsion**	**Gesamtantwort („*overall response*“)**
CR	CR	Nein	CR
CR	Inkomplette Antwort/SD	Nein	PR
PR	Non-PD	Nein	PR
SD	Non-PD	Nein	SD
PD	Any	Ja oder Nein	PD
Any	PD	Ja oder Nein	PD
Any	Any	Ja	PD
NA	Non-PD	Nein	NA
Non-PD	NA	Nein	NA

NA = Nicht einschätzbar; CR = *complete response* (komplette Antwort); PR = *partial response* (partielle Antwort); SD = *stable disease* (stabile Erkrankung); PD = *progressive disease* (Krankheitsprogression)

Tabelle 4-18: Bewertung des Verzerrungspotenzials für <Primärer Endpunkt Progressionsfreies Überleben bewertet durch IRF> in RCT mit dem zu bewertenden Arzneimittel (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011)

Studie	**Verzerrungspotenzial auf Studienebene**	**Verblindung Endpunkterheber**	**Adäquate Umsetzung des ITT-Prinzips**	**Hinweise auf ergebnisgesteuerte Berichterstattung**	**Sonstige das Verzerrungspotenzial beeinflussende Punkte**	**Verzerrungspotenzial Endpunkt**
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Bei niedrigem Verzerrungspotential auf Studienebene, wurde das Verzerrungspotential für den primären Endpunkt <PFS, bewertet durch IRF (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011)> wie folgt bewertet: Die Endpunkterheber und die Behandler waren hinsichtlich dieses Endpunktes verblindet. Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Alle Befunde wurden durch das IRF beurteilt, was auch die verblindete Erhebung der Endpunkte sicherstellte. Die Bewertung einer Krankheitsprogression erfolgte mittels RECIST-Kriterien. Das ITT-Prinzip wurde adäquat umgesetzt: alle randomisierten Patienten gingen in die Auswertung ein. Hinweise auf ergebnisgesteuerte Berichterstattung ergaben sich beim Abgleich mit der Publikation der Studie nicht. [49] Auch sonstige, das Verzerrungspotential beeinflussende Punkte wurden nicht beobachtet.

Da die Analyse des primären Endpunktes <PFS, bewertet durch IRF> entsprechend dem Studienprotokoll als finale Analyse zum Zeitpunkt des ersten Datenschnitts geplant war, d.h. damit auch unter Verblindung der Patienten, Ärzte und Endpunkterheber erfolgte, wurde das Verzerrungspotential auf Endpunktebene mit „niedrig“ beurteilt.

Tabelle 4-19: Ergebnisse für<Primärer Endpunkt Progressionsfreies Überleben bewertet durch IRF> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T+D Hazard Ratio [95% KI] p-Wert**	**Ptz+T+D vs Pla+T+D Zeit bis zur Progression in Monaten; Median p-Wert**
	Patienten mit Ereignis [a] n/N (%)	**Zeit bis zur Progression in Monaten; Median [95% KI]**	**Patienten mit Ereignis [a] n/N (%)**	**Zeit bis zur Progression in Monaten; Median [95% KI]**		
Primäre Analyse (IRF basiert)						
Progressionsfreies Überleben bewertet durch IRF [b]	242 / 406 (59,6 %)	12,4 [10; 13]	191 / 402 (47,5 %)	18,5 [15;23]	0,62 [0,51; 0,75] p < 0,0001	< 0,0001

a: Ereignis bezieht sich hier auf das Auftreten einer Progression oder Tod, alle Ereignisse entsprechen der ITT Population

b: Die Ergebnisse sind nach vorherigem Behandlungsstatus und Region stratifiziert. Diese Tabelle stellt die Ergebnisse der Primäranalyse dar.

Datenquelle: Tabelle 16 aus CSR 10 / 2011

Log-Rank Test stratifiziert nach vorherigem Behandlungsstatus und Region, zweiseitiger Test auf 5 % Niveau

Die 1. Interimanalyse (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011) wurde wie geplant nach Auftreten von 381 IRF-bewerteten PFS Ereignissen durchgeführt, danach waren keine weiteren Bewertungen durch das IRF geplant, da diese Auswertung den finalen Analysezeitpunkt für das IRF-bewertete PFS darstellt.

Zum Zeitpunkt des Datenschnittes für die primäre Analyse des PFS, waren insgesamt 433 IRF-bewertete PFS Ereignisse aufgetreten, 242 (59,6 %) im Pla+T+D Behandlungsarm und 191 (47,5 %) im Ptz+T+D Behandlungsarm.

Das IRF-bewertete PFS wurde für Patienten im Behandlungsarm Ptz+T+D im Vergleich zu Patienten, die Pla+T+D bekamen, signifikant verbessert (HR = 0,62, $p < 0.0001$) bei einem Anstieg des PFS im Median von 6,1 Monaten (Median 12,4 für Pla+T+D im Vergleich zu 18,5 Monaten für Ptz+T+D).

Die stratifizierten IRF-PFS Ergebnisse wurden von der unstratifizierten Analyse bestätigt (HR = 0,63 [95% KI: 0,52; 0,76]; $p < 0,0001$) (S.1103 im CSR).

Hinsichtlich des primären Endpunktes PFS waren innerhalb der CLEOPATRA-Studie folgende Sensitivitätsanalysen geplant (CLEOPATRA CSR 10 / 2011, S. 78):

- Ein möglicher Unterschied zwischen den verschiedenen Bewertungsverfahren durch Investigator und IRF wurde analysiert. Zensierungsregel: Das früheste Datum der Dokumentation einer Krankheitsprogression (PD) durch das IRF bzw. den Investigator wurde als Datum für das Auftreten eines PFS-Ereignisses zu Grunde gelegt.
- Zensierung zum Zeitpunkt des Beginns der nächsten Linie der antitumorösen Therapie (*next-line anti-cancer therapy* (NACT)): Patienten, die mit einer NACT vor der IRF-Bewertung einer Progression (innerhalb von 18 Wochen nach der letzten Tumorbewertung bzw. der letzten IRF-verfügbaren Tumorbewertung begannen), wurden zum Datum der letzten verfügbaren IRF-Tumorbewertung vor dem Start einer NACT zensiert.
- Einfluss der Bewertung des PFS durch das IRF auf die Behandlung. Nur IRF-bewertete PFS Ereignisse, die nicht später als 42 Tage nach der letzten Administration jeglicher Studienmedikation stattfanden, wurden in die Analyse einbezogen.
- Potentieller Bias durch unterschiedliche Zeitintervalle zwischen den Tumorbewertungen aufgrund verpasster Visiten: Eine fehlende Bewertung (oder die früheste fehlende Bewertung im Fall mehrerer fehlender Bewertungen) wurden durch eine Bewertung der Krankheitsprogression (PD) ersetzt und die Zeit bis zum Ereignis als der erwartete Tag der fehlenden Bewertung eingesetzt. Fehlte keine Bewertung, wurde das erste Datum, an dem eine PD durch das IRF bewertet wurde, als Datum für das Ereignis herangezogen. Im Fall eines Todes vor dem Auftreten einer PD innerhalb von 18 Wochen nach der letzten Tumorbewertung wurde, falls die letzte Tumorbewertung vor dem Tod fehlte, die fehlende Bewertung durch eine Bewertung der PD ersetzt.
- Zeitpunkt des Todes, inklusive aller Todesfälle, die mehr als 18 Wochen nach der letzten Tumorbewertung auftraten.

- Kontrolle hinsichtlich Behandlungsabbruch aufgrund von Toxizität: Patienten, die sämtliche Studienbehandlung aufgrund Toxizität abbrachen, wurden vom Zeitpunkt der letzten IRF-Bewertung an oder vor dem Beginn des Behandlungsabbruchs zensiert.

Alle Sensitivitätsanalysen basierten auf der ITT-Population und waren nach vorherigem Behandlungsstatus und Region stratifiziert.

In allen Sensitivitätsanalysen wurde PFS für die Patienten, die Ptz+T+D erhielten im Vergleich zu Patienten die Pla+T+D bekamen, signifikant verbessert. Diese Ergebnisse unterstützen die Ergebnisse der primären Analyse.

Das HR für jede Sensitivitätsanalyse war mit dem Gesamt-HR konsistent und reichte von 0,58 bis 0,66 (alle $p < 0,0001$).

Die Ergebnisse der Analyse des PFS, festgestellt durch den Investigator, sind konsistent mit den Resultaten, die auf dem IRF-bewerteten PFS beruhen. Zum Zeitpunkt des klinischen Datenschnittes war ein Investigator-bewertetes PFS Ereignis bei 451 Patienten eingetreten (250 Patienten [61,6 %] im Pla+T+D Behandlungsarm und 201 Patienten [50,0 %] im Ptz+T+D Behandlungsarm). Das Investigator-bewertete PFS wurde für Patienten unter Ptz+T+D im Vergleich zu Patienten unter Pla+T+D signifikant verbessert (HR = 0,65; $p < 0,0001$; Median 12,4 Monate für Pla+T+D bzw. 18,5 Monate für Ptz+T+D) (S.1143-1144 im CSR 10 / 2011).

Die Häufigkeit einer Krankheitsprogression (PD), die ausschließlich aufgrund symptomatischer Verschlechterung diagnostiziert wurde, war mit neun Patienten (2,2 %) im Pla+T+D Behandlungsarm und einem Patient (0,2 %) im Ptz+T+D Behandlungsarm gering (S.1142 im CSR 10 / 2011).

Tabelle 4-20: Ergebnisse für Sensitivitätsanalysen zum Primären Endpunkt Progressionsfreies Überleben bewertet durch IRF während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T+D Hazard Ratio [95% KI] p-Wert**	**Ptz+T+D vs Pla+T +D Zeit bis zur Progression in Monaten; Median p-Wert**
	Patienten mit Ereignis [a] n/N (%)	**Zeit bis zur Progression in Monaten; Median [95% KI]**	**Patienten mit Ereignis [a] n/N (%)**	**Zeit bis zur Progression in Monaten; Median [95% KI]**		
Sensitivitätsanalysen:						
1. Progressionsfreies Überleben bewertet durch IRF [b]	274 / 406 (67,5 %)	10,4 [9; 12]	226 / 402 (56,2 %)	14,6 [13; 17]	0,66 [0,55; 0,79] p < 0,0001	< 0,0001
2. Progressionsfreies Überleben bewertet durch IRF [c]	231 / 406 (56,9 %)	12,3 [10; 13]	173 / 402 (43,0 %)	18,7 [16; 25]	0,58 [0,48; 0,71] p < 0,0001	< 0,0001
3. Progressionsfreies Überleben bewertet durch IRF [d]	220 / 406 (54,2 %)	12,4 [10; 14]	165 / 402 (41,0 %)	20,8 [17; 25]	0,58 [0,47; 0,71] p < 0,0001	< 0,0001
4. Progressionsfreies Überleben bewertet durch IRF [e]	242 / 406 (59,6 %)	12,3 [10; 13]	191 / 402 (47,5 %)	18,5 [15; 23]	0,62 [0,51; 0,75] p < 0,0001	< 0,0001
5. Progressionsfreies Überleben bewertet durch IRF [f]	251 / 406 (61,8 %)	12,4 [10, 13]	200 / 402 (49,8 %)	17,2 [15; 22]	0,63 [0,52; 0,76] p < 0,0001	< 0,0001
6. Progressionsfreies Überleben bewertet durch IRF [g]	237 / 406 (58,4 %)	12,3 [10; 13]	184 / 402 (45,8 %)	18,5 [15; 23]	0,61 [0,50; 0,74] p < 0,0001	< 0,0001
PFS Analyse (Investigator)						
Progressionsfreies Überleben festgestellt durch *Investigator* [h]	250 / 406 (61,6 %)	12,4 [10; 13]	201 / 402 (50,0 %)	18,5 [16; 21]	0,65 [0,54; 0,78] < 0,0001	< 0,0001

a: Ereignis bezieht sich hier auf das Auftreten einer Progression oder Tod, alle Ereignisse entsprechen der ITT Population

b: Sensitivitätsanalyse 1; Das früheste Datum der Dokumentation einer PD durch das IRF bzw. den Investigator wurde als Datum für das Auftreten eines PFS-Ereignisses zu Grunde gelegt; Tabelle aus CSR 10 / 2011 S. 1105

c: Sensitivitätsanalyse 2; Patienten, die mit einer NACT vor der IRF-Bewertung einer Progression (innerhalb von 18 Wochen nach der letzten Tumorbewertung) bzw. der letzten IRF-verfügbaren Tumorbewertung begannen, wurden zum Datum der letzten verfügbaren IRF-Tumorbewertung vor dem Start einer NACT zensiert. Tabelle

aus CSR 10 / 2011 S. 1106,

d: Sensitivitätsanalyse 3; nur IRF-bewertete PFS Ereignisse, die nicht später als 42 Tage nach der letzten Dosisverabreichung stattfanden, wurden in die Analyse einbezogen; Tabelle aus CSR 10 / 2011 S. 1107,

e: Sensitivitätsanalyse 4; Fehlende Bewertungen (oder die früheste fehlende Bewertung im Fall von mehreren fehlenden Bewertungen) wurde durch eine Bewertung des PD ersetzt und die Zeit bis zum Ereignis wurde auf den Tag der erwarteten fehlenden Bewertung gesetzt. Fehlten keine Visiten, wurde das erste PD-Datum, das vom IRF bewertet wurde, als Datum für das Ereignis gewertet. Im Fall eines Todes vor einer PD innerhalb von 18 Wochen nach der letzten Tumorbewertung, wo vor dem Todesfall die letzte Visite fehlte, wurde der Zeitpunkt dieser fehlenden Visite als Zeitpunkt für das Ereignis gewertet. Tabelle aus CSR 10 / 2011 S. 1108,

f: Sensitivitätsanalyse 5; inklusive aller Todesfälle, die mehr als 18 Wochen nach der letzten Tumorbewertung auftraten.; Tabelle aus CSR 10 / 2011 S. 1109,

g: Sensitivitätsanalyse 6; Patienten die aufgrund von Toxizität alle Studienmedikation abgebrochen haben wurden zu Ihrer letzten IRF Tumordokumentation zensiert oder vor Ihrem Behandlungsabbruchsdatum; Tabelle aus CSR 10 / 2011 S. 11010

h: Die Ergebnisse sind nach vorherigem Behandlungsstatus und Region stratifiziert. Diese Tabelle stellt die Ergebnisse der Primäranalyse dar. Tabelle 22 aus CSR 10 / 2011. Investigator-bewertetes PFS war als Zeit zwischen dem Tag der Randomisierung bis zum ersten dokumentierten radiografischen PD, bestimmt durch den Investigator unter Verwendung der RECIST-Kriterien oder Tod jeglicher Ursache definiert, je nachdem welches Ereignis zuerst auftrat.

4.3.1.3.1.2.2 <Endpunkt Gesamtansprechrate> – RCT

Tabelle 4-21: Operationalisierung von <Endpunkt Gesamtansprechrate>

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	Das Gesamtansprechen, die *best overall response rate (RECIST) for objective response* war als komplettes Ansprechen (*complete response* [CR]), oder partielles Ansprechen (*partial response* [PR]) definiert, bestimmt durch das IRF auf der Basis von RECIST-Kriterien an zwei aufeinanderfolgenden Untersuchungen mit einem Abstand von mindestens 4 Wochen (Patienten ohne messbare Krankheit oder mit einer nur auf die Knochen lokalisierten Krankheit wurden nicht in diese Analyse eingeschlossen). **RECIST [54]** **Messbare Erkrankung** : Vorhandensein mindestens einer messbaren Läsion. Liegt nur eine messbare Läsion vor, sollte die neoplastische Eigenschaft durch Zytologie / Histologie bestätigt worden sein. Messbare Läsionen, die nicht in mindestens einer Dimension mit dem längsten Durchmesser ≥ 20 mm ausgemessen werden konnten mittels konventioneller Techniken oder ≥ 10 mm, Dokumentation mittels Spiral-CT, z.B. Knochenläsionen, CT Scan erforderlich. **Nicht-Messbare Erkrankung** : alle anderen Läsionen, inklusive kleinen Läsionen (längster Durchmesser < 20 mm mit konventionellen Methoden oder < 10 mm mit Spiral CT Scan), i.e. Knochenläsionen, Hirnhaut-Erkrankungen, Aszites, pleurale / perikardiale Effusion, entzündliche Brusterkrankungen, Lymphangitis, zystische Läsionen, und auch abdominelle Massen, die nicht bestätigt waren, mussten durch bildgebende Verfahren abgeklärt werden. **Komplettes und partielles Ansprechen** sollten 4 bis 6 Wochen nach erstmaliger Beobachtung des Ansprechens bestätigt werden (28 und 42 Tage). Im Fall stabiler Erkrankung (stable diesease (SD)), musste in den Tumorbewertungen das entsprechende RECIST-Kriterium erfüllt sein, mit einem Minimum-Intervall von 6 Wochen. Komplettes und partielles Ansprechen sollten 4 bis 6 Wochen nach erstmaliger Beobachtung des Ansprechens bestätigt werden (28 und 42 Tage). Im Fall stabiler Erkrankung (stable diesease (SD)), musste in den Tumorbewertungen das entsprechende RECIST-Kriterium erfüllt sein, mit einem Minimum-Intervall von 6 Wochen. **Ansprechkriterien der Tumorbewertung :** **Evaluation der Zielläsionen** Komplettes Ansprechen (*Complete Response – CR*): Verschwinden aller Zielläsionen Partielles Ansprechen (*Partial Response – PR*): Verminderung um mindestens 30 % für die Summe der längsten Durchmesser der Zielläsionen (als Referenz die Summe der längsten Durchmesser zu Studienbeginn (*baseline*)) Fortschreitende Erkrankung (*Progressive disease – PD*): Erhöhung um mindestens 20 % für die Summe der längsten Durchmesser der Zielläsionen (als Referenz die Summe der kürzesten Durchmesser, die seit Behandlungsbeginn gemessen wurde) Stabile Erkrankung (*Stable disease* – SD): Weder ausreichende Verminderung, um als PR eingestuft zu werden, noch ausreichende Erhöhung, um als PD eingestuft zu werden (als Referenz die Summe der kürzesten Durchmesser, die seit Behandlungsbeginn gemessen wurde). **Evaluation der Nicht-Zielläsionen** Komplettes Ansprechen (*Complete Response – CR*): Verschwinden aller Nicht-Zielläsionen und Normalisierung der Tumormarker-Level

Partielles Ansprechen (*Partial Response – PR*)	Bestehenbleiben von einer oder mehreren Nicht-Zielläsionen und/oder Bestehenbleiben eines erhöhten Tumormarkerlevels über den Normalwerten
Fortschreitende Erkrankung (*Progressive disease – PD*)	Eindeutige Zunahme von bestehenden Nicht-Zielläsionen.

Kategorien der Tumorbewertung :

Zielläsion	**Nicht-Zielläsion**	**Neue Läsion**	**Gesamtantwort („*overall response*“)**
CR	CR	Nein	CR
CR	Inkomplette Antwort/SD	Nein	PR
PR	Non-PD	Nein	PR
SD	Non-PD	Nein	SD
PD	Any	Ja oder Nein	PD
Any	PD	Ja oder Nein	PD
Any	Any	Ja	PD
NA	Non-PD	Nein	NA
Non-PD	NA	Nein	NA

NA = *Nicht einschätzbar*; CR = *complete response* (komplettes Ansprechen); PR = *partial response* (partielles Ansprechen); SD = *stable disease* (stabile Erkrankung); PD = *progressive disease* (Krankheitsprogression)

Tabelle 4-22: Bewertung des Verzerrungspotenzials für <Endpunkt Gesamtansprechrate>in RCT mit dem zu bewertenden Arzneimittel (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011)

Studie	**Verzerrungspotenzial auf Studienebene**	**Verblindung Endpunkterheber**	**Adäquate Umsetzung des ITT-Prinzips**	**Hinweise auf ergebnisgesteuerte Berichterstattung**	**Sonstige das Verzerrungspotenzial beeinflussende Punkte**	**Verzerrungspotenzial Endpunkt**
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Bei einem auf Studienebene als „niedrig“ eingestuften Verzerrungspotential wurde das Verzerrungspotential für den Endpunkt <Gesamtansprechrate (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011)> folgendermaßen eingeschätzt: Die Endpunkterheber waren verblindet, die Bewertung durch das IRF erfolgte auf der Basis klinischer wie auch

apparativer Befunde (1. Datenschnitt Mai 2011). Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Die Beurteilung einer Krankheitsprogression erfolgte auf Basis der RECIST-Kriterien. Ausgewertet wurden nur randomisierte Patienten mit messbarer Erkrankung zu Studienbeginn - entsprechend der Vorgaben der RECIST-Kriterien. In der Gesamtgruppe hatten 88,6 % der Patienten eine messbare Erkrankung. Somit kann das ITT-Prinzip als adäquat umgesetzt bewertet werden (weniger als 15 % der randomisierten Patienten wurden nicht ausgewertet). Der Abgleich mit der Publikation der Studie ergab keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. [49]

Auch sonstige, das Verzerrungspotential beeinflussende Punkte, wurden nicht beobachtet. Daher wurde das Verzerrungspotential für den Endpunkt <Gesamtansprechrate> mit „niedrig“ beurteilt.

Reduzierte ITT bei dem Endpunkt Gesamtansprechrate

Für den Endpunkt Gesamtansprechrate ist die ITT-Population kleiner (679 randomisierte Patienten, davon 336 Pla+T+D und 343 Ptz+T+D) als die bei anderen Endpunkte (808 randomisierte Patienten, davon 406 Pla+T+D und 402 Ptz+T+D). Der Hintergrund für diese Verkleinerung der ITT-Population ist, dass für diese bestimmte Teilpopulation (i.e. Patienten mit nicht messbarer Erkrankung, Patienten nur mit Knochenmetastasen) dieser Endpunkt methodisch gar nicht bestimmt werden kann.

Diese Reduktion der ITT-Population ist damit keine Verletzung des ITT-Prinzips, da für alle Endpunkte jeweils alle randomisierten Patienten herangezogen wurden, für die der Endpunkt technisch bestimmt werden konnte. Da in beiden Behandlungsarmen etwa gleichviele nicht-viszerale Patienten waren (17% Pla+T+D, 15% Ptz+T+D) ergibt sich durch die Reduktion auch keine Imbalance zwischen den Behandlungsarmen.

Die Reduktion der ITT-Population für den Endpunkt Ansprechrate hat damit keinen Einfluss auf das Verzerrungspotential, es bleibt bei ‚niedrig‘.

Tabelle 4-23: Ergebnisse für<Endpunkt Gesamtansprechrate> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (konfirmatorische finale Analyse, erster Datenschnitt Mai 2011)

	Placebo+Trastuzumab+Docetaxel	**Pertuzumab+Trastuzumab+Docetaxel**	**Ptz+T+D vs Pla+T +D Gesamtansprechrate Absolute Differenz [95% KI] p-Wert**	**Ptz+T+D vs Pla+T +D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**
	Patienten mit Ereignis [a] n / N (%) [95% KI]	**Patienten mit Ereignis [a] n / N (%) [95% KI]**			
Gesamtansprechrate	233 / 336 (69,3 %) [64,1; 74,2]	275 / 343 (80,2 %) [75,6; 84,3]	10,83 [4,2; 17,5] p = 0,0011	1,16 [1,06; 1,26]	0,0012

a: n = Ereignis bezieht sich hier auf die Anzahl *Responder*, die durch IRF bewertet wurden, basierend auf RECIST. N = bezieht sich auf die Patienten, bei denen eine durch IRF bewertete, feststellbare Erkrankung an der Baseline vorlag. Definition einer messbaren Erkrankung siehe Tabelle 4-21

Bei der Auswertung wurde nach vorherigem Behandlungsstatus und Region stratifiziert. Die Auswertung beruht auf der ITT Population.

Datenquelle: Tabelle 23 aus CSR 10 / 2011 + zusätzliche Analysen der Roche Pharma AG [55]

Mantel-Haenszel Chi Quadrat Test, zweiseitig auf 5% Niveau

Der 1. Datenschnitt (finale konfirmatorische Analyse, Mai 2011) wurde wie geplant nach Auftreten von 381 IRF-bewerteten PFS Ereignissen durchgeführt. Entsprechend dem Studienprotokoll entsprach dies der finalen Auswertung der Gesamtansprechrate, weshalb keine weiteren Auswertungen der Gesamtansprechrate über diesen Zeitpunkt hinaus vorliegen.

Die Gesamtansprechrate (best *overall response rate*, PR oder CR), festgestellt durch das IRF, wurde bei den Patienten der ITT-Population bewertet, die eine durch das IRF bestimmte, messbare Erkrankung zu Studienbeginn hatten. Über 80 % der Patienten in jedem Behandlungsarm waren in dieser Analyse eingeschlossen (336 Patienten im Pla+T+D Behandlungsarm und 343 Patienten im Ptz+T+D Behandlungsarm).

Die IRF-bewerteten Gesamtansprechrate weist eine statistisch signifikante Differenz von fast 11 % zwischen den Behandlungsarmen auf (275 der Patienten unter Ptz+T+D (80,2 %) im Vergleich zu 233 der Patienten unter Pla+T+D (69,3 %), p = 0,0012, die ein Gesamtansprechen erreichten (stratifizierte Analyse)). Die unstratifizierte Analyse bestätigte dieses Ergebnis (p = 0,0014, CSR 10 / 2011 S. 1606). Ein komplettes Ansprechen wurde bei 14 Patienten (4,2 %) im Pla+T+D Behandlungsarm und bei 19 Patienten (5,5 %) im Ptz+T+D Behandlungsarm erreicht - ein partielles Ansprechen bei 219 (65,2 %) vs. 256 Patienten (74,6 %). Bei der Investigator bewerteten Ansprechrate wurde bei 253 Patienten unter Pla+T+D und 284 unter Ptz+T+D dieses Ereignis festgestellt (p = 0,0049 für den Unterschied der Ansprechrate, stratifizierte Analyse (CSR 10 / 2011 S. 1604).

4.3.1.3.1.3 <Lebensqualität> – RCT

4.3.1.3.1.3.1 <Endpunkt Zeit bis Symptomprogression> – RCT

Tabelle 4-24: Operationalisierung von <Endpunkt Zeit bis Symptomprogression>

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	Die Zeit bis Symptomprogression war als Zeit zwischen der Randomisierung und der ersten Symptomprogression definiert, gemessen unter Verwendung des FACT-B Fragebogens (Functional Assessment of Cancer Therapy-Breast Cancer) mittels TOI-PFB (Trial Outcomes Index-Physical / Functional / Breast) - einer 24-Item Skala, die sich aus drei Abschnitten aus dem FACT-B Fragebogen (*physical well-being, functional well-being, und additional concerns BCS*) zusammensetzt. [56] Alle Items des Fragebogens werden durch die Patientin auf einer 5-Punkte-Skala von 0 („überhaupt nicht“) bis 4 („sehr hoch“) beurteilt. Ein höherer Wert der Subskala ist ein Indikator für eine als besser empfundene Lebensqualität. Eine Verminderung um fünf Punkte ist als klinisch relevant definiert und somit als Schwelle für eine Symptomprogression. [57] Ausschließlich weibliche Patientinnen füllten den Fragebogen alle 9 Wochen aus (innerhalb von drei Tagen vor jeder Tumorbewertung) bis zur IRF-bestimmten Krankheitsprogression (PD). Patientinnen sollten den Bogen immer zum geplanten Zeitpunkt ausfüllen, auch wenn die Studienmedikation aufgrund von Toxizität oder Investigator-bewertetem PD (Bewertungen wurden zum IRF-bestimmten Krankheitsprogression gestoppt) nicht mehr gegeben wurde. Für die Bewertung wurden dieselben Kriterien angewendet, wie für die primäre Wirksamkeitsanalyse (PFS bestimmt durch IRF; *IRF-determined PFS*). Eine *Time to event* (Zeit bis zum Ereignis) Analyse mit der Kaplan-Meier Methode wurde zur Testung der Hypothese, dass die Zugabe von Pertuzumab zum „*standard of care*“ keinen bedeutenden Einfluss auf die Lebensqualität hat, durchgeführt. <u>Zensierung</u>: Daten von Patienten, die keine beobachtete Symptomprogression hatten, wurden mit dem letzten aufgenommenen Datum der Beurteilung mittels TOI-PFB zensiert. Wenn kein Baseline-Wert für den TOI-PFB vorlag oder wenn keine Beurteilungen nach der Baseline für den TOI-PFB vorlagen, wurden die Daten mit dem Tag der Randomisierung + 1 Tag zensiert.

Tabelle 4-25: Bewertung des Verzerrungspotenzials für <Endpunkt Zeit bis Symptom-progression>in RCT mit dem zu bewertenden Arzneimittel (1. Datenschnitt Mai 2011)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Auf Studienebene wurde das Verzerrungspotential als „niedrig“ eingestuft. Für die <Zeit bis Symptomprogression> (1. Datenschnitt Mai 2011) wurde dies wie folgt beurteilt: Die Patientinnen, die den Fragebogen ausfüllten, waren verblindet, die Auswertung erfolgte

ebenfalls verblindet. Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Alle randomisierten Patientinnen gingen in die Auswertung ein, das ITT-Prinzip wurde adäquat umgesetzt. In der Publikation der Studie wurde die Erfassung der Lebensqualität nicht erwähnt, im Studienbericht fanden sich keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. [49]

Da dieser Fragebogen nur zur Anwendung bei Frauen vorgesehen ist, vervollständigten nur weibliche Patientinnen den FACT-B Fragebogen. Somit wurden 806 Patientinnen von 808 randomisierten Patienten, d.h. 99,8% aller randomisierten Studienteilnehmer, in die Analysepopulation für die FACT-B Analyse einbezogen. Das Verzerrungspotential für den Endpunkt <Zeit bis Symptomprogression> wurde deshalb – geltend für die verblindet durchgeführten Studienperioden 1 und 2 - mit „niedrig“ beurteilt. [57] Der FACT-B besteht insgesamt aus 5 Subskalen, die bei der Auswertung zu unterschiedlichen Endpunkten zusammengefasst werden. [16;56] Dies sind die generischen Skalen:

- Physical Well-Being (PWB)
- Social/Family Well-Being (SWB)
- Emotional Well-Being (EWB)
- Functional Well-Being (FWB)

sowie die krankheitsspezifische

- Breast Cancer Subscale (BCS)

Alle Subskalen zusammen bilden den FACT-B Gesamtscore. Die Validierung der Version 3 des FACT-B ist durch Brady dokumentiert. [58] Der FACT-G – bestehend aus den 4 generischen Subskalen des FACT-B – wurde von Version 3 auf die Version 4 validiert. [59]

In der CLEOPATRA-Studie wurden zur Messung der Lebensqualität zwei Endpunkte basierend auf dem FACT-B verwendet:

- Zeit bis zur Symptomprogression (TOI-PFB: Bestehend aus den 3 Subskalen PWB, FWB und BCS (24 Fragen) des FACT-B).
- Zeit bis zur Verschlechterung des BCS-Score (BSC-Subskala des FACT-B bestehend aus 10 Fragen).

In der CLEOPATRA-Studie kommt bereits die Version 4 des FACT-B zur Anwendung. Diese ist gegenüber der Version 3 unter anderem durch das Löschen der experimentellen Bewertungsfragen der Subskalen und durch eine zusätzliche Frage im BCS-Score gekennzeichnet, wobei die experimentellen Bewertungsfragen nicht in das Scoring eingehen.

Zur Bewertung der klinisch relevanten Unterschiede (MID) bzw. Veränderungen (MIC) Dabei sind aus dem FACT-B – Skala unterschiedliche Einzelskalen abzuleiten. Für die in der CLEOPATRA-Studie verwendeten Skalen wurden folgende Werte festgelegt:

- TOI-PFB: 5-6 Punkte
- BCS-Score: 2-3 Punkte

Diese Werte wurden für die Version 2 des FACT-B bestimmt. [57] Gegenüber der Version 2 kam bei der, in der CLEOPATRA-Studie eingesetzten, Version 4 zum Score betreffend TOI-PFB und folglich auch BCS-Score lediglich 1 Frage hinzu. Diese Veränderungen der in der CLEOPATRA-Studie verwendeten Version 4 gegenüber den Versionen 2 und 3 sind damit als geringfügig hinsichtlich des Einflusses auf die MID anzusehen.

Tabelle 4-26: Ergebnisse für <Endpunkt Zeit bis Symptomprogression> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (1. Datenschnitt Mai 2011)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T +D Hazard Ratio [95% KI] p-Wert**	**Ptz+T+D vs Pla+T +D Zeit bis zur Progression Median in Wochen p-Wert**
	Patientinnen mit Ereignis n / N (%)	**Zeit bis zur Progression in Wochen Median [95% KI]**	**Patientinnen mit Ereignis n / N (%)**	**Zeit bis zur Progression in Wochen Median [95% KI]**		
Zeit bis zur Symptomprogression	229 / 404 (56,7 %)	18,3 [18; 27]	239 / 402 (59,5 %)	18,4 [18; 27]	0,97 [0,81; 1,16] p = 0,7182	0,7161
Die Ergebnisse sind nach vorherigem Behandlungsstatus und Region stratifiziert. ITT Population. Log Rank Test Es wurden nur weibliche Patienten befragt (und ausgewertet). Datenquelle: Tabelle 25 aus CSR 10 / 2011 und zusätzliche Analysen [55].						

Wie geplant vervollständigten nur weibliche Patientinnen den FACT-B Fragebogen. Somit wurden 806 Patientinnen von 808 randomisierten Patienten, d.h. 99,8% aller randomisierten Studienteilnehmer, in der Analysepopulation für die FACT-B Analyse (404 Patientinnen im Pla+T+D Behandlungsarm und 402 Patientinnen im Ptz+T+D Behandlungsarm) eingeschlossen. Eine Symptomprogression entsprechend FACT-B wurde als eine Abnahme von 5 Punkten oder mehr nach Studienbeginn im TOI-PFB Index definiert. [57]

Es ergab sich kein Anhaltspunkt, dass die zusätzliche Gabe von Pertuzumab zu Trastuzumab und Docetaxel zu einem nachteiligen Effekt auf die Lebensqualität führte: 229 Patientinnen (56,7 %) im Pla+T+D Behandlungsarm und 239 (59,5 %) im Ptz+T+D Behandlungsarm zeigten während der Studie eine Symptomprogression basierend auf TOI-PFB: HR 0,97; 95% KI [0,81; 1,16]) (1. Datenschnitt Mai 2011)

4.3.1.3.1.3.2 <Endpunkt Zeit bis zur Verschlechterung des BCS-Score> – RCT

Tabelle 4-27: Operationalisierung von <Endpunkt Zeit bis zur Verschlechterung des BCS-Score>

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	Die Zeit bis zur Verschlechterung war als Zeit zwischen der Randomisierung und der ersten Verschlechterung definiert, gemessen unter Verwendung des FACT BCS *(Breast Cancer Scale)* - einer 10-Item Subskala, die aus dem letzten Abschnitt aus dem FACT-B (*Functional Assessment of Cancer Therapy-for Patients mit Breast Cancer)* Fragebogen (*additional concerns*) besteht. [56] Eine Verminderung um zwei Punkte ist als klinisch relevant definiert und somit als Schwelle für eine Verschlechterung. [57]. Ausschließlich weibliche Patientinnen füllten den Fragebogen alle 9 Wochen aus (innerhalb von drei Tagen vor jeder Tumorbewertung) bis zur IRF-bestimmten Krankheitsprogression (PD). Patientinnen sollten den Bogen immer zum geplanten Zeitpunkt ausfüllen, auch wenn die Studienmedikation aufgrund von Toxizität oder Investigator-bewertetem PD (Bewertungen wurden zum IRF-bewerteten PD gestoppt) nicht mehr gegeben wurde. Für die Bewertung wurden dieselben Kriterien angewendet, wie für die primäre Effektivitätsanalyse (*IRF-determined PFS*). Eine *Time to event* Analyse mit der Kaplan Meier Methode wurde zur Testung der Hypothese, dass die Zugabe von Pertuzumab zum Behandlungsregime keinen bedeutenden Einfluss auf die HRQoL hat, durchgeführt. Zur Zensierung sind im CSR keine Angaben identifizierbar.

Tabelle 4-28: Bewertung des Verzerrungspotenzials für <Endpunkt Zeit bis zur Verschlechterung des BCS-Score> in RCT mit dem zu bewertenden Arzneimittel (Erster Datenschnitt Mai 2011)

Studie	**Verzerrungspotenzial auf Studienebene**	**Verblindung Endpunkterheber**	**Adäquate Umsetzung des ITT-Prinzips**	**Hinweise auf ergebnisgesteuerte Berichterstattung**	**Sonstige das Verzerrungspotenzial beeinflussende Punkte**	**Verzerrungspotenzial Endpunkt**
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Auf Studienebene wurde das Verzerrungspotential als „niedrig“ eingestuft. Auf Endpunktebene wurde dies wie folgt beurteilt: Die Patientinnen, die den Fragebogen ausfüllten, waren verblindet, die Auswertung erfolgte ebenfalls verblindet. Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Alle randomisierten Patientinnen gingen in die Auswertung ein, das ITT-Prinzip wurde adäquat umgesetzt. In der Publikation der Studie wurde die Erfassung der Lebensqualität nicht erwähnt, im Studienbericht fanden sich keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. [49] Auch sonstige, das Verzerrungspotential beeinflussende Punkte, wurden nicht beobachtet. Da der entsprechende Fragebogen nicht auf die Nutzung durch Männer ausgerichtet ist, wurde er von den an der Studie teilnehmenden Frauen ausgefüllt, wobei diese Patientengruppe mit 806 von 808 Studienteilnehmern 99,8 % der randomisierten Patienten entspricht. Das Verzerrungspotential für den Endpunkt <Zeit bis zur Verschlechterung des BCS-Score> wurde deshalb mit „niedrig“ beurteilt.

Tabelle 4-29: Ergebnisse für <Endpunkt Zeit bis zur Verschlechterung des BCS-Score> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (1. Datenschnitt Mai 2011)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T +D Hazard Ratio [95% KI] p-Wert**	**Ptz+T+D vs Pla+T +D Zeit bis zur Verschlechterung in Wochen Median p-Wert**
	Patientinnen mit Ereignis n/N (%)	**Zeit bis zum Ereignis in Wochen Median [95% KI]**	**Patientinnen mit Ereignis n/N (%)**	**Zeit bis zum Ereignis in Wochen Median [95% KI]**		
Zeit bis zur Verschlechterung des BCS-Score	239 / 404 (59,2 %)	18,3 [18; 27]	222 / 402 (55,2 %)	26,7 [18; 36]	0,77 [0,64; 0,93] p = 0,0058	0,0061

Die Ergebnisse sind nach vorherigem Behandlungsstatus und Region stratifiziert. ITT Population. Log Rank Test, stratifiziert nach vorheriger Behandlung und Region
Es wurden nur weibliche Patientinnen befragt (und ausgewertet).
Ereignis bezieht sich hier auf das Auftreten einer Verschlechterung.
Datenquelle: Tabelle aus CSR 10 / 2011 S.1650 und zusätzliche Analysen [55]

Eine Verschlechterung des BCS-Score entsprechend FACT-B (1. Datenschnitt Mai 2011) wurde als eine Abnahme des BCS-Score von 2 Punkten nach Studienbeginn definiert. [57] Das Hinzufügen von Pertuzumab zu Trastuzumab+Docetaxel verlängerte signifikant die Zeit bis zur Verschlechterung HR = 0,77; 95 %KI [0,64; 0,93] p = 0,0058. 239 (59,2 %) der Patientinnen im Pla+T+D Behandlungsarm und 222 (55,2 %) der Patientinnen im Ptz+T+D Behandlungsarm, erlitten eine Verschlechterung basierend auf dem BCS-Score während der Studie.

4.3.1.3.1.4 <Sicherheit> – RCT

4.3.1.3.1.4.1 <Endpunkt Patienten mit unerwünschtem Ereignis (UE)>– RCT

Tabelle 4-30: Operationalisierung von <Endpunkt Patienten mit unerwünschtem Ereignis>

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	Ein unerwünschtes Ereignis (UE) war als jedes unangenehme und nicht gewünschte Zeichen (inklusive klinischer oder Laborbefunde), Symptom oder eine Krankheit mit temporärem Behandlungsbedarf mit medizinischen Produkten, unabhängig von einer möglichen Verursachung durch die Studienmedikation, definiert. Vorbestehende Konditionen, die sich während der Studie verschlechterten, wurden ebenfalls als UE gewertet. Eine Progression der zu Grunde liegenden Malignität wurde hier nicht gewertet, wenn sie eindeutig der erwarteten Krebsprogression zugeordnet werden konnte. Unerwünschte Ereignisse (Original Begriffe auf den CRFs - Datenerfassungsbogen mittels „preferred terms“ (PT)) wurden gemäß der aktuellsten Version des *Medical Dictionary for Regulatory Activities* (MedDRA, Version 14) kodiert. Die Daten zu den unerwünschten Ereignissen wurden nach Häufigkeit und nach NCI-CTCAE Grad (Version 3.0) dargestellt. Patienten, bei denen dasselbe Ereignis mehr als einmal auftrat, wurden bei der Berechnung der Häufigkeiten nur einmal mit dem höchsten NCI-CTCAE Grad gezählt. Die Auswertung erfolgte für alle Patienten, die mindestens eine Dosis der Studienmedikation erhalten hatten. Daten wurden während der gesamten Behandlungszeit inklusive der letzten Visite erfasst, die vier bis sechs Wochen (28-42 Tage) nach der Gabe der letzten Studienmedikation stattfand.

Tabelle 4-31: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit unerwünschtem Ereignis> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunktherheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Das Verzerrungspotential auf Studienebene wurde als niedrig eingestuft. Für den Endpunkt <Patienten mit unerwünschtem Ereignis> wurde ein mögliches Verzerrungspotential wie folgt beurteilt: Die Befragung der verblindeten Patienten sowie die Beurteilung von Krankheitszeichen und Symptomen erfolgten durch den ebenfalls verblindeten Investigator. Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Die Auswertung der sicherheitsbezogenen Ereignisse erfolgte für die *Safety Analysis Population* (SAP), sie bezog sich nur auf Patienten, die *as treated* ausgewertet

wurden, welche mindestens eine Dosis von einem der Komponenten der Studienbehandlung erhalten hatten. Die Vorgehensweise für die Sicherheitsauswertung wird als adäquat beurteilt. Es zeigten sich im Studienbericht keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. In der Publikation zur Studie werden die unerwünschten Ereignisse nicht als Gesamtsumme berichtet, sondern mit Fokus auf spezifische Ereignisse und den Schweregrad. [49] Andere, das Verzerrungspotential beeinflussende Punkte, wurden nicht beobachtet. Das Verzerrungspotential für den Endpunkt <Patienten mit unerwünschtem Ereignis> wurde insgesamt mit „niedrig“ beurteilt.

Tabelle 4-32: Ergebnisse für <Endpunkt Patienten mit unerwünschtem Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T +D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**
	Patienten mit Ereignis (Anzahl, %)	**Gesamt N**	**Patienten mit Ereignis (Anzahl, %)**	**Gesamt N**		
Patienten mit unerwünschtem Ereignis (UE)	391 (98,74 %)	396	408 (100 %)	408	1,01 [1,00; 1,02]	0,0286
p-Wert wurde mittels 2-seitigen Chi^2 Test bestimmt. Datenquelle: Tabelle S. 342 (und Tabelle 19) aus CSR update 12/2012; *SAP*, und zusätzliche Analysen [55]						

Die Inzidenz der Kategorie alle UE lag unter Pla+T+D bei 98,74 % vs. 100 % der Patienten unter Ptz+T+D (RR: 1,01 [95% KI: 1,00; 1,02]).

Das Ergebnis der Kategorie alle UE bedeutet, dass ein 1%-Unterschied zwischen den Gruppen besteht (knapp zwei Zehntel Prozent bei Betrachtung der unteren 95%-Konfidenzintervallgrenze), der lediglich aufgrund der hohen Fallzahl signifikant wurde. Vor dem Hintergrund des massiven multiplen Testproblems im Falle der vielen post-hoc Analysen ist das Ergebnis für alle AEs als borderline zu betrachten. Der Punktschätzer für das Relative Risiko liegt bei 1,012, die exakte untere Grenze des zweiseitigen 95%-Konfidenzintervalls liegt bei 1,0016, bereits 1 Patient mehr oder weniger in einem der Studienarme hätte dieses Ergebnis nicht signifikant werden lassen. Aus diesem Grund betrachtet Roche das Ergebnis als formal signifikant, aber nicht als medizinisch relevant.

Da die absolute zeitliche Exposition der Patientinnen unter Ptz+T+D mit Pertuzumab länger war als im Kontrollarm, kann eine höhere Anzahl von UE unter Ptz+T+D erwartet werden. Um die Gesamtzahl der UE der Therapiearme vergleichen zu können, wurde die Anzahl von UE bezogen auf die Therapiedauer berechnet. Insgesamt traten unter Pla+T+D 5.535 UE auf (18,72 Ereignisse pro Patient und Jahr) gegenüber mit 6.521 UE (16,88 Ereignisse pro Patient und Jahr) unter Ptz+T+D.

4.3.1.3.1.4.2 <Endpunkt Patienten mit schwerwiegendem unerwünschten Ereignis (SUE)>– RCT

Tabelle 4-33: Operationalisierung von <Endpunkt Patienten mit schwerwiegendem unerwünschten Ereignis

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	Schwerwiegende unerwünschte Ereignisse (SUE) wurden nach den ICH-Kriterien definiert und auf einem separaten SAE-Formular dokumentiert. Ein SUE ist jedes aufgetretene Ereignis, das eine signifikante Bedrohung, eine Kontraindikation, eine Nebenwirkung oder eine Vorsichtsmaßnahme sein könnte: Jedes unerwünschte Ereignis (UE, siehe Kapitel 4.3.1.3.1.4.1) das einen oder mehrere der folgenden Aspekte erfüllt, wurde als SUE klassifiziert (Studienbericht CSR 2. Datenschnitt, Seite 6731): · Das Ereignis ist tödlich (d.h. es resultiert ein Todesfall. Anmerkung: "Tod" ist ein Endpunkt, nicht ein Ereignis.) · Das Ereignis ist lebensbedrohlich (Anmerkung: Die Bezeichnung "lebensbedrohlich - *life threatening*" bezieht sich auf ein Ereignis, durch welches der Patient direkt tödlich bedroht ist zum Zeitpunkt des Ereignisses. Ereignisse, die hypothetisch, also z.B. bei größerem Schweregrad, eine Todesfolge haben könnten, sind nicht gemeint.) · Das Ereignis benötigt eine stationäre Aufnahme oder die Verlängerung einer bestehenden Hospitalisierung. · Das Ereignis resultiert in bleibender signifikanter Behinderung oder Einschränkung. · Es handelt sich um eine kongenitale Anomalie oder einen Geburtsdefekt. · Das Ereignis ist medizinisch signifikant oder benötigt eine Intervention, um eines oder mehrere der oben genannten Ereignisse zu vermeiden. Die Auswertung erfolgte für alle Patienten, die mindestens eine Dosis der Studienmedikation erhalten hatten. Daten wurden während der gesamten Behandlungszeit inklusive der letzten Visite erfasst, die vier bis sechs Wochen (28-42 Tage) nach der Gabe der letzten Studienmedikation stattfand.

Tabelle 4-34: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit schwerwiegendem unerwünschten Ereignis> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Auf Studienebene wurde das Verzerrungspotential als niedrig eingestuft. Auf Endpunktebene wurden mögliche Verzerrungsfaktoren für den Endpunkt <Patienten mit schwerwiegendem unerwünschten Ereignis> wie folgt beurteilt: Die Befragung der verblindeten Patienten sowie die Bewertung von Krankheitszeichen und Symptomen erfolgte durch den Investigator, der ebenfalls verblindet war. Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai

2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Ein SUE wurde anhand der Kriterien der ICH bewertet. Die Auswertung der sicherheitsbezogenen Ereignisse erfolgte für die *Safety Analysis Population* (SAP), sie bezog sich nur auf Patienten, welche mindestens eine Dosis von einer der Komponenten der Studienbehandlung erhalten hatten. Eine Auswertung der Safety-Endpunkte mittels des ITT-Prinzips ist nicht lege artis und würde auch der Intention der Safety-Analyse widersprechen. Die vollständige Auswertung erfolgte durch die Safety Population als „as treated“. Es zeigten sich weiterhin keine Hinweise auf eine ergebnis-gesteuerte Berichterstattung beim Abgleich mit der Publikation der Studie. [49] Auch sonstige, das Verzerrungspotential beeinflussende Punkte, wurden nicht gesehen. Das Verzerrungspotential für den Endpunkt <Patienten mit schwerwiegendem unerwünschtem Ereignis> wurde insgesamt mit „niedrig“ beurteilt.

Tabelle 4-35: Ergebnisse für <Endpunkt Patienten mit schwerwiegendem unerwünschten Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T +D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**
	Patienten mit Ereignis (Anzahl, %)	**Gesamt N**	**Patienten mit Ereignis (Anzahl, %)**	**Gesamt N**		
Patienten mit schwerwiegendem unerwünschtem Ereignis	115 (29,04 %)	396	148 (36,27 %)	408	1,25 [1,02; 1,53]	0,0289
Datenquelle: Tabelle S. 342 (und Tabelle 19) aus CSR update 12/2012; *SAP*, und zusätzliche Analysen [55] p-Wert wurde mittels Chi^2 Test bestimmt.						

Die Inzidenz der SUE lag bei Pla+T+D bei 29,0 % vs. 36,3 % der Patienten unter Ptz+T+D (RR 1,25; 95 % KI [1,02; 1,53]). Der Unterschied zwischen den beiden Behandlungsarmen in der Rate von SUE war hauptsächlich durch eine höhere Inzidenz febriler Neutropenien im Pertuzumab-Arm (11,3 % der Patientinnen) als im Kontrollarm (5,0 % der Patientinnen) bedingt, die eine Verlängerung des Krankenhausaufenthalts erforderlich machten.

Laut GCP-Verordnung und deutschem Arzneimittelgesetz bzw. Medizinproduktegesetz gelten alle unerwünschten Ereignisse als schwerwiegend, die:

- den Tod eines Probanden zur Folge haben (I)
- unmittelbar lebensbedrohend sind (II)
- einen unvorhergesehenen Krankenhausaufenthalt oder die Verlängerung eines Krankenhausaufenthalts erforderlich machen (III)
- eine kongenitale Anomalie oder einen Geburtsfehler nach sich ziehen (IV)
- eine bleibende oder schwerwiegende Behinderung oder Invalidität zur Folge haben (V)
- das Ereignis ist medizinisch signifikant oder benötigt eine Intervention, um eines oder mehrere der oben genannten Ereignisse zu vermeiden. (II, III und VI)

I.) Grad 5 UE waren in beiden Therapiearmen sehr selten. Im Kontrollarm traten 12 (3 % der Patienten) und im Pertuzumab-Arm 8 (2 %der Patienten) Grad 5 UE auf (RR: 0,65; 95% KI: [0,28; 1,57]. Auch wenn bedingt durch die geringe Fallzahl kein signifikanter Unterschied zwischen den Therapiearmen besteht, so gibt es einen deutlichen metrischen Trend zugunsten des Pertuzumab-Arms.

II.) Unmittelbar lebensbedrohliche SUE entsprechen dem Grad 4 NCI-CTCAE, Version 3.0. In beiden Behandlungsarmen traten, bei einem vergleichbaren Anteil von Patienten, UE von Grad 4 auf.

III.) Unter anderem führte die höhere Rate an febrilen Neutropenien im Pertuzumab-Arm zu mehr unvorhergesehenen oder verlängerten Krankenhausaufenthalten. Dies ist unangenehm für die Patienten, jedoch tritt dadurch keine bleibende oder dauerhafte Beeinträchtigung ein. Insgesamt lassen sich die febrilen Neutropenien in der Regel gut behandeln und sind von begrenzter Dauer.

IV.) Für Pertuzumab wurde bisher über keinen Fall berichtet.

V.) Ein Hinweis auf eine durch Pertuzumab-verursachte Veränderung des Sicherheitsprofils gegenüber Trastuzumab und Docetaxel liegt nicht vor.

VI.) Grad 3 NCI-CTCAE Ereignisse gelten als medizinisch signifikant. In beiden Behandlungsarmen traten, bei einem vergleichbaren Anteil von Patienten, UE von Grad ≥ 3 auf.

4.3.1.3.1.4.3 <Endpunkt Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis (UE)> – RCT

Tabelle 4-36: Operationalisierung von <Endpunkt Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis>

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	Alle Behandlungsabbrüche, die aufgrund eines UE erfolgten. Ein unerwünschtes Ereignis (UE) war als jedes unangenehme und nicht gewünschte Zeichen (inklusive klinischer oder Laborbefunde), Symptom oder eine Krankheit mit temporärem Behandlungsbedarf mit medizinischen Produkten, unabhängig von einer möglichen Verursachung durch die Studienmedikation, definiert. Vorbestehende Konditionen, die sich während der Studie verschlechterten, wurden ebenfalls als UE gewertet. Eine Progression der zu Grunde liegende Malignität wurde hier nicht gewertet, wenn sie eindeutig der erwarteten Krebsprogression zugeordnet werden konnte.

Tabelle 4-37: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Das Verzerrungspotential auf Studienebene wurde als niedrig eingestuft. Für den Endpunkt <Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis> wurde das Verzerrungspotential wie folgt beurteilt: Die Befragung der verblindeten Patienten sowie die Bewertung von Krankheitszeichen und Symptomen erfolgte durch die ebenfalls verblindeten Investoren. Die Verblindung in Studienperiode 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Die Auswertung der sicherheitsbezogenen Ereignisse erfolgte für die *Safety Analysis Population* (SAP), sie bezieht sich nur auf Patienten, welche mindestens eine Dosis von einer der Komponenten der Studienbehandlung erhalten hatten (*as treated*). Die Vorgehensweise für die Sicherheitsauswertung wird als adäquat bewertet. Es zeigten sich im Studienbericht keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. In der Publikation zur Studie werden die Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis nicht berichtet. [49] Sonstige, das Verzerrungspotential beeinflussende Punkte, wurden nicht beobachtet. Das Verzerrungspotential für den Endpunkt <Patienten mit unerwünschtem Ereignis> wurde insgesamt mit „niedrig“ beurteilt.

Tabelle 4-38: Ergebnisse für <Endpunkt Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T +D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**
	Patienten mit Ereignis (Anzahl, %)	**Gesamt N**	**Patienten mit Ereignis (Anzahl, %)**	**Gesamt N**		
Patienten mit Behandlungs-abbruch wegen UE [a]	114 (28,79 %)	396	125 (30,64 %)	408	1,06 [0,86; 1,32]	0,5665
Patienten mit Behandlungs-abbruch wegen UE sofern nicht Docetaxel bedingt [b]	24 (6,06 %)	396	32 (7,84 %)	408	1,29 [0,78; 2,16]	0,3212

a: Tabelle S. 342 (und Tabelle 19) aus CSR update 12/2012; *SAP*, und zusätzliche Analysen [55]
b: Tabelle S. 1424 aus CSR update 12/2012; *SAP*, und zusätzliche Analysen [55]
p-Wert wurde mittels Chi^2 Test bestimmt.

Die Inzidenz der UE, die zum Behandlungsabbruch führten lag unter Pla+T+D bei 28,8 % (n = 114) gegenüber 30,6 % (n = 125) unter Ptz+T+D (RR 1,06; 95 % KI [0,86; 1,32]).

Wenn Ereignisse, die zum Abbruch der Therapie mit Docetaxel führten, ausgeschlossen werden, traten nur bei 6,1 % der Patienten im Kontroll-Arm UE auf, die zum Absetzen der Studienmedikation führten, und bei 7,8 % der Patienten im Pertuzumab-Arm. Patienten die in dieser Studie Docetaxel absetzen, konnten die Therapie mit Pertuzumab und Trastuzumab fortsetzen. Dagegen mussten die Patienten bei Ereignissen, die zum Absetzen von Pertuzumab oder Trastuzumab führten, sämtliche Studienmedikamente absetzen.

4.3.1.3.1.4.4 <Endpunkt Patienten mit UE ≥ Grad 3> – RCT

Tabelle 4-39: Operationalisierung von <Endpunkt Patienten mit unerwünschtem Ereignis ≥ Grad 3>

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	Der Schweregrad aller UE wurde entsprechend der Vorgaben der NCI-CTCAE, Version 3.0, auf einer fünf-Punkte Skala (Grad 1 bis 5) eingeordnet. UE, die nicht im CTCAE gelistet waren, wurden als "mild, moderate, severe, life-threatening, oder fatal" eingeordnet. Ein unerwünschtes Ereignis (UE) war als jedes unangenehme und nicht gewünschte Zeichen (inklusive klinischer oder Laborbefunde), Symptom oder eine Krankheit mit temporärem Behandlungsbedarf mit medizinischen Produkten, unabhängig von einer möglichen Verursachung durch die Studienmedikation, definiert. Vorbestehende Konditionen, die sich während der Studie verschlechterten, wurden ebenfalls als UE gewertet. Eine Progression der zu Grunde liegende Malignität wurde hier nicht gewertet, wenn sie eindeutig der erwarteten Krebsprogression zugeordnet werden konnte. Die Auswertung erfolgte für alle Patienten, die mindestens eine Dosis der Studienmedikation erhalten hatten. Daten wurden während der gesamten Behandlungszeit inklusive der letzten Visite erfasst, die vier bis sechs Wochen (28-42 Tage) nach der Gabe der letzten Studienmedikation stattfand.

Tabelle 4-40: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit unerwünschtem Ereignis ≥ Grad 3> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunktherheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Auf Studienebene wurde das Verzerrungspotential als niedrig eingestuft. Auf Endpunktebene wurden mögliche Verzerrungsfaktoren für den Endpunkt<Patienten mit unerwünschtem Ereignis ≥ Grad 3> wie folgt beurteilt: Die Befragung der verblindeten Patienten sowie die Bewertung von Krankheitszeichen und Symptomen erfolgte verblindet durch Investigatoren, die Bewertung des Schweregrads basierte auf objektiven, prädefinierten Kriterien (u.a. NCI-CTCAE, Version 3.0). Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Die Auswertung der sicherheitsbezogenen Ereignisse erfolgte für die *Safety Analysis Population* (SAP), sie bezog sich nur auf Patienten, welche mindestens eine Dosis einer der Komponenten der Studienbehandlung erhalten hatten

(*as treated*). Die Vorgehensweise für die Sicherheitsauswertung wird als adäquat bewertet. Im Studienbericht zeigten sich keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. In der Publikation der Studie wurden die unerwünschten Ereignisse ≥ Grad 3 nicht als Gesamtsumme berichtet, sondern mit Fokus auf Ereignisse von speziellem Interesse, wie z.B. febrile Neutropenie oder Diarrhoe. [49] Auch sonstige, das Verzerrungspotential beeinflussende Punkte, wurden nicht beobachtet. Das Verzerrungspotential für den Endpunkt <Patienten mit unerwünschtem Ereignis ≥ Grad3 > wurde insgesamt mit „niedrig“ beurteilt.

Tabelle 4-41: Ergebnisse für<Endpunkt Patienten mit unerwünschtem Ereignis ≥ Grad 3> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T +D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**
	Patienten mit Ereignis (Anzahl, %)	**Gesamt N**	**Patienten mit Ereignis (Anzahl, %)**	**Gesamt N**		
Patienten mit UE ≥ Grad 3	291 (73,48 %)	396	311 (76,23 %)	408	1,04 [0,96; 1,12]	0,3707

Die Einteilung der UE erfolgte nach NCI-CTCAE (1-5). Bei vermehrtem Auftreten eines UE bei einer Patientin wurde das Ereignis mit dem höchsten Schweregrad gezählt.
Tabelle S. 342 + 349 (und Tabelle 19) aus CSR update 12/2012; *SAP*, und zusätzliche Analysen [55]
p-Wert wurde mittels Chi^2 Test bestimmt.

Obwohl die meisten UE von Schweregrad 1 oder 2 waren, trat bei der Mehrzahl der Patientinnen mindestens ein UE von Grad ≥ 3 auf (bei n = 291 (73,5 %) der Patienten unter Pla+T+D und bei n = 311 (76,2 %) der Patienten unter Ptz+T+D. Somit war die Inzidenz von UE von Grad ≥ 3 zwischen den Behandlungsarmen ausgewogen (RR: 1,04 [95% KI: 0,96; 1,12].

4.3.1.3.1.4.5 <Endpunkt Patienten mit UE Grad 3> – RCT

Tabelle 4-42: Operationalisierung von <Endpunkt Patienten mit UE Grad 3>

Studie	Operationalisierung
<CLEOPATRA (WO20698)>	Der Schweregrad aller UE wurde entsprechend der Vorgaben der NCI-CTCAE, Version 3.0, auf einer fünf-Punkte Skala (Grad 1 bis 5) eingeordnet. UE, die nicht im CTCAE gelistet waren, wurden als "mild, moderate, severe, life-threatening, oder fatal" eingeordnet (Grad 1-5). Ein unerwünschtes Ereignis (UE) war als jedes unangenehme und nicht gewünschte Zeichen (inklusive klinischer oder Laborbefunde), Symptom oder eine Krankheit mit temporärem Behandlungsbedarf mit medizinischen Produkten, unabhängig von einer möglichen Verursachung durch die Studienmedikation, definiert. Vorbestehende Konditionen, die sich während der Studie verschlechterten, wurden ebenfalls als UE gewertet. Eine Progression der zu Grunde liegende Malignität wurde hier nicht gewertet, wenn sie eindeutig der erwarteten Krebsprogression zugeordnet werden konnte. Die Auswertung erfolgte für alle Patienten, die mindestens eine Dosis der Studienmedikation erhalten hatten. Daten wurden während der gesamten Behandlungszeit inklusive der letzten Visite erfasst, die vier bis sechs Wochen (28-42 Tage) nach der Gabe der letzten Studienmedikation stattfand.

Tabelle 4-43: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit UE Grad 3> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Auf Studienebene wurde das Verzerrungspotential als niedrig eingestuft. Auf Endpunktebene wurden mögliche Verzerrungsfaktoren für den Endpunkt<Patienten mit UE Grad 3> wie folgt beurteilt: Die Befragung der verblindeten Patienten sowie die Bewertung von Krankheitszeichen und Symptomen erfolgte verblindet durch Investigatoren, die Bewertung des Schweregrads basierte auf objektiven, prädefinierten Kriterien (u.a. NCI-CTCAE, Version 3.0). Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Die Auswertung der sicherheitsbezogenen Ereignisse erfolgte für die *Safety Analysis Population* (SAP), sie bezog sich nur auf Patienten, welche mindestens eine Dosis einer der Komponenten der Studienbehandlung erhalten hatten (*as treated*). Die Vorgehensweise für die Sicherheitsauswertung wird als adäquat bewertet. Im Studienbericht zeigten sich keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. In

der Publikation der Studie wurden die UE Grad 3 nicht als Gesamtsumme berichtet, sondern die der UE ≥ Grad 3 mit Fokus auf Ereignisse von speziellem Interesse, wie z.B. febrile Neutropenie oder Diarrhoe. [49] Auch sonstige, das Verzerrungspotential beeinflussende Punkte, wurden nicht beobachtet. Das Verzerrungspotential für den Endpunkt <Patienten mit UE Grad 3> wurde insgesamt mit „niedrig“ beurteilt.

Tabelle 4-44: Ergebnisse für<Endpunkt Patienten mit UE Grad 3> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T +D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**
	Patienten mit Ereignis (Anzahl, %)	**Gesamt N**	**Patienten mit Ereignis (Anzahl, %)**	**Gesamt N**		
Patienten mit UE Grad 3	228 (57,58 %)	396	254 (62,25 %)	408	1,08 [0,97; 1,21]	0,1761

Die Einteilung der UE erfolgte nach NCI-CTCAE (1-5). Bei vermehrtem Auftreten eines UE bei einer Patientin wurde das Ereignis mit dem höchsten Schweregrad gezählt.
Tabelle S. 342 + 349 (und Tabelle 19) aus CSR update 12/2012; *SAP*, und zusätzliche Analysen [55]
p-Wert wurde mittels Chi^2 Test bestimmt.

UE Grad 3 traten in beiden Behandlungsarmen mit vergleichbarer Häufigkeit auf. Weder aus medizinischer, noch aus statistischer Sicht finden sich Hinweise auf ein erhöhtes Risikoprofil von Pertuzumab.

4.3.1.3.1.4.6 <Endpunkt Patienten mit UE Grad 4> – RCT

Tabelle 4-45: Operationalisierung von <Endpunkt Patienten mit UE Grad 4>

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	Der Schweregrad aller UE wurde entsprechend der Vorgaben der NCI-CTCAE, Version 3.0, auf einer fünf-Punkte Skala (Grad 1 bis 5) eingeordnet. UE, die nicht im CTCAE gelistet waren, wurden als "mild, moderate, severe, life-threatening, oder fatal" eingeordnet (Grad 1-5). Ein unerwünschtes Ereignis (UE) war als jedes unangenehme und nicht gewünschte Zeichen (inklusive klinischer oder Laborbefunde), Symptom oder eine Krankheit mit temporärem Behandlungsbedarf mit medizinischen Produkten, unabhängig von einer möglichen Verursachung durch die Studienmedikation, definiert. Vorbestehende Konditionen, die sich während der Studie verschlechterten, wurden ebenfalls als UE gewertet. Eine Progression der zu Grunde liegende Malignität wurde hier nicht gewertet, wenn sie eindeutig der erwarteten Krebsprogression zugeordnet werden konnte. Die Auswertung erfolgte für alle Patienten, die mindestens eine Dosis der Studienmedikation erhalten hatten. Daten wurden während der gesamten Behandlungszeit inklusive der letzten Visite erfasst, die vier bis sechs Wochen (28-42 Tage) nach der Gabe der letzten Studienmedikation stattfand.

Tabelle 4-46: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit UE Grad 4> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Auf Studienebene wurde das Verzerrungspotential als niedrig eingestuft. Auf Endpunktebene wurden mögliche Verzerrungsfaktoren für den Endpunkt <Patienten mit UE Grad 4> wie folgt beurteilt: Die Befragung der verblindeten Patienten sowie die Bewertung von Krankheitszeichen und Symptomen erfolgte verblindet durch Investigatoren, die Bewertung des Schweregrads basierte auf objektiven, prädefinierten Kriterien (u.a. NCI-CTCAE, Version 3.0). Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Die Auswertung der sicherheitsbezogenen Ereignisse erfolgte für die *Safety Analysis Population* (SAP), sie bezog sich nur auf Patienten, welche mindestens eine Dosis einer der Komponenten der Studienbehandlung erhalten hatten (*as treated*). Die Vorgehensweise für die Sicherheitsauswertung wird als adäquat bewertet. Im Studienbericht zeigten sich keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. In

der Publikation der Studie wurden die UE Grad 4 nicht als Gesamtsumme berichtet, sondern die der UE ≥ Grad 3 mit Fokus auf Ereignisse von speziellem Interesse, wie z.B. febrile Neutropenie oder Diarrhoe. [49] Auch sonstige, das Verzerrungspotential beeinflussende Punkte, wurden nicht beobachtet. Das Verzerrungspotential für den Endpunkt <Patienten mit UE Grad 4> wurde insgesamt mit „niedrig“ beurteilt.

Tabelle 4-47: Ergebnisse für<Endpunkt Patienten mit UE Grad 4> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T +D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**
	Patienten mit Ereignis (Anzahl, %)	**Gesamt N**	**Patienten mit Ereignis (Anzahl, %)**	**Gesamt N**		
Patienten mit UE Grad 4	159 (40,15 %)	396	165 (40,44 %)	408	1,01 [0,85; 1,19]	0,9333

Die Einteilung der UE erfolgte nach NCI-CTCAE (1-5). Bei vermehrtem Auftreten eines UE bei einer Patientin wurde das Ereignis mit dem höchsten Schweregrad gezählt.
Tabelle S. 342 + 349 (und Tabelle 19) aus CSR update 12/2012; *SAP*, und zusätzliche Analysen [55]
p-Wert wurde mittels Chi^2 Test bestimmt.

UE Grad 4 traten in beiden Behandlungsarmen mit vergleichbarer Häufigkeit auf. Weder aus medizinischer, noch aus statistischer Sicht finden sich Hinweise auf ein erhöhtes Risikoprofil von Pertuzumab.

4.3.1.3.1.4.7 <Endpunkt Patienten mit UE Grad 5> – RCT

Tabelle 4-48: Operationalisierung von <Endpunkt Patienten mit UE Grad 5>

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	Der Schweregrad aller UE wurde entsprechend der Vorgaben der NCI-CTCAE, Version 3.0, auf einer fünf-Punkte Skala (Grad 1 bis 5) eingeordnet. UE, die nicht im CTCAE gelistet waren, wurden als "mild, moderate, severe, life-threatening, oder fatal" eingeordnet (Grad 1-5). Ein unerwünschtes Ereignis (UE) war als jedes unangenehme und nicht gewünschte Zeichen (inklusive klinischer oder Laborbefunde), Symptom oder eine Krankheit mit temporärem Behandlungsbedarf mit medizinischen Produkten, unabhängig von einer möglichen Verursachung durch die Studienmedikation, definiert. Vorbestehende Konditionen, die sich während der Studie verschlechterten, wurden ebenfalls als UE gewertet. Eine Progression der zu Grunde liegende Malignität wurde hier nicht gewertet, wenn sie eindeutig der erwarteten Krebsprogression zugeordnet werden konnte. Die Auswertung erfolgte für alle Patienten, die mindestens eine Dosis der Studienmedikation erhalten hatten. Daten wurden während der gesamten Behandlungszeit inklusive der letzten Visite erfasst, die vier bis sechs Wochen (28-42 Tage) nach der Gabe der letzten Studienmedikation stattfand.

Tabelle 4-49: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit UE Grad 5> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012)

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Hinweise auf ergebnisgesteuerte Berichterstattung	Sonstige das Verzerrungspotenzial beeinflussende Punkte	Verzerrungspotenzial Endpunkt
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Auf Studienebene wurde das Verzerrungspotential als niedrig eingestuft. Auf Endpunktebene wurden mögliche Verzerrungsfaktoren für den Endpunkt<Patienten mit UE Grad 5> wie folgt beurteilt: Die Befragung der verblindeten Patienten sowie die Bewertung von Krankheitszeichen und Symptomen erfolgte verblindet durch Investigatoren, die Bewertung des Schweregrads basierte auf objektiven, prädefinierten Kriterien (u.a. NCI-CTCAE, Version 3.0). Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Die Auswertung der sicherheitsbezogenen Ereignisse erfolgte für die *Safety Analysis Population* (SAP), sie bezog sich nur auf Patienten, welche mindestens eine Dosis einer der Komponenten der Studienbehandlung erhalten hatten (*as treated*). Die Vorgehensweise für die Sicherheitsauswertung wird als adäquat bewertet. Im Studienbericht zeigten sich keine Hinweise auf eine ergebnisgesteuerte Berichterstattung. In

der Publikation der Studie wurden die UE Grad 5 nicht als Gesamtsumme berichtet, sondern die der UE ≥ Grad 3 mit Fokus auf Ereignisse von speziellem Interesse, wie z.B. febrile Neutropenie oder Diarrhoe [49] Auch sonstige, das Verzerrungspotential beeinflussende Punkte, wurden nicht beobachtet. Das Verzerrungspotential für den Endpunkt <Patienten mit UE Grad 5> wurde insgesamt mit „niedrig“ beurteilt.

Tabelle 4-50: Ergebnisse für<Endpunkt Patienten mit UE Grad 5> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T +D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**
	Patienten mit Ereignis (Anzahl, %)	**Gesamt N**	**Patienten mit Ereignis (Anzahl, %)**	**Gesamt N**		
Patienten mit UE Grad 5	12 (3,03 %)	396	8 (1,96 %)	408	0,65 [0,27; 1,57]	0,3306

Die Einteilung der UE erfolgte nach NCI-CTCAE (1-5). Bei vermehrtem Auftreten eines UE bei einer Patientin wurde das Ereignis mit dem höchsten Schweregrad gezählt.
Tabelle S. 342 + 349 (und Tabelle 19) aus CSR update 12/2012; *SAP*, und zusätzliche Analysen [55]
p-Wert wurde mittels Chi^2 Test bestimmt.

Tödliche UE waren sehr selten. Die Hinzugabe von Pertuzumab zu Trastuzumab und Docetaxel wurde gut vertragen und der Anteil der Patienten, die die Therapie „aufgrund Tod“ abbrachen, war in beiden Behandlungsarmen vergleichbar (Pla+T+D: n = 12; 3,0 % vs. Ptz+T+D: n = 8; 2,0%).

4.3.1.3.1.4.8 <Endpunkt Patienten mit UE von besonderem Interesse>– RCT

Tabelle 4-51: Operationalisierung von <Endpunkt Patienten mit UE von besonderem Interesse>

Studie	Operationalisierung
<CLEOPATRA (WO20698) >	**Unerwünschtes Ereignis von besonderem Interesse:** Diese Ereignisse wurden – wenn möglich – mittels standardisierter MedDRA-Abfragen (Queries) (SMQ) analysiert, da diese mit einem Set an AE konsistent sind, die besonders im Interesse von globalen regulatorischen Behörden stehen. Waren keine SMQ verfügbar, wurden Gruppen von MedDRA Adverse Event Grouped Terms (AEGTs) verwendet:

UE von besonderem Interesse	Vorgehen zur Sicherheitsanalyse
Diarrhoe	PT *diarrhea*
Exanthem (rash)	Roche Standard AEGT „*EGRF associated rash*“
Leukopenie, Neutropenie	SMQ (Eingrenzung) ‘*Hematopoietic leucopenia*’
Febrile Neutropenie	PT ‘*Febrile neutropenia*’ – Subgruppe der Suche nach ‘*leucopenia*’
Leukopenische Infektion	Ereignisse aus der Gruppe ‘Infections and Infestations’
Febrile neutropenische Infektion	SOC mit einem Beginn ≤ 14 Tage nach dem Startdatum eines ≥ Grad 3 Ereignisses aus den SMQ (Eingrenzung), ‘*Haematopoietic leucopenia*’ und für Infektionen, die nach dem PT ‘*febrile neutropenia*’ folgten – Subgruppe der Suche nach ‘Leukopenic infection’
Interstitielle Lungenerkrankung	SMQ (Eingrenzung) *interstitial lung disease*
Anaphylaxie und Hypersensibilitätsreaktionen	Roche Standard AEGT ‘*Anaphylaxis and hypersensitivity*’, enthalten in MedDRA SMQ (Eingrenzung) ‘*Anaphylactic reaction*’ plus alle MedDRA PTs, die ‘*hypersensitivity*’ enthalten.
Mukositis	Roche Standard AEGT ‘*Mucositis of gastrointestinal tract*’
Behandlungsbedingte Leberstörungen	SMQ (ausgedehnt) ‘Drug related hepatic disorders – umfassende Suche’. Diese SMQ besteht aus 8 individuellen SMQs*.
Kardiale Dysfunktion / SAE, die auf eine CHF hindeuten	SAE aus der SMQ (ausgedehnt) ‘*Cardiac failure*’
QT Verlängerung	SMQ (ausgedehnt) ‘*Torsade de pointes* / QT Verlängerung

SMQ = *Standard MedDRA queries*; AEGT = *Adverse events group terms* ; PT = *preferred term*; EGFR = *Epidermal growth factor receptor;* SOC = *System Organ Class*; CHF = *congestive heart failure* (chronische Herzinsuffizienz)

* Hepatotoxizität (hepatotoxicity) wurde durch die folgenden acht SMQs definiert: SMQ ‘Cholestase und Gelbsucht mit hepatischer Ursache‘ (‘*Cholestasis and jaundice of hepatic origin*’); SMQ ‘hepatische Störungen mit Relation zur Medikation – nur schwere Ereignisse‘ (‘*Drug related hepatic disorders - severe events only*’); SMQ ‘Leberversagen, Fibrose und Zirrhose und andere Konditionen, die in Verbindung mit einer Leberschädigung stehen‘ (‘*Hepatic failure, fibrosis and cirrhosis and other liver damage-related conditions*’); SMQ ‘Hepatitis, nicht infektiöse‘ (‘*Hepatitis, non infectious*’); SMQ ‘benigne Leberneoplasmen (inkl. Zysten und Polypen)‘ (‘*Liver neoplasms benign (incl cysts and polyps))*; SMQ ‘maligne und unspezifizierte Leberneoplasmen‘ (‘*Liver neoplasms, malignant and unspecified*’); SMQ Untersuchungen, Zeichen und Symptome, die mit der Leber im Zusammenhang stehen (‘*Liver related investigations, signs and symptoms*’) und SMQ ‘Störungen der Koagulation und Blutungen, die in Beziehung zur Leber stehen‘ (‘*Liver related coagulation and bleeding disturbances*’)

Datenquelle: CSR 10 / 2011 Tabelle 7, Seite 88

LVEF (Linksventrikuläre Ejektionsfraktion)
Der LVEF-Wert zu Studienbeginn (gemessen mittels ECHO oder MUGA) und das Maximum der absoluten Verminderung (oder das Minimum der absoluten Verbesserung, im Fall einer Erhöhung der Werte nach Studienbeginn) für LVEF-Messwerte nach der Baseline wurden summiert. Die Differenz für das Maximum der absoluten Verminderung für LVEF-Messwerte zwischen den beiden Behandlungsarmen wurde mittels des Wilcoxon Rank Summen-Tests beurteilt.

Die Investigatoren wurden instruiert, einen symptomatischen Abfall der linksventrikulären systolischen Funktion als solchen zu berichten. Patienten mit symptomatischem LVSD konnten nicht weiter an der Studie teilnehmen. Jeder Fall von LVSD wurde ebenfalls als SUE berichtet und entsprechend der Vorgaben von NCI-CTCAE (Version 3) und der NYHA Klassifikation in einen Schweregrad eingeteilt.

NCI-CTCAE Grad 3 symptomatisches LVSD wurde bis zu drei Jahre nach Beendigung der Studienbehandlung berichtet und dann bis zur Wiederherstellung / einer Stabilisierung / dem Tod bzw. einer Bestätigung, dass keine Verbesserung mehr zu erwarten ist oder dem Ende der Überlebens-Nachbeobachtungszeit nachverfolgt.

Ein asymptomatischer Abfall der LVEF wurde nur dann als UE berichtet, wenn der Abfall mindestens 10 % seit Studienbeginn betrug und unter 50 % oder im Falle der Notwendigkeit einer Behandlung oder des Abbruchs der Behandlung mit der Studienmedikation. Auch diese Ereignisse wurden nach NCI-CTCAE (Version 3.0) bzgl. ihres Schweregrads beurteilt.

Potentielle kardiale Ereignisse, deren Auflistung im eCRF zusätzlich zu den Patienten-Quell-Dokumenten und Datenprofilen durch Quintiles Verfügung gestellt wurde, wurden auch dem CRC zur Verfügung gestellt. Das CRC reviewte diese Listen entsprechend der Definitionen für kardiale Ereignisse, die in die folgenden Kategorien unterteilt wurden:

- Symptomatische LVSD (nicht tödlich)
- Symptomatische LVSD (tödlich)
- Anderes kardiales, nicht tödliches Ereignis
- Nicht-LVSD kardialer Tod
- Nicht-kardialer kardiovaskulärer Tod
- Nicht-kardiovaskulärer Tod
- Wahrscheinlicher (probable) kardialer Tod
- Nicht eruierbar

Tabelle 4-52: Bewertung des Verzerrungspotenzials für <Endpunkt Patienten mit UE von besonderem Interesse> in RCT mit dem zu bewertenden Arzneimittel (2. Datenschnitt Mai 2012)

Studie	**Verzerrungspotenzial auf Studienebene**	**Verblindung Endpunkterheber**	**Adäquate Umsetzung des ITT-Prinzips**	**Hinweise auf ergebnisgesteuerte Berichterstattung**	**Sonstige das Verzerrungspotenzial beeinflussende Punkte**	**Verzerrungspotenzial Endpunkt**
<CLEOPATRA (WO20698) >	niedrig	ja	ja	nein	nein	niedrig

Auf Studienebene wurde das Verzerrungspotential als niedrig eingestuft. Auf Endpunktebene wurden mögliche Verzerrungsfaktoren für den Endpunkt <Endpunkt Patienten mit UE von besonderem Interesse> wie folgt beurteilt: Die Befragung der verblindeten Patienten sowie die Bewertung von Krankheitszeichen und Symptomen erfolgte durch den Investigator, der ebenfalls verblindet war. Die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden. Ein SUE wurde anhand der Kriterien der ICH bewertet. Die Auswertung der sicherheitsbezogenen Ereignisse erfolgte für die *Safety Analysis Population* (SAP), sie bezog sich nur auf Patienten, welche mindestens eine Dosis von einer der Komponenten der Studienbehandlung erhalten hatten. Eine Auswertung der Safety-Endpunkte mittels des ITT-Prinzips ist nicht lege artis und würde auch der Intention der Safety-Analyse widersprechen. Die vollständige Auswertung erfolgte durch die Safety Population als „as treated“. Es zeigten sich weiterhin keine Hinweise auf eine ergebnisgesteuerte Berichterstattung beim Abgleich mit der Publikation der Studie. [49] Auch sonstige, das Verzerrungspotential beeinflussende Punkte, wurden nicht gesehen. Das Verzerrungspotential für den Endpunkt <Patienten mit UE von besonderem Interesse> wurde insgesamt mit „niedrig“ beurteilt.

Tabelle 4-53: Zusammenfassung für Patienten mit UE von besonderem Interesse während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

	Placebo+Trastuzumab +Docetaxel		**Pertuzumab+Trastuzumab +Docetaxel**		**Ptz+T+D vs Pla+T +D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**
	Patienten mit Ereignis (Anzahl, %)	**Gesamt N**	**Patienten mit Ereignis (Anzahl, %)**	**Gesamt N**		
Diarrhö	191 (48,23 %)	396	278 (68,14 %)	408	1,41 [1,25; 1,60]	< 0,0001
Exanthem	144 (36,36 %)	396	194 (47,55 %)	408	1,31 [1,11; 1,54]	0,0013
Leukopenie	231 (58,33 %)	396	255 (62,5 %)	408	1,07 [0,96;1,20]	0,2271
Febrile Neutropenie	30 (7,58 %)	396	56 (13,73 %)	408	1,81 [1,19; 2,76]	0,0048
Febrile neutropenische Infektion	3 (0,76 %)	396	14 (3,43 %)	408	4,53 [1,31; 15,64]	0,0120[a]
Interstitielle Lungenerkrankung	6 (1,52 %)	396	10 (2,45 %)	408	1,62 [0,59; 4,41]	0,3425
Anaphylaxie und Hypersensibilitäts-reaktionen	36 (9,09 %)	396	45 (11,03 %)	408	1,21 [0,80; 1,84]	0,3616
Behandlungsbedingte Leberstörungen	43 (10,86 %)	396	42 (10,29 %)	408	0,95 [0,63; 1,42]	0,7948
Mukositis	150 (37,88 %)	396	203 (49,75 %)	408	1,31 [1,12; 1,54]	0,0007
Kardiale Dysfunktion / SUEs, die auf eine CHF hindeuten	8 (2,02 %)	396	6 (1,47 %)	408	0,73 [0,25; 2,08]	0,5517
QT Verlängerung	5 (1,26 %)	396	9 (2,21 %)	408	1,75 [0,59; 5,17]	0,3070

Quelle: Tabelle S. 342-348 aus CSR update 12/2012; *SAP*, und zusätzliche Analysen [55]
p-Wert (CMH Chi^2 Test) da niedrige Ereignisrate
CHF = congestive heart failure (chronische Herzinsuffizienz)

Es wird in folgendem Text nur auf signifikante Ergebnisse weiter eingegangen.

Diarrhö

Die Inzidenz der Diarrhö lag unter Pla+T+D bei 48,2 % vs. 68,1 % unter Ptz+T+D (RR 1,41; 95 % KI [1,25; 1,60]). Die meisten Diarrhöen waren vom Schweregrad 1–2 (88 %). Grad ≥ 3 Diarrhöen traten bei 5,1 % (n = 20) der Patientinnen im Kontrollarm und bei 9,1 % (n = 37) der Patienten im Pertuzumab-Arm auf. In keinem der beiden Therapiearmen wurden fatale Verläufe beobachtet. In beiden Behandlungsarmen war der Anteil der Patienten mit einer Diarrhö als UE im ersten Therapiezyklus am höchsten und nahm danach stetig ab. Das Auftreten einer Diarrhö nahm am deutlichsten zwischen Zyklus 1 und Zyklus 2 ab. Nach Zyklus 10 trat bei weniger als 4 % der Patienten im Kontroll-Arm und bei weniger als 5 % der Patienten im Pertuzumab-Arm eine Diarrhö auf. Sie ist, wie präklinische Studien gezeigt haben, eine häufige Nebenwirkung von Substanzen, die auf den HER-Rezeptor abzielen. [60-62] In den meisten Fällen war die Diarrhö auch in den präklinischen Studien reversibel und behandelbar. Zudem wurden bei der mikroskopischen Untersuchung keine Anomalien in der Darmschleimhaut festgestellt und es fand sich kein Hinweis auf eine Gewebeschädigung. Im Zusammenhang mit der Brustkrebstherapie auftretende Diarrhoen lassen sich ohne Probleme durch Standardtherapien (Loperamid) behandeln. [63] Sie stellen klinisch eine sehr gut beherrschbare Nebenwirkung der Therapie mit Ptz+T+D (RMP 2.1) oder Pla+T+D (Trastuzumab + Taxan, hier Docetaxel) dar.

Eine undifferenzierte statistische Betrachtung der UE-Kategorie Diarrhö führt aus medizinischer Sicht zu einem verzerrten Bild. Die medizinische Gesamtbetrachtung der UE des Kontrollarms und des Pertuzumab-Arms zeigt, dass im Pertuzumab-Arm zwar statistisch mehr UE beobachtet wurden, diese aber in der Regel aber sehr gut mit einer Standardtherapie behandelbar waren. Weder im Kontrollarm noch im Pertuzumab-Therapiearm wurden fatale Verläufe beobachtet. Unter anderem führte die höhere Inzidenz im Pertuzumab-Arm zu mehr unvorhergesehenen oder verlängerten Krankenhausaufenthalten. Dies ist unangenehm für die Patienten, es tritt jedoch keine bleibende oder dauerhafte Beeinträchtigung ein.

Exanthem

Die Inzidenz der Exantheme lag unter Pla+T+D bei 36,4 % vs. 47,5 % unter Ptz+T+D (RR 1,31; 95 % KI [1,11; 1,54]). Dabei war der Anteil an Exanthemen mit Schweregrad ≥ 3 sehr gering (Kontrollarm n = 5; 1,3 % vs Pertuzumab-Therapie-Arm n=12; 2,9 %).

Exanthem ist ein Klasseneffekt von EGFR-Hemmern. EGFR wird in der Haut vor allem in undifferenzierten proliferierenden Keratinozyten der Basalschicht der Epidermis und der äußeren Schicht der Haarfollikel exprimiert und ist entscheidend für die normale Entwicklung und Physiologie der Epidermis. Keratinozyten durchlaufen einen hoch regulierten Prozess der Differenzierung und Migration zur Hautoberfläche. [64] Exantheme traten meistens in den ersten beiden Therapiezyklen auf, in späteren Therapiezyklen waren sie selten. Ein Grad 5 Exanthem oder Rash trat in keinem Therapiearm auf.

Das Exanthem wird derzeit nicht als bedeutsames identifiziertes Risiko klassifiziert und gilt nicht als medizinisch signifikant (RMP 2.1). Exantheme lassen sich mit Hydrocortison, Bethamethason, Clindamycin oder Hydroxyzin sehr gut behandeln.

Eine undifferenzierte statistische Betrachtung der UE-Kategorie Exanthem führt aus medizinischer Sicht zu einem verzerrten Bild. Die medizinische Gesamtbetrachtung der UE-Kategorie Exanthem des Kontrollarms und des Pertuzumab-Arms zeigt, dass im Pertuzumab-Arm zwar statistisch mehr Exantheme beobachtet wurden, diese aber in der Regel aber sehr gut mit einer Standardtherapie behandelbar waren. Weder im Kontroll- noch im Pertuzumab-Arm starben Patienten an einem Exanthem oder Rash. Grad ≥ 3 Exantheme waren in beiden Therapiearmen selten.

Febrile Neutropenie, febrile neutropenische Infektion

Hinsichtlich febriler Neutropenie lag die Inzidenz bei 7,6 % (n = 30) unter Pla+T+D vs. 13,8 % (n = 56) unter Ptz+T+D: RR 1,81; 95 % KI [1,19; 2,76]. Febrile neutropenische Infektionen traten n = 3 unter Pla+T+D und n = 14 unter Ptz+T+D auf (RR = 4,53; 95 % KI [1,31; 15,64]).

In der Fachinformation zu Docetaxel werden febrile Neutropenie und Neutropenie als sehr häufige Ereignisse genannt. Laut der Fachinformation tritt eine Neutropenie Grad 4 bei sogar 54,2 % der mit 75 mg / m^2 Docetaxel allein behandelten Patienten und bei 76 % der mit 100 mg / m^2 Docetaxel allein behandelten Patienten auf. [65]

Bei asiatischen Patienten wurde im Vergleich zu Patienten anderer ethnischer Herkunft und aus anderen geographischen Regionen in beiden Behandlungsgruppen eine höhere Inzidenz febriler Neutropenien beobachtet. Bei den asiatischen Patienten war auch die Inzidenz febriler Neutropenien ≥ Grad 3 in der mit Pertuzumab-behandelten Gruppe höher (n = 33; 25,78 %) als bei „kaukasischen“ Patienten (n = 20; 8,0 %). [46;47] Asiatische Patienten sind wahrscheinlich empfindlicher für die durch Docetaxel bedingte Hämatotoxizität, da sie wegen des niedrigeren Körpergewichts, des niedrigeren Body Mass Index (BMI) und der geringeren Körperoberfläche (BSA) einer höheren Exposition gegenüber Docetaxel ausgesetzt sind.

Die Exazerbation einer Chemotherapie- / Docetaxel-assoziierten Neutropenie ist ein identifiziertes Risiko bei der Anwendung von Pertuzumab, welches durch eine konsequente Therapie jedoch klinisch sehr gut beherrscht wird (RMP 2.1). Auch die Fachinformation für Pertuzumab führt die febrile Neutropenie als Therapierisiko für die Kombination Chemotherapie, z.B. Docetaxel, Trastuzumab und Pertuzumab auf: „Patienten, die mit Perjeta, Trastuzumab und Docetaxel behandelt werden, haben im Vergleich zu Patienten, die mit Pla+T+D, Trastuzumab und Docetaxel behandelt werden, vor allem während der ersten 3 Zyklen der Behandlung ein erhöhtes Risiko für eine febrile Neutropenie (siehe Abschnitt 4.8).“ Da der Nadir der Neutrophilenzahl bei Patienten, die mit Perjeta behandelt wurden, und bei Patienten, die mit Pla+T+D behandelt wurden, ähnlich war, könnte die höhere Inzidenz febriler Neutropenien bei Patienten, die mit Perjeta behandelt wurden, bei diesen Patienten mit einer höheren Inzidenz von Mukositis und Diarrhö assoziiert sein. Eine symptomatische

Behandlung der Mukositis und der Diarrhö sollte in Erwägung gezogen werden. In der Studie CLEOPATRA wurden keine Ereignisse von febriler Neutropenie berichtet, nachdem Docetaxel abgesetzt wurde.

Welcher Mechanismus dem Phänomen zugrunde liegt, dass Pertuzumab die Inzidenz leukopenischer / neutropenischer Ereignisse bei mit Docetaxel plus Trastuzumab behandelten Patienten erhöht, ist nicht bekannt. Jedoch scheint dies im Gegensatz zu dem Anstieg bei Zugabe von Trastuzumab zu Docetaxel weder durch einen Anstieg der Inzidenz von Neutropenien Grad 3 / 4 bedingt zu sein (basierend auf Labordaten) noch führt Pertuzumab zu einer Verlängerung einer mit Docetaxel assoziierten Neutropenie. Möglicherweise ist die Zunahme leukopenischer Ereignisse durch die höhere Rate von Mukositis und / oder Diarrhö bei Zugabe von Pertuzumab zu Docetaxel plus Trastuzumab bedingt, da Läsionen von Schleimhautoberflächen das Eindringen von Mikroorganismen in den Kreislauf ermöglichen können, was bei Patienten mit Neutropenie eine höhere Inzidenz von Fieber / Infektionen zur Folge hat. [66]

In beiden Behandlungsgruppen war der Anteil von Patienten mit einer febrilen Neutropenie im ersten Therapiezyklus am höchsten und nahm danach stetig ab. Am deutlichsten nahm die Häufigkeit zwischen Zyklus 1 und Zyklus 2 ab. Nach Zyklus 10 trat in beiden Behandlungsarmen bei nicht mehr als einem Patienten in irgendeinem Behandlungszyklus eine febrile Neutropenie auf. UE im Zusammenhang mit einer febrilen Neutropenie wurden in den ersten 5 Therapiezyklen im Pertuzumab-Arm häufiger angegeben, obgleich die Unterschiede in der Häufigkeit gering waren. Danach war die Häufigkeit febriler Neutropenien in beiden Behandlungsarmen vergleichbar niedrig.

Mukositis

Die Häufigkeit einer Mukositis lag unter Pla+T+D bei 37,9 % vs. 49,8 % unter Ptz+T+D: RR 1,31; 95 % KI [1,12, 1,54]. Dabei war der Anteil an Mukositis mit Schweregrad ≥ 3 gering (Kontrollarm 2,0 %, n = 8 vs. Pertuzumab-Therapiearm 3,2 %, n = 13). Eine Mukositis trat häufig beim Nadir der Neutrophilenzahl auf. Dies passt zu der Beobachtung in der klinischen Praxis, dass bei Patienten, die sich einer antineoplastischen Therapie unterziehen, Mukositis und Neutropenie häufig zusammen auftreten. Wahrscheinlich infolge ähnlicher zytotoxischer Wirkungen der Chemotherapie auf das Knochenmark und mukosale Vorläuferzellen kommen diese Nebenwirkungen zusammen vor. [67]

Das Auftreten einer Mukositis unter einer Brustkrebstherapie mit Trastuzumab, Docetaxel und Pertuzumab beeinflusste die Therapie nur unwesentlich und ließ sich durch Standardmaßnahmen im Rahmen einer Chemotherapie sehr gut beherrschen. [63]

Eine undifferenzierte statistische Betrachtung der UE-Kategorie Mukositis führt aus medizinischer Sicht zu einem verzerrten Bild. Die medizinische Gesamtbetrachtung der UE-Kategorie Mukositis des Kontrollarms und des Pertuzumab-Arms zeigt, dass im Pertuzumab-Arm zwar statistisch mehr Schleimhautentzündungen beobachtet wurden, diese aber in der Regel sehr gut mit einer Standardtherapie behandelbar waren. Weder im Kontrollarm, noch

im Pertuzumab-Therapiearm starben Patienten an einer Mukositis. Grad 3 Mukositis waren in beiden Therapiearmen selten.

Tabelle 4-54: Ergebnisse für<Patienten mit UE von besonderem Interesse: linksventrikuläre Dysfunktion> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region

	Placebo+Trastuzumab+Docetaxel		**Pertuzumab+Trastuzumab+Docetaxel**		**Ptz+T+D vs Pla+T +D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**
	Patienten mit Ereignis (Anzahl, %)	**Gesamt N**	**Patienten mit Ereignis (Anzahl, %)**	**Gesamt N**		
Patienten mit UE Links-ventrikuläre Dysfunktion [a]	34 (8,59 %)	396	22 (5,39 %)	408	0,63 [0,37; 1,05]	0,0755
Patienten mit UE symptomatische LVSD, festgestellt durch CRC [b]	4 (1,01 %)	396	4 (0,98 %)	408	0,97 [0,24; 3,85]	1,0000*
Patienten mit UE symptomatische LVSD festgestellt durch *Investigator*[c]	7 (1,77 %)	396	5 (1,23 %)	408	0,69 [0,22; 2,17]	0,5264
Patienten mit UE asymptomatische LVSD [d]	27 (6,82 %)	396	17 (4,17 %)	408	0,61 [0,34; 1,10]	0,0986

a, b, c, d: Datenquelle: Tabelle S. 342 (und Tabelle 19) aus CSR update 12/2012; *SAP*, und zusätzliche Analysen [55]
(Grad ≥ 3 Ereignisse sind symptomatic, folglich lässt sich daraus dann die Anzahl der asymptomatischen herausrechnen).
*: exakter p-Wert gezeigt, da wenige Ereignisse.
p-Wert wurde mittels Chi^2 Test bestimmt.

Der Anteil der Patientinnen, bei denen kardiale UE auftraten, war zwischen den Behandlungsarmen vergleichbar (n = 69; 17,4% im Kontroll-Arm vs. n = 63, 15,4 % im Pertuzumab-Arm) (CSR 12/2012 S.100).

Das häufigste kardiale UE war eine Linksventrikuläre Dysfunktion (LVSD), die im Arm Pla+T+D häufiger (n = 34; 8,6 % der Patientinnen) angegeben wurde als im Pertuzumab-Arm (n = 22; 5,4 % der Patientinnen). Im Kontrollarm hatten auch mehr Patientinnen (n = 7; 1,8 %) nach Bewertung des Prüfarztes eine symptomatische LVSD als im Pertuzumab-Arm (n = 5 Patientinnen; 1,2 %). Das Auftreten der LVEF Grad ≥ 3 fielen ebenfalls zugunsten des Pertuzumab-Arms aus: bei 13 Patientinnen (3,3 %) im Kontrollarm und 5 Patientinnen (1,2 %) im Pertuzumab-Arm kam es zu einer LVEF Grad ≥ 3. Diese Ergebnisse wurden vom unabhängigen Cardiac Review Committee (CRC) bestätigt, das die kardialen Daten von über 600 Patienten in der Studie prüfte. Aus welchen Gründen Pertuzumab, das wie Trastuzumab am HER2-Rezeptor angreift, die Wirksamkeit, aber nicht die Kardiotoxizität von Trastuzumab erhöht, ist bisher nicht geklärt. Dieses Phänomen wurde jedoch bereits früher in klinischen Studien mit Trastuzumab dokumentiert. Der Tyrosinkinase-Hemmer Lapatinib, der an HER1 und HER2 angreift, wurde ebenfalls zusammen mit Trastuzumab verabreicht, ohne dass es zu einem größeren Anstieg des kardialen Risikos kam. [68]

Die Inzidenz der asymptomatischen linksventrikulären Dysfunktionen lag bei mit Pla+T+D behandelten Patienten bei 6,8 % und bei den mit Pertuzumab behandelten Patienten bei 4,2 %. Dies entspricht einem Relativen Risiko von 0,61 bei einem 95 % KI von [0,34; 1,10].

4.3.1.3.2 Subgruppenanalysen – RCT

Für die tabellarische Darstellung der Ergebnisse aus Subgruppenanalysen gelten die gleichen Anforderungen wie für die tabellarische Darstellung von Ergebnissen aus Gesamtpopulationen in Abschnitt 4.3.1.3.1.

Beschreiben Sie die Ergebnisse von Subgruppenanalysen (einschließlich der Interaktionsterme). Stellen Sie dabei die Ergebnisse in den Subgruppen zunächst für jede einzelne Studie in tabellarischer Form dar. Diese Anforderung gilt sowohl für Subgruppenanalysen auf Basis individueller Patientendaten als auch für solche auf Basis aggregierter Daten.

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse 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 bzw. wenn Sie nicht alle Studien in die Meta-Analyse einschließen.

Da bei der Suche nach Studien für diese Nutzenbewertung nur eine Studie identifiziert wurde, entfällt die Durchführung von Meta-Analysen.

In den Dossiers sollen Interaktionstests für alle Subgruppenanalysen (Treatment-by-subgroup) für alle klinischen Endpunkte dargestellt werden. Diese Forderung hat Implikationen und hängt auch mit der Planung der Studie zusammen.

Im Zuge des Designs der Studie gibt es einige wesentliche statistische Aspekte, die Rückwirkung auf die Interaktionstests haben.

A) Betrachtung studienbezogen bzw. über alle Endpunkte hinweg:

- **Post-hoc Definition reduziert die Generalisierbarkeit.**

  Post-hoc gebildete Subgruppen haben keinen Beweischarakter. Deren Ergebnisse basieren auf keiner methodisch korrekten Prüfung der entsprechenden Hypothesen. [1] U.a. aus folgenden (lediglich auf die post-hoc Definition) bezogenen Gründen:

  - Nichtbeachtung des Selektionsbias. Dessen Einfluss ist umso größer, je kleiner die Subgruppen sind.
  - Subgruppen werden bei nicht statistisch signifikanten Unterschieden von prognostischen Faktoren als vergleichbar angenommen.
  - Die Schätzgenauigkeit in den Subgruppen ist aufgrund der geringen Fallzahl meist sehr gering. Damit sind auch die Schätzungen nicht sehr belastbar und die Ergebnisse können in einem sehr breiten Rahmen variieren.

Den Goldstandard für Subgruppenanalysen stellen Auswertungen bezüglich a priori gebildeter Subgruppen dar, wobei eine anhand der Subgruppen stratifizierte Randomisierung

und eine für die Datenanalyse adäquate statistische Methode (Homogenitätstest, Interaktionstest) vorgenommen bzw. angewandt werden. Dieser Goldstandard ist in der CLEOPATRA-Studie ausschließlich für zwei Subgruppen erfüllt:

- Vorhergehender Behandlungsstatus (de novo vs vorangegangene adjuvante oder neoadjuvante Therapie)
- Region (Europa, Nord-Amerika, Süd Amerika und Asien)

· **Multiplizität / Typ I Fehlerinflation**

Die Anzahl der post-hoc durchgeführten Interaktionstests lässt sich anhand Tabelle 4-55 nachvollziehen.

Tabelle 4-55: Anzahl der post-hoc durchgeführten Interaktionstests

Kategorie	**Anz. Variablen**	**Prädefiniert**	**Post-hoc extern (G-BA)**	**Post-hoc**
Efficacy	9	9	2	3
Safety	7	2		

Damit ergeben sich 140 Einzeltests. Dies entspricht nicht dem ursprünglich geplanten globalen zweiseitigen 5%-Signifikanzniveau, sondern gemäß:

$\alpha^* = 1 - (1 - \alpha)^k$ mit $\alpha = 0{,}05$ (zweiseitig) und k = Anzahl der Vergleiche

einem Niveau von 99,9 %, das heißt, die Wahrscheinlichkeit von mindestens einer falsch positiven Entscheidung beträgt 99,9 %. Wird nicht adjustiert, ergeben sich genau 7 falsch-positive Testergebnisse.

Nach der post-hoc Durchführung der Interaktionstests weisen im vorliegenden Dossier folgende abgestuften p-Werte für den Signifikanztest auf.

Lediglich 4 Tests weisen ein signifikantes Ergebnis unterhalb des Signifikanzniveaus von 5 % auf. Zwei weitere Tests liegen unterhalb des 10 %igen Signifikanzniveaus. Diese 6 Tests liegen numerisch weit unterhalb dessen, was zufällig an signifikanten Interaktionstests zu erwarten gewesen wäre und sind damit sehr gut mit der Hypothese ‚Zufallsbefunde' vereinbar, selbst wenn man das ‚Underpowering' der Interaktionstests berücksichtigt. Dies gilt auch für die weiteren 7 Interaktionstests oberhalb des 10 % Signifikanzniveaus.

Tabelle 4-56: Interaktionstest: zu erwartende Zufallsbefunde bei verschiedenen, gegebenen Signifikanzniveaus

Interaktionstest p<0.05 (7 zu erwartende Zufallsbefunde bei gegebenem Signifikanzniveau)			
OS	viszeral, nicht viszeral	p = 0,0138	
PFS	viszeral, nicht viszeral	p = 0,0332	
	FISH Status	p = 0,0254	
SUE	Rasse	p = 0,0259	
0.05 ≤ p<0.1 (7 zusätzlich zu erwartende Zufallsbefunde bei gegebenem Signifikanzniveau)			
OS	Lokalrezidive	p = 0,073	
ORR	Vorausgehender Behandlungsstatus	p = 0,0919	
0.1 ≤ p<0.2 (14 zusätzlich zu erwartende Zufallsbefunde bei gegebenem Signifikanzniveau)			
OS	FISH Status	p = 0,1900	
PFS	IHC-HER2-Status (3+,2+)	p = 0,1252	
	ER / PgR-Status	p = 0,1277	
ORR	vorherige (neoajduvante oder adjuvante) Behandlung mit Trastuzumab	p = 0,1491	
	Dosiseskalation mit Docetaxel auf 100 mg / m^2	p = 0,1482	
FACT-B TOI	Subgruppen Altersgruppe (< 75; ≥ 75)	p = 0,1448	
	IHC-HER2-Status (3+,2+)	p = 0,1566	

Post-hoc definierte Subgruppen haben per se keinen Beweischarakter. Auch für alle anderen gilt, dass sie aufgrund der guten Verträglichkeit der Anzahl der positiven Interaktionstests mit der Zufallshypothese auf Studienebene, kaum sinnvoll in den inhaltlichen Kategorien des IQWiG (Beleg“ für eine Interaktion bei p-Wert von ≤ 0,05; „Hinweis“ bei p-Wert ≤ 0,2) interpretiert werden können, wenn nicht weitere Aspekte auf Ebene der Einzelvariablen gegeben sind. Dann und nur dann kann die Kategorisierung seitens des IQWiG mit immer noch gebührender Vorsicht überhaupt zum Einsatz kommen.

B) Bezogen auf Einzelvariablen sind drei weitere Aspekte zu beachten

- **Umfang der verfügbaren bzw. zusätzlichen Information (z.B. Events)**

  Für die Subgruppen der Patienten mit viszeralen Metastasen *Viszeral* versus Patienten mit nicht-viszeralen Metastasen *Nicht Viszeral* beruht die Analyse auf n = 90 im Arm Pla+T+D bzw. n = 88 im Arm Ptz+T+D. Für diese Subgruppen ändern sich sowohl das Hazard-Ratio (HR) wie auch das Konfidenzintervall zwischen den beiden Datenschnitten erheblich. Hierbei ist zu beachten, dass die Zunahme an zusätzlicher Information (in diesem Fall Events) zwischen den Datenschnitten aufgrund der sehr geringen Eventrate in der Subgruppe *Viszeral* sehr gering ist. Hier sind insgesamt 19 Ereignisse in beiden Armen zusammen eingetreten (7 unter Pla+T+D, 12 unter Ptz+T+D), so dass insgesamt

33 Ereignisse vorliegen. Ein solches Ergebnis ist deshalb aufgrund der geringen verfügbaren und der noch geringeren zusätzlichen Information wenig belastbar.

Tabelle 4-57: Subgruppen der Patienten mit viszeralen Metastasen *Viszeral* versus Patienten mit nicht-viszeralen Metastasen *Nicht Viszeral* für das Gesamtüberleben

Viszeral			
	Pla+T+D	Ptz+T+D	Total
Patienten	90	88	178
Erster Datenschnitt			
Events (%)	7 (7,8)	7 (8,0)	14
Zweiter Datenschnitt			
Events (%)	14 (15,5)	19 (21,6)	33

Diese Einschätzung kann durch eine Simulation unterlegt werden: Dabei wurden die zensierten Patienten des Pla+T+D Arms zufällig ausgewählt und ihr Status auf ‚Ereignis' gesetzt. Das ‚neue' Hazard-Ratio wurde dann berechnet und ferner ein mittleres HR, das auf 1.000 Wiederholungen der Simulation beruht.

In der ersten Simulation wurden 5 Patienten zufällig ausgewählt, was zur gleichen Anzahl von Ereignissen in beiden Studienarmen führt. Hierbei war das mittlere HR über die 1.000 Analysen 1,084, was keinen Unterschied zwischen den Behandlungsarmen entspräche.

Wird die Anzahl der zufällig ausgewählten zensierten Patienten unter Pla+T+D, die auf ‚Ereignis' gesetzt werden verringert, so ergeben sich die HR-Werte gemäß Tabelle 4-58.

Tabelle 4-58: Ergebnis der Simulation des OS: Einfluss zusätzlicher Ereignisse auf die geschätzten HR, Ergebnissicherheit

Zusätzliche Anzahl Ereignisse*	**Angenommene Anzahl Ereignisse in der Analyse**	**Mittleres HR§**	**SD**	**Min**	**Max**
1	15 (Pla) vs 19 (Ptz)	1,338	0,0211	1,322	1,423
2	16 (Pla) vs 19 (Ptz)	1,263	0,0276	1,236	1,423
3	17 (Pla) vs 19 (Ptz)	1,200	0,0294	1,163	1,371
4	18 (Pla) vs 19 (Ptz)	1,140	0,0313	1,099	1,305
5	19 (Pla) vs 19 (Ptz)	1,084	0,0320	1,039	1,258

* Gibt die Anzahl der ursprünglich zensierten Patienten wider, die zufällig aus dem Pla+T+D Arm gezogen wurden und für die Simulation als Ereignisse gewertet wurden.
§: Anzahl der Simulationen, jeweils 1.000.

Diese explorativen Simulationen beschreiben eine hypothetische Situation, die jedoch den Schluss nahelegen, dass das HR von 1,42, das in der Subgruppe *Nicht Viszeral* beobachtet wurde, sehr sensitiv für Veränderungen ist, was eine Konsequenz der sehr geringen Zahl an

Ereignissen in dieser Subgruppe darstellt. Selbst wenn nur 1 oder 2 Ereignisse unter Pla+T+D hinzukommen verändert sich das HR erheblich. Deshalb sollte dieser Schätzwert mit großer Vorsicht interpretiert werden.

Für die Subgruppe ‚FISH' ist das Bild über die Endpunkte OS und PFS weniger konsistent, so dass hier der Schluss auf einen Zufallsbefund auch ohne Simulationen nahe liegt.

- **Konsistenz der Ergebnisse über einzelnen Endpunkte hinweg**

Ein sich in der Realität abbildender Subgruppeneffekt muss sich bei allen Endpunkten manifestieren, die in die ‚gleiche inhaltliche Richtung' zeigen (z.B. bei den Wirksamkeitsvariablen).

- **Vorhandensein einer biologischen Rationale, um eine sinnvolle Interpretation zu ermöglichen.**

Subgruppe „nicht-viszerale Erkrankung"

Zwei der analysierten Baseline-Faktoren waren die viszerale und nicht-viszerale Erkrankung. Als nicht-viszerale Läsionen galten Tumore in Brust, Knochen, Knochenmark, Lymphknoten, Haut und Weichteilen. Alle anderen Tumorlokalisationen wurden als viszerale Metastasen eingestuft. Patienten, die sowohl viszerale als auch nicht-viszerale Läsionen aufwiesen (z.B. ein Patient mit einem Tumor in der Leber und einer Knochenmetastase), wurden für diese Analyse der Gruppe mit viszeraler Erkrankung zugeordnet.

Die Subgruppenanalyse für das Gesamtüberleben (OS) ergab für die Subgruppe der Patienten mit ausschließlich nicht-viszeraler Erkrankung ein HR von 1,42 mit einer sehr hohen Variabilität um die Punktschätzung: 95 % KI [0,71; 2,84]. Denn auch bei der finalen Analyse des Gesamtüberlebens waren in dieser Subgruppe nur relativ wenige Todesfälle zu verzeichnen (33 Todesfälle bei insgesamt 178 Patienten (18,5 %)). Die 1-Jahres-Überlebensraten waren in beiden Behandlungsarmen hoch (0,93 bzw. 0,95).

Mögliche Ursachen für dieses Ergebnis lassen sich in den demografischen Charakteristika und Krankheitsmerkmalen zu Studienbeginn in der Subgruppe der Patienten mit nicht-viszeraler Erkrankung finden:

Mit der Untersuchung der Baseline-Daten sollte festgestellt werden, ob die beobachteten Ergebnisse für Patienten mit nicht-viszeraler Erkrankung auf ein verdecktes Ungleichgewicht der demografischen Charakteristika und Krankheitsmerkmale zu Studienbeginn zwischen Patienten mit nicht-viszeraler Erkrankung und Patienten mit viszeralen Erkrankung (Tabelle 4-59) oder zwischen den Behandlungsarmen (Tabelle 4-60) zurückzuführen sind.

Unabhängig vom Behandlungsarm wurden die demografischen Charakteristika und Krankheitsmerkmale zu Studienbeginn für Patienten mit nicht-viszeraler und mit viszeraler Erkrankung verglichen (Tabelle 4-59). Wie zu erwarten, waren Metastasen bei Patienten mit nicht-viszeraler Erkrankung (gemäß o. g. Definition) weniger wahrscheinlich, sie hatten weniger Läsionen, es waren weniger Organe betroffen und es lag eine niedrigere messbare

Tumorlast vor als bei Patienten mit viszeraler Erkrankung. Es bestand bei ihnen außerdem die Tendenz zu einem besseren Wert in der ECOG-Skala. Das Verteilungsmuster der Metastasen unterschied sich beträchtlich zwischen diesen beiden Patientengruppen. **Diese Unterschiede, die durch die Definition der viszeralen Erkrankung zu erwarten waren erklären den relativ günstigen Krankheitsverlauf bei Patienten mit nicht-viszeraler Erkrankung im Vergleich zu Patienten mit viszeraler Erkrankung (unabhängig vom Behandlungsarm).**

Tabelle 4-59: Demografische Charakteristika und Krankheitsmerkmale zu Studienbeginn bei Patienten mit nicht-viszeraler Erkrankung vs. Patienten mit viszeraler Erkrankung

	Nicht-viszeral n = 178	**Viszeral n = 630**
Metastasen	164 (92,1)	623 (99,2)
Mediane Anzahl der Läsionen (Ziel- + Nicht-Zielveränderungen)	3,0	6,0
Nur 1 Läsion	48 (27,0)	22 (3,5)
2 Läsionen	27 (15,2)	53 (8,4)
3 Läsionen	37 (20,8)	73 (11,6)
> 4 Läsionen	66 (37,0)	482 (76,5)
Anzahl betroffener Organe		
1 Organ	73 (41,0)	70 (11,1)
2 Organe	67 (37,6)	158 (25,1)
3 Organe	37 (20,8)	183 (29,0)
> 4 Organe	1 (0,6)	219 (34,8)
Mediane messbare Tumorlast (Bewertung durch Prüfarzt)	68,0 mm	83,0 mm
ECOG 0	126 (70,8)	396 (62,9)
ECOG 1	52 (29,2)	230 (36,5)
ER / PR +	92 (51,7)	296 (47,0)
IHC 3+	156 (88,1)	565 (89,8)
IHC 2+	19 (10,7)	60 (9,5)
Vorangegangene (neo)adjuvante Therapie	83 (46,6)	293 (46,5)
Vorangegangene Behandlung mit Trastuzumab	22 (12,4)	66 (10,5)
Medianes behandlungsfreies Intervalls (Monate)*	31,5	29,5

*Behandlungsfreies Intervall ist definiert als Zeitraum vom Ende der systemischen Therapie (Chemotherapie oder zielgerichtete -Therapie) bis zum Auftreten von Metastasen bei Patienten, die eine vorangegangene (neo)adjuvante Therapie erhielten.

In Tabelle 4-60 werden wesentliche demografische Charakteristika und Krankheitsmerkmale zu Studienbeginn nach Behandlungsarm verglichen. Der auffälligste Unterschied zwischen den beiden Behandlungsarmen innerhalb der Subgruppe der Patienten mit nicht-viszeraler Erkrankung war die Dauer des behandlungsfreien Intervalls (Median 42,0 Monate bei Patienten im Pla+T+D-Arm vs. 28,0 Monate bei Patienten im Ptz+T+D-Arm). Dies war auch erheblich länger als das behandlungsfreie Intervall bei Patienten mit viszeraler Erkrankung (28,0 Monate im Pla+T+D-Arm und 32,0 Monate im Ptz+T+D-Arm). Dieses Ergebnis deutet darauf hin, dass die Patienten der Subgruppe mit nicht-viszeraler Erkrankung im Pla+T+D-Arm im Vergleich zu anderen Patienten einen eher indolenten Krankheitsverlauf aufwiesen und / oder besonders gut auf die vorangegangene (neo-)adjuvante Therapie angesprochen hatten. Es legt außerdem den Schluss nahe, dass sie eine Chemotherapie-sensible Erkrankung hatten.

Patienten in der Subgruppe mit nicht-viszeraler Erkrankung im Pla+T+D-Arms hatten weniger Läsionen (Ziel- und Nicht-Zielläsionen), bei ihnen waren weniger Organe betroffen und ihre messbare Tumorlast war geringer als bei Patienten des Ptz+T+D-Arms in der Subgruppe mit nicht-viszeraler Erkrankung. Es gab weitere Baseline-Faktoren in den beiden Behandlungsarmen der Subgruppe mit nicht-viszeraler Erkrankung, für die ein Unterschied von > 5 % in der Inzidenz nachgewiesen werden konnte:

- Mehr Patienten im Pla+T+D-Arm hatten nur Knochenmetastasen (26,7 % vs. 15,9 %).
- Mehr Patienten im Pla+T+D-Arm hatten eine IHC3+-Erkrankung (92,2 % vs. 83,9 %)
- Mehr Patienten im Ptz+T+D-Arm hatten eine IHC2+-Erkrankung (13,8 % vs. 7,8 %).
- Mehr Patienten im Ptz+T+D-Arm hatten eine ER / PR+-Erkrankung (54,5 % vs. 48,9 %).
- Mehr Patienten im Ptz+T+D-Arm hatten eine vorangegangene Behandlung mit Trastuzumab erhalten (15,9 % vs. 8,9 %)

Tabelle 4-60: Wichtige demografische Charakteristika und Krankheitsmerkmale zu Studienbeginn in den Behandlungsarmen (nach viszeraler / nicht-viszeraler Erkrankung)

	Nicht-viszeral (Anzahl, %)		**Viszeral (Anzahl, %)**	
	Pla+T+D n=90	**Ptz+T+D n=88**	**Pla+T+D n=316**	**Ptz+T+D n=314**
ECOG 0	65 (72,2)	61 (69,3)	183 (57,9)	213 (67,8)
ECOG 1	25 (27,8)	27 (30,7)	132 (41,8)	98 (31,2)
ER / PR +	44 (48,9)	48 (54,5)	155 (49,1)	141 (44,9)
ER / PR -	42 (46,7)	39 (44,3)	154 (48,7)	173 (55,1)
IHC 3+	83 (92,2)	73 (83,9)	288 (91,4)	277 (88,2)
IHC 2+	7 (7,8)	12 (13,8)	25 (7,9)	35 (11,1)
Vorangegangene (neo-)adjuvante Therapie	41 (45,6)	42 (47,7)	151 (47,8)	142 (45,2)
Vorangegangene Therapie mit Trastuzumab	8 (8,9)	14 (15,9)	33 (10,4)	33 (10,5)
Medianes behandlungsfreies Intervalls (Monate)*	42,0	28,0	28,0	32,0
Nur Knochenmetastasen	24 (26,7)	14 (15,9)	0	0
Knochen und andere Lokalisation	25 (27,8)	33 (37,5)	k. A.	k. A.
Weichteilmetastasen	41 (45,6)	41 (46,6)	k. A.	k. A.
Gesamtanzahl metastatischer Läsionen**				
1	30 (33,3)	18 (20,5)	10 (3,2)	12 (3,8)
2	14 (15,6)	13 (14,8)	27 (8,5)	26 (8,3)
3	14 (15,6)	23 (26,1)	34 (10,8)	39 (12,4)
≥ 4	32 (35,6)	34 (38,6)	245 (77,5)	237 (75,5)
Anzahl betroffener Organe				
1	42 (46,7)	31 (35,2)	37 (11,7)	33 (10,5)
2	32 (35,6)	35 (39,8)	88 (27,8)	70 (22,3)
3	16 (17,8)	21 (23,9)	87 (27,5)	96 (30,6)
≥ 4	0	1 (1,1)	104 (32,9)	115 (36,6)
Mediane messbare Tumorlast	65,5 mm	70,0 mm	84,0 mm	82,5 mm

k. A.: nicht untersucht

*Behandlungsfreies Intervall ist definiert als Zeitraum vom Ende der systemischen Therapie (Chemotherapie oder zielgerichtete Therapie) bis zum Auftreten von Metastasen bei Patienten, die eine vorangegangene (neo)adjuvante Therapie erhielten.

** Ziel + Nicht-Zielläsionen

Insgesamt lässt sich feststellen, dass das in der Studie beobachtete HR von 1,42 (95-%-KI [0,71; 2,84]) für das Gesamtüberleben in der Subgruppe der Patienten mit nicht-viszeraler Erkrankung möglicherweise durch ein zufälliges Ungleichgewicht in den Baseline-Faktoren bedingt ist. Die Patienten des Pla+T+D-Arms in der Subgruppe mit nicht-viszeraler Erkrankung schienen eine besonders begünstigte Subgruppe zu sein, die von einem längeren behandlungsfreien Intervall (deutet auf eher indolente Tumoren und / oder besseres Ansprechen auf Chemotherapie hin), weniger Läsionen, weniger betroffenen Organen und einer geringeren Tumorlast profitierten im Vergleich zu den Patienten im Ptz+T+D-Arm. Außerdem hatten diese Patienten mit höherer Wahrscheinlichkeit HER2-Tumoren mit einem immunhistochemische Wert von 3+ (IHC3+) und eine auf die Knochen beschränkte Metastasierung sowie mit geringerer Wahrscheinlichkeit ER / PR-positive Tumoren und eine vorangegangene Behandlung mit Trastuzumab erhalten als Patienten im Ptz+T+D-Arm. Zusammengenommen führen diese Faktoren wahrscheinlich zu der langsamen Ereignisrate bei dem Gesamtüberleben, dem verlängerten Gesamtüberleben und dem guten Ansprechen auf die Therapie aus Trastuzumab+Docetaxel in der Subgruppe mit nicht-viszeraler Erkrankung im Pla+T+D-Arm. Das Konfidenzintervall für die Hazard Ratio in der Subgruppe mit nicht-viszeraler Erkrankung in der CLEOPATRA Studie schließt dabei nicht aus, dass die tatsächliche Hazard Ratio in dieser Subgruppe konsistent mit der ITT-Population ist.

Grundsätzlich gibt es keinen nachvollziehbaren biologischen Grund, warum die zusätzliche Gabe von Pertuzumab eine geringere Wirkung bei Patienten mit nicht-viszeralem, HER2-positivem, metastasierten Mammakarzinom zeigen sollte. In der Zulassungsstudie von Trastuzumab in Kombination mit Docetaxel waren bei Patienten mit viszeraler und nicht-viszeraler Erkrankung konsistente Verbesserungen in den Ansprechraten beobachtet worden, wenn Trastuzumab zusätzlich zu Docetaxel verabreicht wurde. [69] Daten aus der Studie WO20697 (NEOSPHERE) zeigten deutlich, dass Ptz+T+D die Wirksamkeit bei Patienten verbesserte, die ausschließlich nicht-viszerale Metastasen (Weichteile und Lymphknoten) aufwiesen. [70] Beim Vergleich der pathologischen Vollremissionsraten (pCR-Rate) zwischen Patienten unter T+D versus Patienten unter Ptz+T+D zeigte sich eine klinisch relevante und statistisch signifikante Verbesserung der pCR-Rate in der Brust bei zusätzlicher Gabe von Pertuzumab (pCR-Rate 29,0 % im T+D-Arm vs. 45,8 % im Ptz+T+D-Arm; Differenz in den pCR-Raten = 16,8 %; 95 % KI [3,5; 30,1], p = 0,0141). Dieser Vorteil war auch nachweisbar, wenn die pCR-Rate in der Brust und in den Lymphknoten beurteilt wurde (pCR-Rate 21,5 % unter Pla+T+D vs. 39,3 % unter Ptz+T+D).

Ausgehend von den oben genannten Überlegungen und Unsicherheiten bezüglich der richtigen Schätzung der Therapiewirkung in der Subgruppe der Patienten mit nicht-viszeraler Erkrankung kann nicht davon ausgegangen werden, dass Patienten mit ausschließlich nicht-viszeraler Erkrankung einen geringeren Vorteil von der zusätzlichen Gabe von Pertuzumab haben. Basierend auf diesen Überlegungen, sieht die Zulassungsbehörde EMA keinen Grund Patienten mit nicht-viszeraler Erkrankung aus der Indikation auzuschließen. [71]

Subgruppe „Vorangegangene (neo-)adjuvante Therapie mit Trastuzumab“ und „Vorangegangene (neo-)adjuvante Therapie mit Taxan“

Es besteht Konsens, dass bei einem krankheitsfreien Intervall von mindestens einem Jahr eine erneute Therapie mit einem Taxan oder Trastuzumab erfolgen kann. [37] Patienten konnten nach einer (neo-)adjuvanten Vortherapie in die CLEOPATRA-Studie aufgenommen werden, wenn zwischen Ende dieser Therapie ein krankheitsfreies Intervall von mindestens 12 Monaten lag.

Um die Wirksamkeit von Pertuzumab nach einer vorangegangenen (neo-)adjuvanten Therapie mit Trastuzumab und / oder Taxanen beurteilen zu können, wurden entsprechende Subgruppen-Analysen durchgeführt. Diese zeigen, dass auch Patienten, die zuvor (neo-) adjuvant mit Trastuzumab und / oder Taxanen behandelt wurden, maßgeblich von einer Ptz+T+D-Behandlung profitieren. Somit zeigt sich bei der Auswertung zum Gesamtüberleben für die Kombinationsbehandlung mit Pertuzumab gegenüber der Standardtherapie T+D ein vergleichbares Ausmaß wie in der gesamten Studienpopulation.

4.3.1.3.2.1 <Mortalität> – RCT

4.3.1.3.2.1.1 <Endpunkt Gesamtüberleben> - RCT

Tabelle 4-61: Subgruppenergebnisse für<Endpunkt Gesamtüberleben> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (finale konfirmatorische Analyse. 2. Datenschnitt Mai 2012)

Parameter	Subgruppenmerkmal	Placebo+Trastuzumab+Docetaxel	Pertuzumab+Trastuzumab+Docetaxel	Ptz+T+D vs Pla+T+D	Interaktions-test
		Patienten mit Ereignis: n/N (%)	Patienten mit Ereignis n/N (%)	Hazard Ratio [95% KI]	p-Wert
Vorausgehender Behandlungsstatus[a]	De novo	80 / 214 (37,4 %)	60 / 218 (27,5 %)	0,66 [0,47; 0,93]	0,9892
	Adjuv. od. neoadjuvante Behandlung	74 / 192 (38,5 %)	53 / 184 (28,8 %)	0,66 [0,46; 0,94]	
Region[a]	Europa	55 / 152 (36,2 %)	44 / 154 (28,6 %)	0,72 [0,48; 1,07]	0,8575
	Nordamerika	23 / 68 (33,8 %)	16 / 67 (23,9 %)	0,68 [0,36; 1,28]	
	Südamerika	30 / 58 (51,7 %)	19 / 56 (33,9 %)	0,55 [0,31; 0,98]	
	Asien	46 / 128 (35,9 %)	34 / 125 (27,2 %)	0,64 [0,41; 1,00]	
Altersgruppe[a]	<65	126 / 339 (37,2 %)	97 / 342 (28,4 %)	0,70 [0,53; 0,91]	0,3190
	≥65	28 / 67 (41,8 %)	16 / 60 (26,7 %)	0,51 [0,27; 0,95]	
Altersgruppe[a]	<75	147 / 392 (37,5 %)	110 / 397 (27,7 %)	0,66 [0,52; 0,85]	0,6862
	≥75	7 / 14 (50,0 %)	3 / 5 (60,0 %)	0,72 [0,15; 3,50]	
Rasse[a]	Kaukasisch	91 / 235 (38,7 %)	73 / 245 (29,8 %)	0,70 [0,51; 0,95]	0,6425
	Afro-amerikanisch	10 / 20 (50,0 %)	3 / 10 (30,0 %)	0,52 [0,14; 1,91]	
	Asiatisch	47 / 133 (35,3 %)	35 / 128 (27,3 %)	0,66 [0,43; 1,03]	
	Weitere	6 / 18 (33,3 %)	2 / 19 (10,5 %)	0,29 [0,06; 1,43]	
Art der Erkrankung[a]	Viszeral	140 / 316 (44,3 %)	94 / 314 (29,9 %)	0,57 [0,44; 0,74]	0,0138
	Nicht-viszeral	14 / 90 (15,6 %)	19 / 88 (21,6 %)	1,42 [0,71; 2,84]	

Parameter	Subgruppenmerkmal	Placebo+Trastuzumab+Docetaxel	Pertuzumab+Trastuzumab+Docetaxel	Ptz+T+D vs Pla+T+D	Interaktions-test
		Patienten mit Ereignis: n/N (%)	Patienten mit Ereignis n/N (%)	Hazard Ratio [95% KI]	p-Wert
IHC-HER2-Status[a]	3+	139 / 371 (37,5 %)	97 / 350 (27,7 %)	0,66 [0,51; 0,85]	0,6987
FISH-Status[a]	FISH Positiv	143 / 383 (37,3 %)	106 / 384 (27,6 %)	0,67 [0,52; 0,86]	0,1900
ER/PgR-Status[a]	ER und/oder PgR Positiv ER und PgR Negativ Unbekannt	65 / 199 (32,7 %) 88 / 196 (44,9 %) 1 / 11 (9,1 %)	48 / 189 (25,4 %) 64 / 212 (30,2 %) 1 / 1 (100,0 %)	0,73 [0,50; 1,06] 0,57 [0,41; 0,79] 8,94 [0,56; 143,6]	0,3272
IHC-HER2-Status[a]	2+	13 / 32 (40,6 %)	14 / 47 (29,8 %)	0,68 [0,32; 1,44]	0,6987
Baseline ECOG[b]	0 1+				0,5854
Vorherige (neoajduvante oder adjuvante) Behandlung mit Trastuzumab[a]	Ja Nein	14 / 41 (34,1%)	10 / 47 (21,3 %)	0,68 [0,30; 1,55]	0,9207
Dosiseskalation mit Docetaxel auf 100 mg/m^2[b]	Ja Nein				0,8124
Vorherige (neo-) adjuvante Taxan-behandlung[a]	Ja Nein	30 / 94 (31,9 %) 124 / 312 (39,7 %)	22 / 91 (24,2 %) 91 / 311 (29,3 %)	0,72 [0,41; 1,25] 0,64 [0,49; 0,84]	0,7076

Ereignis bezieht sich hier auf den Tod
Kursive Werte wurden nachträglich ausgerechnet auf Basis der Daten aus dem CSR
a: CSR update 12/2012 S.305-307, ITT Population, + und zusätzliche Analysen [55]
b: Post-hoc OS Analysen wurden nur für G-BA vorgeschlagene Subgruppen zusätzlich durchgeführt.

Der p-Wert als Ergebnis des Interaktionstests für das Merkmal „Lokalrezidiv“ liegt bei 0,0703.

Für diesen Endpunkt sind über alle Subgruppen hinweg keine Interaktionentestergebnisse vorhanden, die einen wirklichen Hinweis auf eine mögliche Interaktion geben könnten. Das Auftreten signifikanter Interaktionen kann mit hoher Wahrscheinlichkeit zufällig sein, da insgesamt weniger signifikante Interaktionen zu verzeichnen sind, als alleine aufgrund Zufall zu erwarten wären.

Zudem sind für die wenigen Interaktionstestergebnisse, die formal in eine der beiden IQWiG-Kategorien für Interaktionstests fallen, auf Einzelvariablenebene in Kapitel 4.3.1.3.2 genannte Kriterien nicht erfüllt:

- Merkmal Viszeral / nicht-viszeral Viszeral:
    - Wenige Daten vorhanden
    - Biologische Rationale fehlt
- Merkmal Fish-Status: Kein konsistentes Bild über die Endpunkte
- Merkmal Lokal-Rezidiv: Kein konsistentes Bild über die Endpunkte

4.3.1.3.2.2 <Morbidität> – RCT

4.3.1.3.2.2.1 <Primärer Endpunkt Progressionsfreies Überleben bewertet durch IRF> - RCT

Tabelle 4-62: Subgruppenergebnisse für<Primärer Endpunkt Progressionsfreie Überleben bewertet durch IRF> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (finale konfirmatorische Analyse, 1. Datenschnitt Mai 2011)

Parameter	**Subgruppen-merkmal**	**Placebo+Trastuzumab +Docetaxel**	**Pertuzumab+Trastuzumab +Docetaxel**	**Ptz+T+D vs Pla+T +D Hazard Ratio [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**	**Interaktions-test p-Wert[o]**
		Patienten mit Ereignis n/N (%)	**Patienten mit Ereignis n/N (%)**			
Vorausgehender Behandlungsstatus[a]	De novo	125 / 214 (58,4 %)	105 / 218 (48,2 %)	0,63 [0,49; 0,82]	0,006	0,8960
	Adjuv. od. neoadjuvante Behandlung	117 / 192 (60,9 %)	86 / 184 (46,7 %)	0,61 [0,46; 0,81]	0,006	
Region[b]	Europa	92 / 152 (60,5 %)	80 / 154 (51,9 %)	0,72 [0,53; 0,97]	0,0327	0,3381
	Nordamerika	41 / 68 (60,3 %)	27 / 67 (40,3 %)	0,51 [0,31; 0,84]	0,0075	
	Südamerika	36 / 58 (62,1 %)	22 / 56 (39,3 %)	0,46 [0,27; 0,78]	0,0040	
	Asien	73 / 128 (57,0 %)	62 / 125 (49,6 %)	0,68 [0,48 0,95]	0,0246	
Altersgruppe[c]	<65	204 / 339 (60,2 %)	165 / 342 (48,2 %)	0,65 [0,53; 0,80]	< 0,0001	0,4180
	≥65	38 / 67 (56,7 %)	26 / 60 (43,3 %)	0,52 [0,31; 0,86]	0,0112	
Altersgruppe [d]	<75	232 / 392 (59,2 %)	189 / 397 (47,6 %)	0,64 [0,53; 0,78]	< 0,0001	0,3104
	≥75	10 / 14 (71,4 %)	2 / 5 (40,0 %)	0,55 [0,12; 2,54]	0,4439	
Rasse[e]	Kaukasisch	143 / 235 (60,9 %)	117 / 245 (47,8 %)	0,62 [0,49; 0,80]	0,0001	0,8037
	Afro-Amerikanisch	14 / 20 (70,0 %)	5 / 10 (50,0 %)	0,64 [0,23; 1,79]	0,3927	
	Asiatisch	76 / 133 (57,1 %)	64 / 128 (50,0 %)	0,68 [0,49; 0,95]	0,0248	
	Weitere	9 / 18 (50,0 %)	5 / 19 (26,3 %)	0,39 [0,13; 1,18]	0,0960	
Art der Erkrankung[f]	Viszeral	205 / 316 (64,9 %)	155 / 314 (49,4 %)	0,55 [0,45; 0,68]	< 0,0001	0,0332
	Nicht-viszeral	37 / 90 (41,1 %)	36 / 88 (40,9 %)	0,96 [0,61; 1,52]	0,8629	

Parameter	**Subgruppen-merkmal**	**Placebo+Trastuzumab +Docetaxel**	**Pertuzumab+Trastuzumab +Docetaxel**	**Ptz+T+D vs Pla+T +D Hazard Ratio [95% KI]**	**Ptz+T+D vs Pla+T +D p-Wert**	**Interaktions-test p-Wert[o]**
		Patienten mit Ereignis n/N (%)	**Patienten mit Ereignis n/N (%)**			
IHC-HER2-Status[g]	3+	214 / 371 (57,7 %)	158 / 350 (45,1 %)	0,60 [0,49; 0,74]	< 0,0001	0,1252
FISH-Status[h]	FISH Positiv	225 / 383 (58,7 %)	182 / 384 (47,4 %)	0,64 [0,53; 0,78]	< 0,0001	0,0254
ER / PgR-Status[i]	ER und/od. PgR+ ER und PgR - Unbekannt	113 / 199 (56,8 %) 122 / 196 (62,2 %) 7 / 11 (63,6 %)	88 / 189 (46,6 %) 102 / 212 (48,1 %) 1 / 1 (100,0 %)	0,72 [0,55; 0,95] 0,55 [0,42; 0,72] >100 [0,00; n.d.]	0,0217 < 0,0001 0,9964	0,1277
IHC-HER2-Status[j]	2+	25 / 32 (78,1 %)	31 / 47 (66,0 %)	0,90 [0.53; 1.54]	0,7087	0,1252
Baseline ECOG[k]	0 1+	131 / 248 (52,8 %) 111 / 158 (70,3 %)	123 / 274 (44,9 %) 68 / 128 (53,1 %)	0,70 [0,55; 0,90] 0,55 [0,40; 0,74]	0,0048 0,0001	0,2132
Vorherige (neo-ajduvante oder adjuvante) Behandlung mit Trastuzumab[l]	Ja Nein	28 / 41 (68,3 %) 89 / 151 (58,9 %)	24 / 47 (51,1 %) 62 / 137 (45,3 %)	0,62 [0,35; 1,07] 0,60 [0,43; 0,83]	0,0878 < 0,0001	0,9263
Dosiseskalation mit Docetaxel auf 100 mg / m^2[m]	Ja Nein	37 / 63 (58,7 %) 205 / 343 (59,8 %)	19 / 46 (41,3 %) 172 / 356 (48,3 %)	0,65 [0,37; 1,13] 0,62 [0,50; 0,76]	0,1233 < 0,0001	0,8543
Vorherige (neo-) adjuvante Taxan-behandlung[n]	Ja Nein	55 / 94 (58,5 %) 187 / 312 (59,9 %)	41 / 91 (45,1 %) 150 / 311 (48,2 %)	0,64 [0,43; 0,97] 0,62 [0,50; 0,76]	0,0347 < 0,0001	0,7970

Parameter	Subgruppen-merkmal	Placebo+Trastuzumab +Docetaxel	Pertuzumab+Trastuzumab +Docetaxel	Ptz+T+D vs Pla+T +D Hazard Ratio [95% KI]	Ptz+T+D vs Pla+T +D p-Wert	Interaktions-test p-Wert[o]
		Patienten mit Ereignis n/N (%)	Patienten mit Ereignis n/N (%)			
Ereignis bezieht sich hier auf das Auftreten einer Progression a-e: Tabelle 26 aus CSR 10/2011; ITT Population f: Tabelle aus CSR 10/2011 S. 1116 und 1117; ITT Population g: Tabelle aus CSR 10/2011 S. 1641 und 1115; ITT Population h: Tabelle 26 aus CSR 10/2011; ITT Population i: Tabelle 26 aus CSR 10/2011; ITT Population j: Tabelle aus CSR 10/2011 S. 1640 und 1115; ITT Population k: Tabelle aus CSR 10/2011 S. 1644, 1645 und 1115; ITT Population l: Tabelle aus CSR 10/2011 S. 1642 und 1643; ITT Population m: Tabelle aus CSR 10/2011 S. 1649 und 1648; ITT Population n: zusätzliche Analysen [55] o: Interaktionsterme: Tabelle aus CSR 10/2011 S. 1115, und zusätzliche Analysen [55]						

Der p-Wert als Ergebnis des Interaktionstests für das Merkmal „Lokalrezidiv“ liegt bei 0,4245.

Für diesen Endpunkt sind über alle Subgruppen hinweg keine Interaktionentestergebnisse vorhanden, die einen wirklichen Hinweis auf eine mögliche Interaktion geben könnten. Das Auftreten signifikanter Interaktionen kann mit hoher Wahrscheinlichkeit zufällig sein, da insgesamt weniger signifikante Interaktionen zu verzeichnen sind, als alleine aufgrund Zufall zu erwarten wären.

Zudem sind für die wenigen Interaktionstestergebnisse, die formal in eine der beiden IQWiG-Kategorien für Interaktionstests fallen, auf Einzelvariablenebene in Kapitel 4.3.1.3.2 genannte Kriterien nicht erfüllt:

- Merkmal Viszeral / nicht-viszeral Viszeral:
  - Wenige Daten vorhanden
  - Biologische Rationale fehlt

- Merkmal Fish-Status: Kein konsistentes Bild über die Endpunkte
- Merkmal ER / PgR: Kein konsistentes Bild über die Endpunkte
- Merkmal IHC-HER2-Status: Kein konsistentes Bild über die Endpunkte

4.3.1.3.2.2.2 <Endpunkt Gesamtansprechrate> - RCT

Tabelle 4-63: Subgruppenergebnisse für<Endpunkt Gesamtansprechrate> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (finale konfirmatorische Analyse, 1. Datenschnitt Mai 2011)

Parameter	**Subgruppenmerkmal**	**Placebo+Trastuzumab +Docetaxel**	**Pertuzumab+Trastuzumab +Docetaxel**	**Ptz+T+D vs Pla+T+D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T+D p-Wert**	**Interaktions-test p-Wert**
		Patienten mit Ereignis n/N (%)	**Patienten mit Ereignis n/N (%)**			
Vorausgehender Behandlungsstatus[a]	De novo	126 / 185 (68,11 %)	166 / 199 (83,42 %)	1,22 [1,09; 1,38]	0,0005	0,0919
	Adjuv. od. neoadjuvante Behandlung	107 / 151 (70,86 %)	109 / 144 (75,69 %)	1,07 [0,93; 1,23]	0,3495	
Region[b]	Europa	79 / 121 (65,29 %)	102 / 130 (78,46 %)	1,20 [1,03; 1,41]	0,0203	0,7035
	Nordamerika	37 / 55 (67,27 %)	42 / 57 (73,68 %)	1,10 [0,86; 1,39]	0,4588	
	Südamerika	33 / 50 (66,00 %)	38 / 45 (84,44 %)	1,28 [1,01; 1,62]	0,0399	
	Asien	84 / 110 (76,36 %)	93 / 111 (83,78 %)	1,10 [0,96; 1,25]	0,1682	
Altersgruppe[c]	<65	189 / 278 (67,99 %)	233 / 293 (79,52 %)	1,17 [1,06; 1,29]	0,0017	0,8646
	≥65	44 / 58 (75,86 %)	42/ 50 (84,00 %)	1,11 [0,92; 1,34]	0,2973	
Altersgruppe [d]	<75	224 / 323 (69,35 %)	272 / 338 (80,47 %)	1,16 [1,06; 1,27]	0,001	0,3729
	≥75	9 / 13 (69,23 %)	3 / 5 (60,00 %)	0,87 [0,39; 1,93]	1,000	
Rasse[e]	Kaukasisch	120 / 189 (63,49 %)	158 / 204 (77,45 %)	1,22 [1,07; 1,39]	0,0024	0,4834
	Afro-amerikanisch	13 / 18 (72,22 %)	8 / 9 (88,89 %)	1,23 [0,85; 1,78]	0,6279	
	Asiatisch	87 / 115 (75,65 %)	95 / 113 (84,07 %)	1,11 [0,97; 1,27]	0,1140	
	Weitere	13 / 14 (92,86 %)	14 / 17 (82,35 %)	0,89 [0,68; 1,15]	0,6067	
Art der Erkrankung[f]	Viszeral	191 / 280 (68,21 %)	224 / 279 (80,29 %)	1,18 [1,07; 1,30]	0,0011	0,4391
	Nicht-viszeral	42 / 56 (75,00 %)	51 / 64 (79,69 %)	1,06 [0,87; 1,29]	0,5413	
IHC-HER2-Status [g]	3+	214 / 305 (70,16 %)	250 / 304 (82,24 %)	1,17 [1,07; 1,28]	0,0005	0,5540
FISH-Status[h]	Positiv	221 / 316 (69,94 %)	262 / 326 (80,37 %)	1,15 [1,05; 1,26]	0,0022	0,3795

Parameter	Subgruppenmerkmal	Placebo+Trastuzumab +Docetaxel	Pertuzumab+Trastuzumab +Docetaxel	Ptz+T+D vs Pla+T+D Relatives Risiko [95% KI]	Ptz+T+D vs Pla+T+D p-Wert	Interaktions-test p-Wert
		Patienten mit Ereignis n/N (%)	Patienten mit Ereignis n/N (%)			
ER / PgR-Status[i]	ER und/oder PgR Positiv ER und PgR Negativ	118 / 162 (72,84 %) 106 / 164 (64,63 %)	122 / 153 (79,74 %) 153 / 189 (80,95 %)	1,09 [0,97; 1,24] 1,25 [1,10; 1,43]	0,1514 0,0006	0,2065
IHC-HER2-Status[j]	2+	17 / 28 (60,71 %)	22 / 35 (62,86 %)	1,04 [0,70; 1,53]	0,8629	0,5540
Baseline ECOG[k]	0 1+	146 / 202 (72,28 %) 87 / 134 (64,93 %)	192 / 238 (80,67 %) 83 / 105 (79,05 %)	1,12 [1,00; 1,24] 1,22 [1,04; 1,43]	0,0378 0,0170	0,5205
Vorherige (neoajduvante oder adjuvante) Behandlung mit Trastuzumab [l]	Ja Nein	20 / 34 (58,82 %) 213 / 302 (70,53 %)	22 / 38 (57,89 %) 253 / 305 (82,95 %)	0,98 [0,67; 1,45] 1,18 [1,08; 1,29]	0,9368 0,0003	0,1491
Dosiseskalation mit Docetaxel auf 100 mg / m^2 [m]	Ja Nein	44 / 53 (83,02 %) 189 / 283 (66,78 %)	34 / 42 (80,95 %) 241 / 301 (80,07 %)	0,98 [0,81; 1,18] 1,20 [1,09; 1,32]	0,7952 0,0003	0,1482
Vorherige (neo-) adjuvante Taxanbehandlung [n]	Ja Nein	54 / 76 (71,05 %) 179 / 260 (68,85 %)	52 / 70 (74,29 %) 223 / 273 (81,68 %)	1,05 [0,86; 1,28] 1,19 [1,07; 1,31]	0,6628 0,0006	0,2066

Ereignis bezieht sich hier auf die Anzahl *Responder*, die durch IRF bewertet wurden, basierend auf RECIST. Die Anzahl N der Patienten bezieht sich auf diejenigen, bei denen eine durch IRF bewertete, feststellbare Erkrankung an der Baseline vorlag.
p-Wert wurde mittels Chi^2 Test bestimmt
a-n zusätzliche Analysen [55]

Der p-Wert als Ergebnis des Interaktionstests für das Merkmal „Lokalrezidiv“ liegt bei 0,4245.

Für diesen Endpunkt sind über alle Subgruppen hinweg keine Interaktionentestergebnisse vorhanden, die einen wirklichen Hinweis auf eine mögliche Interaktion geben könnten. Das Auftreten signifikanter Interaktionen kann mit hoher Wahrscheinlichkeit zufällig sein, da insgesamt weniger signifikante Interaktionen zu verzeichnen sind, als alleine aufgrund Zufall zu erwarten wären.

Zudem sind für die wenigen Interaktionstestergebnisse, die formal in eine der beiden IQWiG-Kategorien für Interaktionstests fallen, auf Einzelvariablenebene in Kapitel 4.3.1.3.2 genannte Kriterien nicht erfüllt:

- Merkmal Trastuzumab-Vorbehandlung: Kein konsistentes Bild über die Endpunkte
- Merkmal Dosiseskalation: Kein konsistentes Bild über die Endpunkte

4.3.1.3.2.3 <Lebensqualität> – RCT

4.3.1.3.2.3.1 <Endpunkt Zeit bis Symptomprogression (Gesundheitsbezogene Lebensqualität)> - RCT

Tabelle 4-64: Subgruppenergebnisse für<Endpunkt Zeit bis Symptomprogression (Gesundheitsbezogene Lebensqualität)> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (1. Datenschnitt Mai 2011)

Parameter	Subgruppenmerkmal	Placebo+Trastuzumab +Docetaxel	Pertuzumab+Trastuzumab +Docetaxel	Ptz+T+D vs Pla+T+D Hazard Ratio [95% KI]	Ptz+T+D vs Pla+T+D p-Wert	Inter-aktionstest p-Wert
		Patienten mit Ereignis n/N (%)	Patienten mit Ereignis n/N (%)			
Vorausgehender Behandlungsstatus[a]	De novo	117 / 213 (54,9 %)	136 / 218 (62,4 %)	1,02 [0,80; 1,31]	0,8740	0,5461
	Adjuv. od. neoadjuvante Behandlung	112 / 191 (58,6 %)	103 / 184 (56,0 %)	0,89 [0,68; 1,17]	0,4125	
Region[b]	Europa	79 / 152 (52,0 %)	78 / 154 (50,6 %)	0,81 [0,59; 1,11]	0,1819	0,4047
	Nordamerika	33 / 67 (49,3 %)	41 / 67 (61,2 %)	1,25 [0,79; 1,99]	0,3462	
	Südamerika	31 / 57 (54,4 %)	35 / 56 (62,5 %)	1,07 [0,66; 1,74]	0,7818	
	Asien	86 / 128 (67,2 %)	85 / 125 (68,0 %)	0,99 [0,73; 1,34]	0,9535	
Altersgruppe[c]	<65	196 / 337 (58,2 %)	205 / 342 (59,9 %)	0,94 [0,77; 1,14]	0,5090	0,4080
	≥65	33 / 67 (49,3 %)	34 / 60 (56,7 %)	1,23 [0,76; 2,00]	0,3998	
Altersgruppe [d]	<75	222 / 390 (56,9 %)	236 / 397 (59,4 %)	0,96 [0,80; 1,15]	0,6670	0,1448
	≥75	7 / 14 (50,0 %)	3 / 5 (60,0 %)	2,53 [0,60; 10,69]	0,2055	
Rasse[e]	Kaukasisch	119 / 233 (51,1 %)	133 / 245 (54,3 %)	0,99 [0,77; 1,26]	0,9169	0,4819
	Afro-amerikanisch	12 / 20 (60,0 %)	5 / 10 (50,0 %)	0,36 [0,12; 1,07]	0,0651	
	Asiatisch	88 / 133 (66,2 %)	88 / 128 (68,8 %)	1,00 [0,75; 1,35]	0,9771	
	Weitere	10 / 18 (55,6 %)	13 / 19 (68,4 %)	1,11 [0,49; 2,55]	0,7998	
Art der Erkrankung[f]	Viszeral	175 / 314 (55,7 %)	189 / 314 (60,2 %)	0,99 [0,81; 1,22]	0,9217	0,5869
	Nicht-viszeral	54 / 90 (60,0 %)	50 / 88 (56,8 %)	0,90 [0,61; 1,32]	0,5786	

Parameter	Subgruppenmerkmal	Placebo+Trastuzumab +Docetaxel	Pertuzumab+Trastuzumab +Docetaxel	Ptz+T+D vs Pla+T+D Hazard Ratio [95% KI]	Ptz+T+D vs Pla+T+D p-Wert	Inter-aktionstest p-Wert
		Patienten mit Ereignis n/N (%)	Patienten mit Ereignis n/N (%)			
IHC-HER2-Status[g]	3+	212 / 370 (57,3 %)	203 / 350 (58,0 %)	0,92 [0,76; 1,12]	0,4027	0,1566
FISH-Status[h]	Positiv	213 / 381 (55,9 %)	225 / 384 (58,6 %)	0,96 [0,79; 1,16]	0,6613	0,3704
ER / PgR-Status[i]	ER und/oder PgR Positiv ER und PgR Negativ Unbekannt	109 / 197 (55,3 %) 114 / 196 (58,2 %) 6 / 11 (54,5 %)	111 / 189 (58,7 %) 127 / 212 (59,9 %) 1 / 1 (100,0 %)	0,97 [0,74; 1,26] 0,96 [0,75; 1,24] 1,25 [0,14; 10,81]	0,7945 0,7707 0,8418	0,9850
IHC-HER2-Status[j]	2+	16 / 31 (51,6 %)	33 / 47 (70,2 %)	1,39 [0,76; 2,52]	0,2839	0,1566
Baseline ECOG[k]	0 1+	144 / 248 (58,1 %) 85 / 156 (54,5 %)	172 / 274 (62,8 %) 67 / 128 (52,3 %)	0,94 [0,75; 1,17] 1,00 [0,72; 1,37]	0,5664 0,9764	0,7782
Vorherige (neoajduvante oder adjuvante) Behandlung mit Trastuzumab [l]	Ja Nein	22 / 41 (53,7 %) 207 / 363 (57,0 %)	27 / 47 (57,4 %) 212 / 355 (59,7 %)	1,14 [0,64; 2,04] 0,95 [0,79; 1,16]	0,6497 0,6325	0,5716
Dosiseskalation mit Docetaxel auf 100 mg / m^2 [m]	Ja Nein	33 / 63 (52,4 %) 196 / 341 (57,5 %)	29 / 46 (63,0 %) 210 / 356 (59,0 %)	1,10 [0,67; 1,81] 0,94 [0,77; 1,14]	0,7141 0,5223	0,5503
Vorherige (neo-) adjuvante Taxanbehandlung [n]	Ja Nein	55 / 94 (58,5 %) 174 / 310 (56,1 %)	58 / 91 (63,7 %) 181 / 311 (58,2 %)	0,96 [0,66; 1,39] 0,97 [0,79; 1,19]	0,8247 0,7648	0,9845

Ereignis bezieht sich hier auf das Auftreten einer Symptomprogression
a-n zusätzliche Analysen [55]
Kursiv gezeigte Werte wurden nachträglich auf Basis der angegebenen Daten berechnet.

Der p-Wert als Ergebnis des Interaktionstests für das Merkmal „Lokalrezidiv“ liegt bei 0,6352.

Für diesen Endpunkt sind über alle Subgruppen hinweg keine Interaktionentestergebnisse vorhanden, die einen wirklichen Hinweis auf eine mögliche Interaktion geben könnten. Das Auftreten signifikanter Interaktionen kann mit hoher Wahrscheinlichkeit zufällig sein, da insgesamt weniger signifikante Interaktionen zu verzeichnen sind, als alleine aufgrund Zufall zu erwarten wären.

Zudem sind für die wenigen Interaktionstestergebnisse, die formal in eine der beiden IQWiG-Kategorien für Interaktionstests fallen, auf Einzelvariablenebene in Kapitel 4.3.1.3.2 genannte Kriterien nicht erfüllt:

- Merkmal Altersgruppe <75 / ≥75:
  - Wenige Daten vorhanden
  - Kein konsistentes Bild über die Endpunkte
- Merkmal IHC-HER2-Status: Kein konsistentes Bild über die Endpunkte

4.3.1.3.2.3.2 <Endpunkt Zeit bis zur Verschlechterung des BCS-Score (Gesundheitsbezogene Lebensqualität)> - RCT

Tabelle 4-65: Subgruppenergebnisse für<Endpunkt Zeit bis zur Verschlechterung (Gesundheitsbezogene Lebensqualität)> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (1. Datenschnitt Mai 2011)

Parameter	**Subgruppenmerkmal**	**Placebo+Trastuzumab +Docetaxel**	**Pertuzumab+Trastuzumab +Docetaxel**	**Ptz+T+D vs Pla+T+D**	**Ptz+T+D vs Pla + T +D**
		Patienten mit Ereignis n/N (%)	**Patienten mit Ereignis n/N (%)**	**Hazard Ratio [95% KI]**	**p-Wert**
Vorausgehender Behandlungsstatus[a]	De novo	122 / 213 (57,3 %)	122 / 218 (56,0 %)	0,80 [0,63; 1,04]	0,0909
	Adjuv. od. neoadjuvante Behandlung	117 / 191 (61,3 %)	100 / 184 (54,3 %)	0,73 [0,56; 0,96]	0,0238
Region[b]	Europa	82 / 152 (53,9 %)	77 / 154 (50,0 %)	0,78 [0,57; 1,07]	0,1236
	Nordamerika	32 / 67 (47,8 %)	34 / 67 (50,7 %)	0,88 [0,54; 1,44]	0,6192
	Südamerika	33 / 57 (57,9 %)	30 / 56 (53,6 %)	0,78 [0,47; 1,28]	0,3240
	Asien	92 / 128 (71,9 %)	81 / 125 (64,8 %)	0,73 [0,54; 0,99]	0,0436
Altersgruppe[c]	<65	202 / 337 (59,9 %)	190 / 342 (55,6 %)	0,78 [0,64; 0,96]	0,0166
	≥65	37 / 67 (55,2 %)	32 / 60 (53,3 %)	0,63 [0,38; 1,05]	0,0743
Altersgruppe [d]	<75	233 / 390 (59,7 %)	219 / 397 (55,2 %)	0,76 [0,63; 0,92]	0,0038
	≥75	6 / 14 (42,9 %)	3 / 5 (60,0 %)	6,33 [1,05; 38,10]	0,0441
Rasse[e]	Kaukasisch	124 / 233 (53,2 %)	121 / 245 (49,4 %)	0,78 [0,61; 1,01]	0,0565
	Afro-amerikanisch	10 / 20 (50,0 %)	5 / 10 (50,0 %)	0,68 [0,23; 2,03]	0,4883
	Asiatisch	94 / 133 (70,7 %)	84 / 128 (65,6 %)	0,76 [0,57; 1,02]	0,0682
	Weitere	11 / 18 (61,1 %)	12 / 19 (63,2 %)	0,72 [0,32; 1,64]	0,4364
Art der Erkrankung[f]	Viszeral	184 / 314 (58,6 %)	180 / 314 (57,3 %)	0,81 [0,66; 0,99]	0,0437
	Nicht-viszeral	55 / 90 (61,1 %)	42 / 88 (47,7 %)	0,65 [0,44; 0,98]	0,0383
IHC-HER2-Status[g]	3+	223 / 370 (60,3 %)	190 / 350 (54,3 %)	0,73 [0,60; 0,89]	0,0015
FISH-Status[h]	Positiv	225 / 381 (59,1 %)	212 / 384 (55,2 %)	0,77 [0,64; 0,93]	0,0060
ER / PgR-Status[i]	ER und / oder PgR Positiv	108 / 197 (54,8 %)	113 / 189 (59,8 %)	0,98 [0,76; 1,28]	0,9019

Parameter	Subgruppenmerkmal	Placebo+Trastuzumab +Docetaxel	Pertuzumab+Trastuzumab +Docetaxel	Ptz+T+D vs Pla+T+D	Ptz+T+D vs Pla + T +D
		Patienten mit Ereignis n/N (%)	Patienten mit Ereignis n/N (%)	Hazard Ratio [95% KI]	p-Wert
	ER und PgR Negativ	126 / 196 (64,3 %)	109 / 212 (51,4 %)	0,60 [0,46; 0,78]	0,0001
	Unbekannt	5 / 11 (45,5 %)	0 / 1 (0 %)	0,00 [0,00;.]	0,9971
IHC-HER2-Status[j]	2+	15 / 31 (48,4 %)	30 / 47 (63,8 %)	1,28 [0,69; 2,39]	0,4353
Baseline ECOG[k]	0	150 / 248 (60,5 %)	157 / 274 (57,3 %)	0,73 [0,58; 0,92]	0,0064
	1+	89 / 156 (57,1 %)	65 / 128 (50,8 %)	0,84 [0,61; 1,16]	0,2988
Vorherige (neoajduvante oder adjuvante) Behandlung mit Trastuzumab [l]	Ja	24 / 41 (58,5 %)	26 / 47 (55,3 %)	0,90 [0,51; 1,60]	0,7248
	Nein	215 / 363 (59,2 %)	196 / 355 (55,2 %)	0,76 [0,62; 0,92]	0,0053
Dosiseskalation mit Docetaxel auf 100 mg / m^2 [m]	Ja	34 / 63 (54,0 %)	26 / 46 (56,5 %)	0,97 [0,58; 1,62]	0,9053
	Nein	205 / 341 (60,1 %)	196 / 356 (55,1 %)	0,74 [0,61; 0,90]	0,0025
Vorherige (neo-) adjuvante Taxanbehandlung [n]	Ja	56 / 94 (59,6 %)	56 / 91 (61,5 %)	0,93 [0,64; 1,35]	0,7095
	Nein	183 / 310 (59,0 %)	166 / 311 (53,4 %)	0,74 [0,60; 0,91]	0,0044

Ereignis bezieht sich hier auf das Auftreten einer Verschlechterung
a-n zusätzliche Analysen [55]
Kursiv gezeigte Werte wurden nachträglich auf Basis der angegebenen Daten berechnet.

Es liegen keine Interaktionstests für die verschiedenen Subgruppen vor. Vor dem Hintergrund der bereits jetzt sehr großen Zahl von durchgeführten post-hoc Interaktionstests, wurde auf Interaktionstests bei Subscores, Domains und Einzelfragen mit den jeweiligen Subgruppen verzichtet.

4.3.1.3.2.4 <Sicherheit> - RCT

4.3.1.3.2.4.1 <Endpunkt Patienten mit unerwünschtem Ereignis (UE)> - RCT

Tabelle 4-66: Subgruppenergebnisse für<Endpunkt Patienten mit unerwünschtem Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

Parameter	Subgruppenmerkmal	Placebo+Trastuzumab +Docetaxel	Pertuzumab+Trastuzumab +Docetaxel	Ptz+T+D vs Pla+T+D	Ptz+T+D vs Pla+T+D	Inter-aktions-test
		Patienten mit Ereignis n/N (%)	Patienten mit Ereignis n /N (%)	Relatives Risiko [95% KI]	p-Wert	p-Wert
Altersgruppe[a]	<65	328 / 332 (98,8)	346 / 346 (100,0)	1,01 [1,00; 1,02]	0,0570*	0,9995
	≥65	63 / 64 (98,4)	62 / 62 (100,0)	1,02 [0,99; 1,05]	1,0000*	
Rasse[b]	Kaukasisch	223 / 226 (98,7)	250 / 250 (100,0)	1,01 [0,998; 1,03]	0,1063*	1,0000
	Afro-amerikanisch	20 / 20 (100,0)	10 / 10 (100,0)			
	Asiatisch	132 / 133 (99,2)	128 / 128 (100,0)	1,01 [0,99; 1,02]	1,0000*	
	Weitere	16 / 17 (94,1)	20 / 20 (100,0)	1,06 [0,94; 1,20]	0,4595*	

a: Tabelle aus CSR update 12/2012 S. 1769 und 1776; *SAP* Mai 2012 und zusätzliche Analysen [55]
b: Tabelle aus CSR update 12/2012 S. 1783, 1791, 1799 und 1807; *SAP* Mai 2012 und zusätzliche Analysen [55]
* Exakter p-Wert (CMH Chi^2 Test) da niedrige Ereignisrate

Weder aus medizinischer, noch aus statistischer Sicht finden sich Hinweise auf ein erhöhtes Risikoprofil von Pertuzumab.

4.3.1.3.2.4.2 <Endpunkt Patienten mit schwerwiegendem unerwünschten Ereignis (SUE)> - RCT

Tabelle 4-67: Subgruppenergebnisse für<Endpunkt Patienten mit schwerwiegendem unerwünschten Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

Parameter	**Subgruppenmerkmal**	**Placebo+Trastuzumab +Docetaxel**	**Pertuzumab+Trastuzumab +Docetaxel**	**Ptz+T+D vs Pla+T +D**	**Ptz+T+D vs Pla+T +D**	**Inter-aktionstest**
		Patienten mit Ereignis n/N(%)	**Patienten mit Ereignis n/N(%)**	**Relatives Risiko [95% KI]**	**p-Wert**	**p-Wert**
Altersgruppe[a]	<65	91 / 332 (27,4)	119 / 346 (34,4)	1,25 [1,00; 1,58]	0,0495	0,9344
	≥65	24 / 64 (37,5)	29 / 62 (46,8)	1,25 [0,83; 1,89]	0,2937	
Rasse[b]	Kaukasisch	56 / 226 (24,78)	75 / 250 (30,00)	1,21 [0,90; 1,63]	0,2033	0,0259
	Afro-amerikanisch	12 / 20 (60,00)	2 / 10 (20,00)	0,33 [0,09; 1,21]	0,0577	
	Asiatisch	40 / 133 (30,08)	61 / 128 (47,66)	1,58 [1,15; 2,17]	0,0036	
	Weitere	7 / 17 (41,18)	10 / 20 (50,00)	1,21 [0,59; 2,49]	0,5965	

a: Tabelle aus CSR update 12/2012 S. 1769 und 1776; *SAP* und zusätzliche Analysen [55]
b: Tabelle aus CSR update 12/2012 S. 1783, 1791, 1799 und 1807; *SAP* und zusätzliche Analysen [55]

Für diesen Endpunkt sind über alle Subgruppen hinweg keine Interaktionentestergebnisse vorhanden, die einen wirklichen Hinweis auf eine mögliche Interaktion geben könnten. Das Auftreten signifikanter Interaktionen kann mit hoher Wahrscheinlichkeit zufällig sein, da insgesamt weniger signifikante Interaktionen zu verzeichnen sind, als alleine aufgrund Zufall zu erwarten wären.

Zudem sind für die wenigen Interaktionstestergebnisse, die formal in eine der beiden IQWiG-Kategorien für Interaktionstests fallen, auf Einzelvariablenebene in Kapitel 4.3.1.3.2 genannte Kriterien nicht erfüllt.

Für diesen Endpunkt findet sich aus medizinischer Sicht kein Hinweis für ein Safety Signal aufgrund der zusätzlichen Gabe von Pertuzumab. Für das Subgruppenmerkmal „Asiatisch“ kann die schlechter vertragene Docetaxel-Therapie und damit die größere Anzahl an febrilen Neutropenien ursächlich für das vermehrte Auftreten von SUE sein.

4.3.1.3.2.4.3 <Endpunkt Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis (UE)> - RCT

Tabelle 4-68: Subgruppenergebnisse für<Endpunkt Patienten mit Behandlungsabbruch wegen unerwünschtem Ereignis> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

Parameter	**Subgruppenmerkmal**	**Placebo+Trastuzumab +Docetaxel**	**Pertuzumab+Trastuzumab +Docetaxel**	**Ptz+T+D vs Pla+T +D**	**Ptz+T+D vs Pla+T +D**	**Inter-aktionstest**
		Patienten mit Ereignis n/N (%)	**Patienten mit Ereignis n/N (%)**	**Relatives Risiko [95% KI]**	**p-Wert**	**p-Wert**
Altersgruppe[c]	<65	94 / 332 (28,3)	106 / 346 (30,6)	1,08 [0,86; 1,37]	0,5077	0,9490
	≥65	20 / 64 (31,3)	19 / 62 (30,7)	0,98 [0,58; 1,65]	0,9417	
Rasse[e]	Kaukasisch	42 / 226 (18,6)	63 / 250 (25,2)	1,36 [0,96; 1,92]	0,0825	0,2138
	Afro-amerikanisch	8 / 20 (40,0)	2 / 10 (20,0)	0,50 [0,13; 1,93]	0,4195*	
	Asiatisch	59 / 133 (44,4)	54 / 128 (42,2)	0,95 [0,72; 1,26]	0,7237	
	Weitere	5 / 17 (29,4)	6 / 20 (30,0)	1,02 [0,38; 2,76]	0,9693	

a: Tabelle aus CSR update 12/2012 S. 1769 und 1776; *SAP* und zusätzliche Analysen [55]
b: Tabelle aus CSR update 12/2012 S. 1783, 1791, 1799 und 1807; *SAP* und zusätzliche Analysen [55]
p-Wert wurde mittels CMH Chi^2 Test bestimmt
*Exakter p-Wert da niedrige Ereignisrate

Für Patienten mit Behandlungsabbruch wegen UE - sofern nicht Docetaxel bedingt - ergab sich für den Interaktionstest für die Analyse der Altersgruppen ein p = 0,1784 und für die Rasse p = 0,49.

Weder aus medizinischer, noch aus statistischer Sicht finden sich Hinweise auf ein erhöhtes Risikoprofil von Pertuzumab.

4.3.1.3.2.4.4 <Endpunkt Patienten mit UE ≥ Grad 3> - RCT

Tabelle 4-69: Subgruppenergebnisse für<Endpunkt Patienten mit UE ≥Grad 3> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

Parameter	**Subgruppenmerkmal**	**Placebo+Trastuzumab +Docetaxel**	**Pertuzumab+Trastuzumab +Docetaxel**	**Ptz+T+D vs Pla+T+D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T+D p-Wert**	**Inter-aktionstest p-Wert**
		Patienten mit Ereignis n/N (%)	**Patienten mit Ereignis n/N (%)**			
Altersgruppe[a]	<65	243 / 332 (73,2)	261 / 346 (75,4)	1,03 [0,94; 1,13]	0,5046	0,4948
	≥65	48 / 64 (75,0)	50 / 62 (80,6)	1,08 [0,89; 1,30]	0,4479	
Rasse[b]	Kaukasisch	167 / 226 (73,9)	183 / 250 (73,2)	0,99 [0,89; 1,10]	0,8641	0,3187
	Afro-amerikanisch	16 / 20 (80,0)	7 / 10 (70,0)	0,88 [0,55; 1,39]	0,6573	
	Asiatisch	98 / 133 (73,7)	107 / 128 (83,6)	1,13 [1,00; 1,29]	0,0517	
	Weitere	10 / 17 (58,8)	14 / 20 (70,0)	1,19 [0,73; 1,94]	0,4839	

a: Tabelle aus CSR update 12/2012 S. 1769 und 1776; *SAP* und zusätzliche Analysen [55]
b: Tabelle aus CSR update 12/2012 S. 1783, 1791, 1799 und 1807; *SAP* und zusätzliche Analysen [55]

Für diesen Endpunkt findet sich aus medizinischer Sicht kein Hinweis für ein Safety Signal aufgrund der zusätzlichen Gabe von Pertuzumab. Für das Subgruppenmerkmal „Asiatisch“ kann die schlechter vertragene Docetaxel-Therapie und damit die größere Anzahl an febrilen Neutropenien ursächlich für das vermehrte Auftreten von UE ≥ Grad 3 sein.

4.3.1.3.2.4.5 <Endpunkt Patienten mit UE Grad 3> - RCT

Tabelle 4-70: Subgruppenergebnisse für<Endpunkt Patienten mit UE Grad 3> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

Parameter	**Subgruppenmerkmal**	**Placebo+Trastuzumab +Docetaxel**	**Pertuzumab+Trastuzumab +Docetaxel**	**Ptz+T+D vs Pla+T+D**	**Ptz+T+D vs Pla+T+D**	**Inter-aktionstest**
		Patienten mit Ereignis n/N (%)	**Patienten mit Ereignis n/N (%)**	**Relatives Risiko [95% KI]**	**p-Wert**	**p-Wert**
Altersgruppe[a]	<65	188 / 332 (56,6)	213 / 346 (61,6)	1,09 [0,96; 1,23]	0,1917	0,9898
	≥65	40 / 64 (62,5)	41 / 62 (66,1)	1,06 [0,82; 1,37]	0,6721	
Rasse[b]	Kaukasisch	130 / 226 (57,5)	146 / 250 (58,4)	1,02 [0,87; 1,18]	0,8465	0,3210
	Afro-amerikanisch	15 / 20 (75,0)	7 / 10 (70,0)	0,93 [0,58; 1,51]	1,000*	
	Asiatisch	74 / 133 (55,6)	89 / 128 (69,5)	1,25 [1,03; 1,51]	0,0208	
	Weitere	9 / 17 (52,9)	12 / 20 (60,0)	1,13 [0,64; 2,01]	0,6701	

a: Tabelle aus zusätzliche Analysen [55] *SAP*
b: Tabelle aus zusätzliche Analysen [55] *SAP*
*Exakter p-Wert da niedrige Ereignisrate

Für diesen Endpunkt findet sich aus medizinischer Sicht kein Hinweis für ein Safety Signal aufgrund der zusätzlichen Gabe von Pertuzumab. Für das Subgruppenmerkmal „Asiatisch“ kann die schlechter vertragene Docetaxel-Therapie und damit die größere Anzahl an febrilen Neutropenien ursächlich für das vermehrte Auftreten von UE Grad 3 sein.

4.3.1.3.2.4.6 <Endpunkt Patienten mit UE Grad 4> - RCT

Tabelle 4-71: Subgruppenergebnisse für<Endpunkt Patienten mit UE Grad 4> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

Parameter	**Subgruppenmerkmal**	**Placebo+Trastuzumab +Docetaxel**	**Pertuzumab+Trastuzumab +Docetaxel**	**Ptz+T+D vs Pla+T+D**	**Ptz+T+D vs Pla+T+D**	**Inter-aktionstest**
		Patienten mit Ereignis n/N (%)	**Patienten mit Ereignis n/N (%)**	**Relatives Risiko [95% KI]**	**p-Wert**	**p-Wert**
Altersgruppe[a]	<65	137 / 332 (41,3)	141 / 346 (40,8)	0,99 [0,82; 1,18]	0,8920	0,4235
	≥65	22 / 64 (34,4)	24 / 62 (	1,13 [0,71; 1,79]	0,6148	
Rasse[b]	Kaukasisch	82 / 226 (36,3)	96 / 250 (38,4)	1,06 [0,84; 1,34]	0,6340	0,7570
	Afro-amerikanisch	12 / 20 (60,0)	4 / 10 (40,0)	0,67 [0,29; 1,54]	0,4421*	
	Asiatisch	60 / 133 (45,1)	59 / 128 (46,1)	1,02 [0,78; 1,33]	0,8739	
	Weitere	5 / 17 (29,4)	6 / 20 (30,0)	1,02 [0,38; 2,76]	0,9693	

a: Tabelle aus zusätzliche Analysen [55] *SAP*
b: Tabelle aus zusätzliche Analysen [55] *SAP*
*Exakter p-Wert da niedrige Ereignisrate

Weder aus medizinischer, noch aus statistischer Sicht finden sich Hinweise auf ein erhöhtes Risikoprofil von Pertuzumab.

4.3.1.3.2.4.7 <Endpunkt Patienten mit UE Grad 5> - RCT

Tabelle 4-72: Subgruppenergebnisse für<Endpunkt Patienten mit UE Grad 5> während der Behandlungszeit aus randomisierter, kontrollierter Studie CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel, stratifiziert nach vorheriger Behandlung und Region (2. Datenschnitt Mai 2012)

Parameter	**Subgruppenmerkmal**	**Placebo+Trastuzumab +Docetaxel**	**Pertuzumab+Trastuzumab +Docetaxel**	**Ptz+T+D vs Pla+T+D Relatives Risiko [95% KI]**	**Ptz+T+D vs Pla+T+D p-Wert**	**Inter-aktionstest p-Wert**
		Patienten mit Ereignis n/N (%)	**Patienten mit Ereignis n/N (%)**			
Altersgruppe[a]	<65	6 / 332 (1,81)	4 / 346 (1,16)	0,64 [0,18; 2,25]	0,5383*	0,8713
	≥65	6 / 64 (9,38)	4 / 62 (6,45)	0,69 [0,20; 2,32]	0,7439*	
Rasse[b]	Kaukasisch	9 / 226 (4,0)	6 / 250 (2,4)	0,60 [0,22, 1,67]	0,3243	0,7133
	Afro-amerikanisch	0 / 20 (0,0)	0 / 10 (0,0)	-	-	
	Asiatisch	2 / 133 (1,5)	2 / 128 (1,6)	1,04 [0,15; 7,27]	1,0000*	
	Weitere	1 / 17 (5,9)	0 / 20 (0,0)	-	0,4595*	

a: Tabelle aus zusätzliche Analysen [55] *SAP*
b: Tabelle aus zusätzliche Analysen [55] *SAP*
*Exakter p-Wert da niedrige Ereignisrate

Weder aus medizinischer, noch aus statistischer Sicht finden sich Hinweise auf ein erhöhtes Risikoprofil von Pertuzumab.

4.3.1.3.3 Zusammenfassung der Ergebnisse aus randomisierten kontrollierten Studien

Der vorliegende Abschnitt soll einen Überblick über die Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen aus randomisierten kontrollierten Studien geben. Die Zusammenfassung soll Aussagen zu allen in Abschnitt 4.3.1.3 präsentierten Endpunkten und Subgruppenanalysen enthalten. Dabei sollen, soweit verfügbar, numerische Ergebnisse aus Meta-Analysen einschließlich Konfidenzintervallen dargestellt werden.

Fassen Sie die Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen aus randomisierten kontrollierten Studien zusammen.

Mortalität

Die finale konfirmatorische Analyse **des Gesamtüberlebens (OS)** im Mai 2012 ergab ein HR von 0,66 (95 % KI [0,52; 0,84]) zugunsten von Ptz. Das mediane OS lag bei Pla+T+D bei 37,6 Monaten (154 Todesfälle; 37,93 %). Demgegenüber wurde die mediane Überlebenszeit unter Ptz+T+D (113 Fälle; 28,11 %) in der finalen Analyse nicht erreicht. Durch die Kombination aus Ptz+T+D wurde das Überleben im Vergleich zu Pla+T+D signifikant verlängert (p = 0,0008).

Morbidität

Die finale konfirmatorische Analyse des **progressionsfreien Überlebens (PFS bewertet durch IRF)** im Mai 2011 ergab ein HR von 0,62 (95 % KI [0,51; 0,75]), entsprechend einer signifikanten Verlängerung des PFS (Median) um 6,1 Monate unter Ptz+T+D (p < 0,0001). Insgesamt traten 433 PFS Ereignisse auf: 242 (59,6 %) unter Pla+T+D und 191 (47,5 %) unter Ptz+T+D.

Die **Gesamtansprechrate (bewertet durch das IRF)** wurde bei den Patienten der ITT-Population bewertet, die eine durch das IRF bestimmte, messbare Erkrankung zu Studienbeginn aufwiesen. Über 80 % der Patienten in jedem Arm waren in diese Analyse eingeschlossen (336 im Pla+T+D Behandlungsarm und 343 im Ptz+T+D Behandlungsarm). Die Gesamtansprechrate konnte von 69,3 % unter Pla+T+D auf 80,2 % unter Ptz+T+D signifikant gesteigert werden: RR = 1,16; 95 % KI [1,06; 1,26]; p = 0,0012.

Die Endpunkte IRF-bewertetes PFS, OS und Gesamtansprechrate stellen auch aus statistischer Sicht insofern wesentliche Endpunkte dar, als für diese drei Endpunkte im Protokoll eine Testhierarchie zur Adjustierung bzgl. multiplen Testens festgelegt worden war. Die statistische Auswertung zeigt nun für jeden dieser 3 wesentlichen Endpunkte eine signifikante Überlegenheit der Behandlung mit Ptz+T+D gegenüber Pla+T+D.

Lebensqualität

Die Zeit bis zur Verschlechterung des **FACT-B BCS-Score** war unter Ptz+T+D signifikant verlängert: HR = 0,77; 95 % KI [0,64; 0,93]. Die mediane Zeit bis zur Verschlechterung lag bei 18,3 Wochen unter Pla+T+D vs. 26,7 Wochen unter Ptz+T+D (p = 0,0058).

Hinsichtlich Symptomprogression – gemessen als Abnahme des **FACT-B TOI-PFB** von mindestens 5 Punkten – zeigte sich kein signifikanter Unterschied zwischen den Studienarmen.

Unerwünschte Ereignisse

Die Inzidenz **aller UE** lag bei Pla+T+D bei 98,74 % vs. 100,0 % der Patienten unter Ptz+T+D (RR: 1,01 [95% KI: 1,00; 1,02]). Das Ergebnis der Kategorie alle UE bedeutet, dass ein 1%-Unterschied zwischen den Gruppen besteht, der lediglich aufgrund der hohen Fallzahl signifikant wurde. Der Punktschätzer für das Relative Risiko liegt bei 1.012, die exakte untere Grenze des zweiseitigen 95%-Konfidenzintervalls liegt bei 1.0016, bereits 1 Patient mehr oder weniger in einem der Studienarme hätte dieses Ergebnis nicht signifikant werden lassen. Aus diesem Grund betrachtet Roche das Ergebnis als formal signifikant aber nicht medizinisch relevant.

Die Inzidenz der **SUE** lag bei Pla+T+D bei 29,0 % vs. 36,3 % der Patienten unter Ptz+T+D (RR 1,25; 95 % KI [1,02; 1,53]). Der Unterschied zwischen beiden Armen war hauptsächlich durch die höhere Inzidenz febriler Neutropenien unter Ptz+T+D bedingt, die eine Verlängerung des stationären Aufenthalts erforderten.

Die Inzidenz der **UE, die zum Behandlungsabbruch** führte, lag unter Pla+T+D bei 28,8 % (n = 114) gegenüber 30,6 % (n = 125) unter Ptz+T+D (RR 1,06; 95 % KI [0,86; 1,32]). Bei Ausschluss von Ereignissen, die zum Abbruch der Docetaxel-Therapie führten, lauten die Ergebnisse 6,1 % der Patienten unter Pla+T+D vs. 7,8 % unter Ptz+T+D.

Obwohl die meisten UE von Schweregrad 1 oder 2 waren, trat bei der Mehrzahl der Patientinnen mindestens ein **UE von Grad ≥ 3** auf (bei n = 291, 73,5 % der Patienten unter Pla+T+D und bei n = 311, 76,2 % der Patienten im Pertuzumab-Arm). Somit war die Inzidenz der UE von Grad ≥ 3 zwischen den Behandlungsarmen ausgewogen (RR: 1,04 [95 % KI: 0,96; 1,12].

Die Betrachtung der Inzidenzen der UE spezifisch in den einzelnen Schweregradgruppen (**Grad 3, Grad 4 und Grad 5**) zeigte keine statistisch signifikanten Unterschiede zwischen den Behandlungsarmen. Die Inzidenz der **UE vom Grad 5** lag unter Pla+T+D bei 3,03 % (n = 12) vs. 1,96 % (n = 8) unter Ptz+T+D (RR 0,65; 95 % KI [0,27; 1,57]).

Die Inzidenz der **Diarrhö** lag unter Pla+T+D bei 48,2 % vs. 68,1 % unter Ptz+T+D (RR 1,41; 95 % KI [1,25; 1,60]). Diarrhö vom Grad ≥ 3 trat bei 5,1 % (n = 20) der Patienten unter Pla+T+D und bei 9,1 % (n = 37) unter Ptz+T+D auf. Weder im Kontrollarm, noch im Pertuzumab-Therapiearm wurden fatale Verläufe beobachtet.

Die Inzidenz der **Exantheme** lag unter Pla+T+D bei 36,4 % vs. 47,5 % unter Ptz+T+D (RR 1,31; 95 % KI [1,11; 1,54]). Exantheme mit Schweregrad ≥ 3 waren sehr selten (n = 5 vs. n = 12). Das Exanthem wird derzeit nicht als bedeutsames identifiziertes Risiko klassifiziert und gilt nicht als medizinisch signifikant Exantheme lassen sich mit Hydrocortison, Bethamethason, Clindamycin oder Hydroxyzin sehr gut behandeln.

Hinsichtlich **febriler Neutropenie** lag die Inzidenz bei 7,6 % (n = 30) unter Pla+T+D vs. 13,8 % (n = 56) unter Ptz+T+D: RR 1,81; 95 % KI [1,19; 2,76]. **Febrile neutropenische Infektionen** traten unter Pla+T+D mit einer Häufigkeit von n = 3 und n = 14 unter Ptz+T+D auf (RR = 4,53; 95 % KI [1,31; 15,64]).

Die Häufigkeit einer **Mukositis** lag unter Pla+T+D bei 37,9 % vs. 49,8 % unter Ptz+T+D: RR 1,31; 95 % KI [1,12, 1,54]. Der Anteil mit Mukositis Grad ≥ 3 war gering (n=8 vs. n=13). Das Auftreten einer Mukositis unter einer Brustkrebstherapie mit Pertuzumab in Kombination mit Trastuzumab und Docetaxel beeinflusste die Therapie nur unwesentlich und ließ sich durch Standardmaßnahmen im Rahmen einer Chemotherapie sehr gut beherrschen.

Hinsichtlich der weiteren unerwünschten Ereignisse von besonderem Interesse (behandlungsbedingte Leberstörungen, Anaphylaxie und Hypersensibilitätsreaktionen, Leukopenie, interstitielle Lungenerkrankung, kardiale Dysfunktion / SAEs, die auf eine CHF hindeuten, QT Verlängerung, Linksventrikuläre Dysfunktion) ergab sich zwischen den Studienarmen kein statistisch signifikanter Unterschied.

Insgesamt traten nicht mehr und keine schwereren Nebenwirkungen auf, wenn Patienten mit HER2-positivem metastasierten oder lokal rezidivierenden inoperablen Brustkrebs mit Pertuzumab behandelt wurden. Das Sicherheitsprofil der beiden Therapiearme (Pla+T+D und Ptz+T+D) wurde durch Ereignisse dominiert, die typischerweise mit einer Behandlung mit Docetaxel assoziiert sind.

Durch das Hinzufügen von Pertuzumab zum im klinischen Alltag etablierten Trastuzumab in Kombination mit Docetaxel Therapieschemata veränderte sich das Sicherheitsprofil der Brustkrebstherapie kaum und ist somit in beiden Therapiearmen vergleichbar.

Subgruppen-Analysen

Insgesamt treten weniger signifikante Interaktionstests auf, als per Zufall erwartbar wären und die scheinbar auf eine Interaktion hindeuten. Für diese Ergebnisse, die formal in eine der beiden IQWiG-Kategorien für Interaktionstests fallen, sind auf Einzelvariablenebene in Kapitel 4.3.1.3.2 genannte Kriterien nicht erfüllt. Somit resultiert keine Differenzierung des Zusatznutzens über einzelne Subgruppen.

Tabelle 4-73: Zusammenfassung der Ergebnisse aus aktiv kontrollierter RCT mit dem zu bewertenden Arzneimittel aus CLEOPATRA (WO20698)

	Placebo+Trastuzumab +Docetaxel n / N (%)	**Pertuzumab+Trastuzumab +Docetaxel n / N (%)**	**Vergleich Effektschätzer [95% KI] p-Wert**
Mortalität			
Gesamtüberleben	154 / 406 (37,9 %)	113 / 402 (28,1 %)	HR:0,66 [0,52 ; 0,84] p = 0,0008
Morbidität			
Progressionsfreies Überleben bewertet durch IRF	242 / 406 (59,6 %)	191 / 402 (47,5 %)	HR: 0,62 [0,51 ; 0,75] p <0,0001
Gesamtansprechrate	233 / 336 (69,3 %)	275 / 343 (80,2 %)	RR: 1,16 [1,06 ; 1,26] p = 0,0012
Lebensqualität			
Anzahl Patienten mit Symptomprogression	229 / 404 (56,7 %)	239 / 402 (59,5 %)	HR: 0,97 [0,81; 1,16] p = 0,7182
Zeit bis zur Symptomprogression [Wochen] Median, 95%KI	18,3 [18; 27]	18,4 [18; 27]	p = 0,7161
Patienten mit Verschlechterung des BCS-Score	239 / 404 (59,2 %)	222 / 402 (55,2 %)	HR: 0,77 [0,64; 0,93] p = 0,0058
Zeit bis zur Verschlechterung des BCS-Score [Wochen] Median, 95%KI	18,3 [18; 27]	26,7 [18; 36]	p = 0,0061
Sicherheit			
Patienten mit unerwünschtem Ereignis	391 / 396 (98,7 %)	408 / 408 (100,0 %)	RR: 1,01 [1,00; 1,02] p = 0,0286
Patienten mit schwerwiegendem unerwünschtem Ereignis	115 / 396 (29,0 %)	148 / 408 (36,3 %)	RR: 1,25 [1,02; 1,53] p = 0,0289
Patienten mit Behandlungsabbruch wegen UE	114 / 396 (28,8 %)	125 / 408 (30,6 %)	RR: 1,06 [0,86; 1,32] p = 0,5665
Patienten mit UE ≥ Grad 3	291/396 (73,5 %)	311 / 408 (76,2 %)	RR: 1,04 [0,96; 1,12] p = 0,3707
Patienten mit UE Grad 3	228 / 396 (57,6 %)	254 / 408 (62,3 %)	RR: 1,08 [0,97; 1,21] p = 0,1761
Patienten mit UE Grad 4	159 / 396 (40,2 %)	165/408 (40,44 %)	RR: 1,01 [0,85; 1,19] p = 0,9333

	Placebo+Trastuzumab +Docetaxel n / N (%)	**Pertuzumab+Trastuzumab +Docetaxel n / N (%)**	**Vergleich Effektschätzer [95% KI] p-Wert**
Patienten mit UE Grad 5	12 / 396 (3,0 %)	8 / 408 (1,96 %)	RR: 0,65 [0,27; 1,57] p = 0,3306
UE von besonderem Interesse			
Diarrhö	191 / 396 (48,23 %)	278 / 408 (68,14 %)	1,41 [1,25; 1,60] p < 0,0001
Exanthem (*Rash*)	144 / 396 (36,36 %)	194 / 408 (47,55 %)	1,31 [1,11; 1,54] p = 0,0013
Leukopenie	231 /396 (58,33 %)	255 / 408 (62,5 %)	1,07 [0,96;1,20] p = 0,2271
Febrile Neutropenie	30/396 (7,58 %)	56/408 (13,73 %)	1,81 [1,19; 2,76] p = 0,0048
Febrile neutropenische Infektionen	3/396 (0,76 %)	14/408 (3,43 %)	4,53 [1,31; 15,64] p = 0,0120
Interstitielle Lungenerkrankung	6/396 (1,52 %)	10/408 (2,45 %)	1,62 [0,59; 4,41] p = 0,3425
Anaphylaxie und Hypersensibilitätsreaktionen	36/396 (9,09 %)	45/408 (11,03 %)	1,21 [0,80; 1,84] p = 0,3616
Behandlungsbedingte Leberstörungen	43/396 (10,86 %)	42/408 (10,29 %)	0,95 [0,63; 1,42] p = 0,7948
Mukositis	150/396 (37,88 %)	203/408 (49,75 %)	1,31 [1,12; 1,54] p = 0,0007
Kardiale Dysfunktion / SUE, die auf CHF hindeuten	8/396 (2,02 %)	6/408 (1,47 %)	0,73 [0,25; 2,08] p = 0,5517
QT Verlängerung	5/396 (1,26 %)	9/408 (2,21 %)	1,75 [0,59; 5,17] p = 0,3070
Patienten mit UE Linksventrikuläre Dysfunktion	34/396 (8,6 %)	22/408 (5,4 %)	RR: 0,63 [0,37; 1,05] p = 0,0755
Patienten mit UE symptomatische LVSD, festgestellt durch CRC	4/396 (1,01 %)	4/408 (0,98 %)	0,97 [0,24; 3,85] p = 1,0000
Patienten mit UE symptomatische LVSD festgestellt durch *Investigator*[c]	7/396 (1,77 %)	5/408 (1,23 %)	0,69 [0,22; 2,17] p = 0,5264
Patienten mit UE asymptomatische LVSD	27/396 (6,82 %)	17/408 (4,17 %)	0,61 [0,34; 1,10] p = 0,0986

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 Vergleichstherapievorliegen oder diese keine ausreichenden Aussagen über den Zusatznutzen zulassen.

4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche

Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung zu Studien für indirekte Vergleiche. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen analog Abschnitt 4.3.1.1 zur Verfügung. Benennen Sie

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Literaturrecherche*
- *Studien aus der Suche in Studienregistern*
- *Resultierender Studienpool aus den einzelnen Suchschritten*

<< Nicht zutreffend >>

4.3.2.1.2 Charakteristika der Studien für indirekte Vergleiche

Charakterisieren Sie nachfolgend die Studien, die für indirekte Vergleiche herangezogen wurden, und bewerten Sie deren Verzerrungspotenzial. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.

<< 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.

Tabelle 4-74: 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 3 Schritten erfolgen: 1) Bewertung des Verzerrungspotenzials auf Endpunktebene pro Studie, 2) tabellarische Darstellung der Ergebnisse der einzelnen Studien, 3) Darstellung des indirekten Vergleichs. Für die Punkte 1 und 2 gelten die gleichen Anforderungen wie für die Darstellung der Ergebnisse der direkten Vergleiche in Abschnitt 4.3.1.3.1.

Geben Sie für den im vorliegenden Abschnitt präsentierten Endpunkt einen Überblick über die in den Studien verfügbaren Vergleiche. Beispielhaft wäre folgende Darstellung denkbar:

Tabelle 4-75: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden

Anzahl Studien	**Referenzen Studien**	**Intervention**	**<Vergleichs-therapie 1>**	**<Vergleichs-therapie 2>**	**<Vergleichs-therapie 3>**
1	<Studie 1>	●		●	●
2	<Studie 2>	●		●	
	<Studie 3>	●		●	
1	<Studie 4>		●	●	●
etc.	etc.	etc.	etc.		

Stellen Sie zusätzlich die Netzwerkstruktur des indirekten Vergleichs grafisch dar.

<< Nicht zutreffend >>

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

Tabelle 4-76: Operationalisierung von <Endpunkt xxx>

Studie	**Operationalisierung**
<Studie 1>	

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

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

Tabelle 4-77: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche

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

Begründen Sie für jede Studie die abschließende Einschätzung.

<< 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.

Tabelle 4-78: 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)**
<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 die übliche Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: Diskutieren Sie das Ausmaß sowie die Gründe für das Auftreten der Heterogenität für alle direkten paarweisen Vergleiche.*
- *Ergebnisse zu den Effekten: Stellen Sie die gepoolten Ergebnisse dar.*
- *Konsistenzprüfung: Stellen Sie die Ergebnisse der Konsistenzprüfung dar. Diskutieren Sie insbesondere Widersprüche zwischen direkter und indirekter Evidenz.*

<< 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 0.

<< 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

Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung zu nicht randomisierten vergleichenden Studien. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen analog Abschnitt 4.3.1.1 zur Verfügung. Benennen Sie

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Literaturrecherche*
- *Studien aus der Suche in Studienregistern*
- *Resultierender Studienpool aus den einzelnen Suchschritten*

<< 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.

Beschreiben Sie die Verzerrungsaspekte der nicht randomisierten vergleichenden Studie auf Studienebene mithilfe des Bewertungsbogens in Anhang 4-G. Fassen Sie die Beschreibung mit den Angaben in der folgenden Tabelle zusammen. Fügen Sie für jede Studie eine neue Zeile ein.

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

Tabelle 4-79: Verzerrungsaspekte auf Studienebene – nichtrandomisierte vergleichende Interventionsstudien

Studie	Zeitliche Parallelität der Gruppen	Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren	Verblindung		Hinweise auf ergebnisgesteuerte Berichterstattung	Andere Aspekte, die Verzerrungen verursachen können
			Patient	Behandler		
<Studie 1>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein / unklar>	<ja / nein>

4.3.2.2.3 Ergebnisse aus nicht randomisierten vergleichenden Studien

4.3.2.2.3.1 <Endpunkt xxx> – nicht randomisierte vergleichende Studien

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

Tabelle 4-80: Operationalisierung von <Endpunkt xxx>

Studie	Operationalisierung
<Studie 1>	

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

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

Tabelle 4-81: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien

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

1) Stellen Sie die Ergebnisse der nichtrandomisierten vergleichenden Studien gemäß den Anforderungen des TREND- bzw. des STROBE-Statements dar.

<< 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 nichtrandomisierten vergleichenden Studien. Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 0.

<< 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

Beschreiben Sie nachfolgend das Ergebnis der Informationsbeschaffung nach Untersuchungen, die nicht in den Abschnitten 4.3.1, 4.3.2.1 und 4.3.2.2 aufgeführt sind. Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.1 (Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel) und stellen Sie Informationen analog Abschnitt 4.3.1.1 zur Verfügung. Benennen Sie

- *Studien des pharmazeutischen Unternehmers*
- *Studien aus der bibliografischen Literaturrecherche*
- *Studien aus der Suche in Studienregistern*
- *Resultierender Studienpool aus den einzelnen Suchschritten*

<< Nicht zutreffend >>

4.3.2.3.2 Charakteristika der weiteren Untersuchungen

Charakterisieren Sie nachfolgend die weiteren Untersuchungen und bewerten Sie deren Verzerrungspotenzial.

Ergebnisse nichtrandomisierter Studien, die keine kontrollierten Interventionsstudien sind, gelten aufgrund ihres Studiendesigns generell als potenziell hoch verzerrt. Trifft das auf die von Ihnen vorgelegten Studien nicht zu, begründen Sie Ihre Einschätzung.

Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2. und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.

<< Nicht zutreffend >>

4.3.2.3.3 Ergebnisse aus weiteren Untersuchungen

4.3.2.3.3.1 <Endpunkt xxx> – weitere Untersuchungen

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

Tabelle 4-82: Operationalisierung von <Endpunkt xxx>

Studie	Operationalisierung
<Studie 1>	

Bewerten Sie das Verzerrungspotenzial für den in diesem Abschnitt beschriebenen Endpunkt. Ergebnisse nichtrandomisierter Studien, die keine kontrollierten Interventionsstudien sind, gelten aufgrund ihres Studiendesigns generell als potenziell hoch verzerrt. Trifft das auf die von Ihnen vorgelegten Studien nicht zu, begründen Sie Ihre Einschätzung.

<< 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.

<< 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 0.

<< Nicht zutreffend >>

4.3.2.4 Zusammenfassung der Ergebnisse aus weiteren Unterlagen

Der vorliegende Abschnitt soll einen Überblick über die Ergebnisse aus weiteren Unterlagen (Abschnitte 4.3.2.1, 4.3.2.2 und 4.3.2.3) geben. Die Zusammenfassung soll Aussagen zu allen in diesen Abschnitten präsentierten Endpunkten und Subgruppenanalysen enthalten. Dabei sollen, soweit verfügbar, numerische Ergebnisse aus Meta-Analysen einschließlich Konfidenzintervallen dargestellt werden.

Fassen Sie die Ergebnisse aus weiteren Unterlagen zusammen.

<< 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.

Die CLEOPATRA-Studie ist eine randomisierte, kontrollierte, internationale, multizentrische klinische Studie der Phase III. Die Patienten erhielten alle den therapeutischen Standard Trastuzumab+Taxan (Docetaxel) als Basis. Patienten und Ärzte waren in dieser Studie verblindet.

In der CLEOPATRA-Studie wurden mittels eines IVRS die Screening-Informationen erfasst und die zur Randomisierung geeigneten Patienten wurden im Verhältnis 1:1 auf einen der beiden Behandlungsarme verteilt. Eine Stratifizierung erfolgte nach vorherigem Behandlungsstatus (de novo vs. vorherige adjuvante oder neoadjuvante Therapie) und Region (Europa, Nordamerika, Süd-Amerika und Asien). Das ITT-Prinzip wurde adäquat umgesetzt, da alle randomisierten Patienten in die Auswertung eingingen. Für die Sicherheitsauswertung wurden alle Patienten entsprechend der tatsächlich eingenommenen Medikation analysiert. Hinweise auf eine ergebnisgesteuerte Berichterstattung ergaben sich beim Abgleich mit der Publikation der Studie bzgl. der dort berichteten Punkte nicht. [49] Die Studie war verblindet - die Verblindung in den Studienperioden 1 (Februar 2008 bis Mai 2011) und 2 (Juni 2011 bis Mai 2012) durfte nur bei Sicherheitsbedenken, die während der Behandlung mit der Studienmedikamentation auftraten, aufgehoben werden. Auch sonstige, das Verzerrungspotential beeinflussende Punkte wurden nicht beobachtet. Das Verzerrungspotential wurde daher sowohl auf Studienebene wie auch für sämtliche Endpunkte mit „niedrig“ bewertet.

Die gewählten Endpunkte sind als valide und patientenrelevant einzustufen. Klinische Endpunkte wurden durch externe Reviewer bzw. Boards zusätzlich bewertet, die Lebensqualität wurde mittels eines validierten Fragebogens erfasst.

Insgesamt ist daher die Aussagekraft der Nachweise als „hoch“ einzustufen, da es sich bei der CLEOPATRA-Studie um den Evidenzgrad Stufe 1b handelt und ein niedriges Verzerrungspotential vorliegt.

Beleglage bei einer pivotalen Studie

Im Rahmen einer Nutzenbewertung ist in der Regel die Mindestanforderung für Aussagen eines Belegs, dass zwei voneinander unabhängig durchgeführte Studien mit endpunktbezogener hoher Ergebnissicherheit und entsprechendem statistisch signifikanten Effekt vorliegen. Dabei ist die Anforderung hinsichtlich der Studienzahl nicht absolut formuliert. Analog dem Vorgehen der EMA besteht auch bei der frühen Nutzenbewertung die Möglichkeit, auf Basis von einer pivotalen Studie, eine hinreichende Ergebnissicherheit zu erreichen, was einem Beleg entspricht. In diesem Fall ist den regulatorischen Anforderungen

an eine pivotale Studie zu genügen. [1] Die Anforderungen an eine solche Studie basieren sowohl bei IQWiG wie auch bei der EMA auf der „CPMP. Points to consider on application with: 1. meta-analyses; 2. one pivotal study". [12]

Die einzelnen Kriterien aus der EMA-Guidance werden im Folgenden für die CLEOPATRA-Studie überprüft. Da die Gewichtung der Kriterien seitens des IQWiG nicht bekannt ist, wurde im Folgenden von einer Gleichgewichtung ausgegangen.

1. Interne Validität (Es sollte keine Anzeichen für mögliche Verzerrung geben)

Das Verzerrungspotenzial auf Studienebene ist niedrig, da es sich um eine adäquat geplante, durchgeführte und ausgewertete doppelblinde RCT handelt (s. hierzu auch die Kapitel im Dossier zu Verzerrungspotential von Studie und Endpunkten). Darüber hinaus weist die Studie in den Dimensionen ‚Mortalität', ‚Morbidität' und ‚Lebensqualität' patientenrelevante Endpunkte (z.B. „Gesamtüberleben", „Progressionsfreies Überleben", „Ansprechrate", „Zeit bis zur Verschlechterung des BCS-Score") auf, deren endpunktübergreifendes und endpunktspezifisches Verzerrungspotential niedrig ist.

2. Externe Validität (Die Studienpopulation soll für eine Übertragung auf die zu behandelnde Population geeignet sein)

Für die Bestimmung der externen Validität existiert kein ähnlich gut ausgearbeitetes Instrumentarium wie für die Bestimmung der internen Validität. [72] Insgesamt ist zu prüfen, ob die Therapieeffekte in der Zielpopulation ähnlich jenen in der Studienpopulation sind. Da es sich damit v.a. um den Bereich des fachlichen Ermessens handelt wird die Frage wie folgt operationalisiert:

(i) Finden sich Einflussfaktoren für eine Effektmodifikation, wobei die Studienpatienten, als ein zeitlich versetzter aber inhaltlich repräsentativer Ausschnitt aus der Zielpopulation gelten können

(ii) Sollte (i) nicht gegeben sein, finden sich Anhaltspunkte, die eine Effektmodifikation durch bestimmte Einflussfaktoren wahrscheinlich bzw. unwahrscheinlich machen

Ad (i)

- Geschlecht: Das Geschlecht ist sowohl bei den Studienpatienten, wie auch bei der Zielpopulation fast ausschließlich auf Patientinnen beschränkt, hier ist demzufolge keine Effektmodifikation zu erwarten.

- Alter: Eine Effektmodifikation über den Einflussfaktor ‚Alter' ist denkbar, da die Studienpatienten jünger sind, als die Zielpopulation (Mittel: 53.5 vs. 65). [38] Allerdings ist durch die inzwischen in allen Bundesländern umgesetzten Brustkrebs-Screeningprogramme für die Zielpopulation eine Senkung des mittleren Alters zu erwarten, so wird bereits für 2012 ein mittleres Alter von 62 [73] berichtet, das noch weiter sinken wird. Damit lässt auch dieser

Einflussfaktor für die Zielpopulation keine, zumindest mittelfristig keine Effektmodifikation erwarten.

Ad (ii)

- Trastuzumab-Vorbehandlung: Der Anteil der Patientinnen mit Trastuzumab-vorbehandlung ist in der Zielpopulation deutlich größer als bei den Studienpatienten. Allerdings sind die Effekte (z.B. bezogen auf OS) in beiden Subgruppen fast identisch, so dass keine Effektmodifikation in der Zielpopulation zu erwarten ist.

- Vorheriger Behandlungsstatus (adjuvant bzw. neo-adjuvant versus de novo): Die Effekte sind in beiden Subgruppen bei den Studienpatienten identisch, eine Verschiebung der Häufigkeiten der Subgruppen in der Zielpopulation lässt damit keine Effektmodifikation in der Zielpopulation erwarten.

Fazit: Die Prüfung von Einflussfaktoren für eine Effektmodifikation gibt im Rahmen des fachlichen Ermessens keinen Anlass an der externen Validität der Studie zu zweifeln.

3. Statistische Signifikanz (Normalerweise ist ein Grad der statistischen Signifikanz besser als das 5%-Niveau nötig)

Die relevanten Wirksamkeitseffekte sind für die betrachteten, unverzerrten Endpunkte allesamt statistisch signifikant und die Effektschätzer mit der Nullhypothese unverträglich. Das zweiseitige Signifikanzniveau von 5 % wird für „Gesamtüberleben“, „Progressionsfreies Überleben“ „Ansprechrate“, „Zeit bis zur Verschlechterung des BCS-Score“) deutlich unterschritten.

Ohne diese hoch signifikanten Ergebnisse wäre eine vorzeitige Beendigung der Studie (wegen außerordentlicher Wirksamkeit) aufgrund der anspruchsvollen statistischen Stopping boundary auch nicht möglich gewesen.

4. Klinische Relevanz (Der Effekt muss groß genug sein, um klinisch relevant zu sein)

Die relevanten Wirksamkeitseffekte sind für die betrachteten, unverzerrten Endpunkt allesamt klinisch relevant (siehe 4.3.1.3.1. & 4.4.2). Die Effektschätzer sind allesamt weit von der Nullhypothese entfernt und stellen grosso modo 20 % - 40 %ige Verbesserungen gegenüber dem Vergleichsarm dar.

Aufgrund dieser überzeugenden Ergebnisse wurde die Studie wegen außerordentlicher Wirksamkeit (1. Datenschnitt: PFS, 2. Datenschnitt: OS) vorzeitig beendet und allen Patienten nach dem zweiten Datenschnitt das „Cross-over“ ermöglicht.

5. Datenqualität

Im CSR verweist Roche zur Sicherstellung der Datenqualität auf eine für die Wahrung der Datengüte in klinischen Studien verfassten SOP (Standard Operating Procedure; gcp-sop003292). Diese SOP nimmt in Kapitel 5.1. wiederum Bezug auf die *ICH E-6-Guidelines for Good Clinical Practice*, die zur Sicherstellung der Datenqualität in klinischen Studien SOPs empfiehlt.

Die erfolgte Zulassung seitens FDA und EMA ist eine Bestätigung für die hohen Standards, die Roche bei der Datenqualität seiner Studien anlegt.

6. Interne Konsistenz (Ähnliche Effekte zeigen sich in verschiedenen vordefinierten Subgruppen, alle relevanten Endpunkte haben ähnliche Ergebnisse)

Alle relevanten Endpunkte in den Dimensionen ‚Mortalität', ‚Morbidität' und ‚Lebensqualität' weisen ähnliche Ergebnisse auf, die auf eine medizinisch relevante Verbesserung durch die neue Behandlung hinweisen (Siehe Kapitel 4.3.1.3.1 und 4.4.2).

Die Subgruppenergebnisse sind recht homogen. Lediglich 3 Tests weisen ein signifikantes Ergebnis unterhalb des Signifikanzniveaus von 5% auf. Für diese Subgruppeneffekte findet sich keine biologische Rationale, gleichzeitig basieren einige auf einer sehr geringen Datenbasis. Ein weiterer Test liegt unterhalb des 10%igen Signifikanzniveaus. Diese 4 Tests liegen numerisch weit unterhalb dessen, was zufällig an signifikanten Interaktionstests zu erwarten gewesen wäre und sind damit sehr gut mit der Hypothese ‚Zufallsbefunde' vereinbar. Dies gilt auch für die weiteren 7 Interaktionstests oberhalb des 10% Signifikanzniveaus.

7. Effekte der Studienzentren („Keines der Studienzentren soll die Gesamtergebnisse dominieren, weder hinsichtlich der Anzahl der Patienten noch hinsichtlich der Effektgröße")

Liegen in einer Studie viele kleine Zentren vor, dann sollten diese für die Analyse gepoolt werden, um die Varianz nicht übermäßig zu vergößern. In dieser Studie wurden deshalb alle Zentren zu vordefinierten übergeordneten regionalen Einheiten gepoolt: Europa, Nordamerika, Südamerika und Asien.

Diese Regionen waren Bestandteil der statistischen Analysemodelle, so dass für mögliche Regionseffekte korrigiert wurde. Bei den Interaktionstests ergab sich kein qualitativer oder quantitativ erheblicher Einfluss auf die Effekte.

8. Hypothesen-Plausibilität („Die getestete Hypothese ist plausibel“)

CLEOPATRA war eine aktiv kontrollierte Studie mit einer Überlegenheitshypothese zum ‚Standard of care‘.

Da in beiden Behandlungsarmen der Standard of Care eingesetzt wird, ist die medizinische Grundlage der Hypothese plausibel. Die Umsetzung in einer Überlegenheitsstudie gegenüber der Standardbehandlung ist hinsichtlich der Wirksamkeit des neuen Produktes maximal fordernd. Damit ist auch die statistische Umsetzung der medizinischen Hypothese korrekt und plausibel.

4.4.2 Fazit

Die vorliegende CLEOPATRA-Studie erfüllt die in der „CPMP: Points to consider on application with: 1. meta-analyses; 2. one pivotal study” genannten Kriterien für Einreichungen mit einer pivotalen Studie. Sie ist damit hinreichend für die Kategorie Beleg bei der Ergebnissicherheit.

Zudem ist darauf hinzuweisen, dass es sich bei Pertuzumab um eine Substanz handelt, bei der ein hohes Maß an Überlebenszeitverlängerung nachgewiesen ist. Eine Wiederholung der Studie wäre ethisch somit nicht durchführbar. Dieser Sachverhalt ist ein weiterer Beweis, dass die Studie als Beleg bewertet werden muss. Sonst wären gerade Substanzen mit einer hohen Wirksamkeit - und einer daraus folgenden Änderung des Therapiestandards - hinsichtlich der Beleglage gemäß dem IQWiG Methodenpapier benachteiligt.

4.4.3 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. Liegt ein Zusatznutzen vor, beschreiben Sie, worin der Zusatznutzen besteht.

Stellen Sie die Wahrscheinlichkeit des Zusatznutzens dar, d. h., beschreiben und begründen Sie unter Berücksichtigung der in Abschnitt 4.4.1 dargelegten Aussagekraft der Nachweise die Ergebnissicherheit der Aussage zum Zusatznutzen.

Beschreiben Sie außerdem das Ausmaß des Zusatznutzens unter Verwendung folgender Kategorisierung (in der Definition gemäß AM-NutzenV):

- *erheblicher Zusatznutzen*

- *beträchtlicher Zusatznutzen*
- *geringer Zusatznutzen*
- *nicht quantifizierbarer Zusatznutzen*
- *kein Zusatznutzen belegbar*
- *der Nutzen des zu bewertenden Arzneimittels ist geringer als der Nutzen der zweckmäßigen Vergleichstherapie*

Berücksichtigen Sie bei den Aussagen zum Zusatznutzen ggf. nachgewiesene Unterschiede zwischen verschiedenen Patientengruppen.

Die folgende Tabelle beinhaltet nur die Ergebnisse der Endpunkte, die hinsichtlich der Bewertung des Zusatznutzens heranzuziehen sind.

Tabelle 4-83: Ausmaß des Zusatznutzens auf Endpunktebene aus der RCT CLEOPATRA (WO20698) mit dem zu bewertenden Arzneimittel auf Ebene der Studienpopulation

	Placebo+Trastuzumab+Docetaxel vs. Pertuzumab+Trastuzumab+Docetaxel **Effektschätzer [95% KI]** **p-Wert** **Wahrscheinlichkeit der Aussage**	**Ableitung des Ausmaßes**
Mortalität		
Gesamtüberleben [a]	HR: 0,66 [0,52; 0,84] p = 0,0008 Wahrscheinlichkeit: Beleg	Endpunktkategorie Überlebenszeit (Mortalität): Zusatznutzen, Ausmaß erheblich
Morbidität		
Progressionsfreies Überleben bewertet durch IRF [b]	HR: 0,62 [0,51; 0,75] p < 0,0001 Wahrscheinlichkeit: Beleg	Endpunktkategorie: schwerwiegende / schwere Symptome / Folgekomplikationen Zusatznutzen, Ausmaß erheblich
Gesamtansprechrate	RR: 1,16 [1,06; 1,26] RR[c]: 0,86 [0,79; 0,94] p = 0,0012 Wahrscheinlichkeit: Beleg	Endpunktkategorie: schwerwiegende / schwere Symptome / Folgekomplikationen Zusatznutzen, Ausmaß beträchtlich
Lebensqualität		
Zeit bis zur Verschlechterung des BCS-Score	HR: 0,77 [0,64; 0,93] p = 0,0058 Wahrscheinlichkeit: Beleg	Endpunktkategorie: Lebensqualität Zusatznutzen / bedeutsame Verbesserung Ausmaß beträchtlich

a: Todesfälle
b: Fälle mit Progression
c: Berechnung mit umgedrehter Effektrichtung, um Anwendung der Schwellenwerte für das Ausmaß des Zusatznutzens zu ermöglichen: Verhältnisse wurden auf das „Gegenereignis“ umgerechnet.

Nicht relevante und / oder statistisch nicht signifikant unterschiedliche Ergebnisse sind in obiger Tabelle nicht dargestellt. Dies betrifft folgende Endpunkte:

- Zeit bis Symptomprogression
- Patienten mit UE
- Patienten mit SUE
- Patienten mit Behandlungsabbruch wegen UE
- Patienten mit UE ≥ Grad 3
- Patienten mit UE Grad 3
- Patienten mit UE Grad 4
- Patienten mit UE Grad 5
- Diarrhö
- Exanthem
- Leukopenie
- Febrile Neutropenie
- Febrile neutropenische Infektionen
- Interstitielle Lungenerkrankung
- Anaphylaxie und Hypersensibilitätsreaktionen
- Mukositis
- Behandlungsbedingte Leberstörungen
- Kardiale Dysfunktion / SUE, die auf eine CHF hindeuten
- QT Verlängerung
- LVSD

Mortalität:

Die finale konfirmatorische Analyse **des Gesamtüberlebens (OS)** im Mai 2012 ergab ein HR von 0,66 (95 % KI [0,52; 0,84]) zugunsten von Ptz. Das mediane OS lag bei Pla+T+D bei 37,6 Monaten (154 Todesfälle; 37,93 %). Demgegenüber wurde die mediane Überlebenszeit unter Ptz+T+D (113 Fälle; 28,11 %) in der finalen Analyse nicht erreicht. Durch die Kombination aus Ptz+T+D wurde das Überleben im Vergleich zu Pla+T+D signifikant verlängert (p = 0,0008). Mit der heutigen Standardtherapie Trastuzumab in Kombination mit Docetaxel wurde ein medianes Gesamtüberleben von 37,6 Monaten erreicht. Dieses liegt damit bereits höher als die in der Zulassungsstudie von Trastuzumab und Docetaxel erreichten

31,2 Monate. [69] Die Definition gemäß AM-NutzenV fordert für einen erheblichen Zusatznutzen eine nachhaltige und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte große Verbesserung des therapierelevanten Nutzens im Sinne einer erheblichen Verlängerung der Überlebensdauer. Der zu erwartende zusätzliche mediane Überlebensvorteil mit Ptz+T+D liegt zwischen 8 und 19,4 Monaten. Damit liegt bereits die erwartete Untergrenze des Überlebensvorteils in der Größenordnung, die durch die Gabe von T+D erzielt wurde. [69] Diese Ergebnisse stellen einen erneuten Therapiedurchbruch bei der Behandlung von HER2-positiven Patienten dar. Somit wird eine erhebliche und bisher nicht erreichte therapierelevante Verlängerung der Überlebensdauer erzielt. Dies entspricht einem **erheblichen Zusatznutzen beim Gesamtüberleben.**

Morbidität:

Für das **progressionsfreie Überleben (PFS bewertet durch IRF)** ergab sich ein HR von 0,62 (95 % KI [0,51; 0,75]), entsprechend einer signifikanten Verlängerung des PFS (Median) um 6,1 Monate unter Ptz+T+D ($p < 0,0001$). Insgesamt traten 433 PFS Ereignisse auf: 242 (59,6 %) unter Pla+T+D und 191 (47,5 %) unter Ptz+T+D. Die Definition gemäß AM-NutzenV fordert für einen erheblichen Zusatznutzen eine nachhaltige und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte große Verbesserung des therapierelevanten Nutzens im Sinne einer weitgehenden Vermeidung von Symptomen und Folgekomplikationen. Ein progressionsfreies Überleben von 18,5 Monaten durch die zusätzliche Gabe von Pertuzumab zu der Standardtherapie wurde bisher für die behandelte Patientenpopulation noch nie erreicht und stellt einen erheblichen Zusatznutzen da. **Damit liegt ein erheblicher Zusatznutzen in der Dimension Morbidität vor.**

Die **Gesamtansprechrate** unter der Standardtherapie aus Docetaxel und Trastuzumab lag bei 69,3 % und entsprach in dieser Höhe den Ansprechraten aus publizierten randomisierten Studien. [69;74;75] Mit der zusätzlichen Gabe von Pertuzumab konnte die objektive Ansprechrate signifikant auf 80,2 % gesteigert werden. Der Unterschied von 10,8 % war signifikant. Die bereits sehr hohe Gesamtansprechrate der Standardtherapie Trastuzumab in Kombination mit Docetaxel konnte damit weiter beträchtlich verbessert werden. Die Definition gemäß AM-NutzenV fordert für einen beträchtlichen Zusatznutzen eine gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte deutliche Verbesserung des therapierelevanten Nutzens im Sinne einer Abschwächung von Symptomen und Folgekomplikationen. **Es liegt ein beträchtlicher Zusatznutzen für die Gesamtansprechrate vor.**

Lebensqualität:

Es wurde eine Analyse der **Zeit bis zur Verschlechterung der im BCS-Score erfassten Brustkrebs-spezifischen Symptome** durchgeführt. Diese zeigte, dass die Kombination aus Pertuzumab, Trastuzumab und Docetaxel in einem verzögerten Auftreten spezifischer Brustkrebs-assoziierter Symptome resultierte. Die Zeit bis zur Verschlechterung des FACT-B BCS-Score war unter Ptz+T+D signifikant verlängert: HR = 0,77; 95 % KI [0,64; 0,93]. Die mediane Zeit bis zur Verschlechterung lag bei 18,3 Wochen unter Pla+T+D vs. 26,7 Wochen

unter Ptz+T+D (p = 0,0058). Die Verbesserung war für alle erfassten Brustkrebssymptome nachzuweisen. Die Definition gemäß AM-NutzenV fordert für einen beträchtlichen Zusatznutzen eine gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte deutliche Verbesserung des therapierelevanten Nutzens im Sinne einer bedeutsamen Verbesserung der Lebensqualität. Durch die zusätzliche Gabe von Pertuzumab zu Trastuzumab und Docetaxel wird eine beträchtliche Verbesserung der Lebensqualität in Bezug auf Brustkrebs-spezifische Symptome erreicht. **Damit liegt ein beträchtlicher Zusatznutzen in der Dimension Lebensqualität vor.**

Unerwünschte Ereignisse:

Eine ausschließlich auf der statistischen Analyse einzelner UE-Kategorien beruhende Beurteilung der UE führt aus medizinischer Sicht zu einem unvollständigen, eventuell sogar verzerrten Bild. Von daher erfolgt die Bewertung der UE aus statistischer als auch medizinischer Sicht.

Zwischen den beiden Therapiearmen zeigt sich insgesamt keine unterschiedliche Inzidenz **unerwünschter Ereignisse (UE)**. Auf die Therapiedauer bezogen, traten unter Pla+T+D 18,72 UE gegenüber 16,88 UE pro Patient und Jahr unter Ptz+T+D auf. **Das Sicherheitsprofil war auch aus medizinischer Sicht vergleichbar**.

Die Gesamtbetrachtung der SUE zeigt, dass unter Ptz+T+D zwar statistisch mehr SUE beobachtet wurden, diese in der Regel aber gut behandelbar waren. Somit war das Sicherheitsprofil von **Ptz+T+D hinsichtlich SUE aus medizinischer Sicht vergleichbar zu Pla+T+D**.

Die Werte für die Kategorie **„Behandlungsabbruch wegen UE“** zwischen beiden Armen waren statistisch nicht unterschiedlich. **Das Sicherheitsprofil war auch aus medizinischer Sicht vergleichbar**.

Die Inzidenz **UE Grad ≥ 3** war zwischen den Behandlungsarmen ausgewogen. Das Sicherheitsprofil von Ptz+T+D hinsichtlich der UE Grad ≥ 3 war **auch aus medizinischer Sicht vergleichbar zu Pla+T+D**.

Die Betrachtung der Inzidenzen der UE spezifisch in den einzelnen Schweregradgruppen (**Grad 3, Grad 4 und Grad 5**) zeigte keine statistisch signifikanten Unterschiede zwischen den Behandlungsarmen.

Todesfälle (UE Grad 5) waren in beiden Armen selten und **das Sicherheitsprofil war aus medizinischer Sicht vergleichbar**.

Aus medizinischer Sicht liegt **kein Unterschied des Sicherheitsprofils** von Ptz hinsichtlich **Diarrhö** vor, auch wenn deren Inzidenz statistisch signifikant erhöht ist. Gleiches gilt für die Exantheme und Mukositis.

Aufgrund der statistisch signifikant häufiger beobachteten **febrilen Neutropenien** und febrilen neutropenischen Infektionen liegt ein Signal vor. In der klinischen Gesamtbetrachtung im Vergleich zu Pla+T+D hat dieses Signal jedoch **auf das Nutzen-Risikoprofil von Ptz+T+D keinen Einfluss**.

Insgesamt traten nicht mehr und keine schwereren Nebenwirkungen auf, wenn Patienten mit HER2-positivem metastasierten oder lokal rezidivierenden inoperablen Brustkrebs zusätzlich mit Ptz behandelt wurden. In der **Gesamtbetrachtung ist das Sicherheitsprofil von Ptz+T+D vergleichbar mit dem Sicherheitsprofil von Pla+T+D.**

Zusammengefasst fordert die Definition gemäß AM-NutzenV für einen erheblichen Zusatznutzen eine nachhaltige und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte große Verbesserung des therapierelevanten Nutzens im Sinne einer erheblichen Verbesserung der Mortalität, Morbidität oder Lebensqualität.

Die zusätzliche Gabe von Pertuzumab zur Standardtherapie resultierte in einer signifikanten und klinisch relevanten erheblichen Verlängerung des **Gesamt- und des progressionsfreien Überlebens,** sowie in einer beträchtlichen Verbesserung der **Gesamtansprechrate**. Diese gehen einher mit einer beträchtlichen Verbesserung der brustkrebsspezifischen **Lebensqualität**. Damit liegt eine Verbesserung durch Pertuzumab in den Dimensionen Mortalität, Morbidität und Lebensqualität vor.

Zusammengefasst zeigt sich unter der Therapie mit Pertuzumab in Kombination mit Trastuzumab und Docetaxel gegenüber der Standardtherapie ein **erheblicher Zusatznutzen**.

4.4.4 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 4.4.3 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 4-84: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens

Bezeichnung der Patientengruppen	**Ausmaß des Zusatznutzens**
Erwachsene Patienten mit HER2-positivem metastasiertem oder lokal rezidivierendem, inoperablem Brustkrebs, die zuvor noch keine anti-HER2-Therapie oder Chemotherapie zur Behandlung ihrer metastasierten Erkrankung erhalten haben.	Erheblich

4.4.5 Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht – Arzneimittel zur Behandlung eines seltenen Leidens

Für Arzneimittel, die zur Behandlung eines seltenen Leidens zugelassen sind, müssen keine Nachweise zum medizinischen Nutzen und medizinischen Zusatznutzen vorgelegt werden (Abschnitt 4.3), solange der Umsatz des Arzneimittels mit der gesetzlichen Krankenversicherung zu Apothekenverkaufspreisen einschließlich Umsatzsteuer in den letzten 12 Monaten einen Betrag von 50 Millionen Euro nicht übersteigt.

Für diese Arzneimittel sind Angaben zu Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, im vorliegenden Abschnitt einzureichen. Dabei soll zunächst das Ausmaß des Zusatznutzens (in der Definition gemäß AM-NutzenV) begründet beschrieben werden. Unter Berücksichtigung des Ausmaßes des Zusatznutzens sollen Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, benannt werden. Weitere Abschnitte von Modul 4 können dabei nach Bedarf bearbeitet werden.

Beschreiben Sie begründet das Ausmaß des Zusatznutzens unter Verwendung der Kategorisierung gemäß AM-NutzenV. Berücksichtigen Sie bei den Aussagen zum Zusatznutzen nachgewiesene Unterschiede zwischen verschiedenen Patientengruppen. Verweisen Sie, wo notwendig, auf weitere Abschnitte von Modul 4.

<< Nicht zutreffend >>

Geben Sie auf Basis des Ausmaßes des Zusatznutzens in der nachfolgenden Tabelle an, für welche Patientengruppen ein therapeutisch bedeutsamer Zusatznutzen besteht. Benennen Sie das Ausmaß des Zusatznutzens in Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen. Fügen Sie für jede Patientengruppe mit therapeutisch bedeutsamem Zusatznutzen eine neue Zeile ein.

Tabelle 4-85: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens –Arzneimittel zur Behandlung eines seltenen Leidens

Bezeichnung der Patientengruppen	**Ausmaß des Zusatznutzens**

4.5 Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte

4.5.1 Begründung für die Vorlage indirekter Vergleiche

Sofern mit dem Dossier indirekte Vergleiche (Abschnitt 4.3.2.1) eingereicht wurden, begründen Sie dies. Begründen Sie dabei auch, warum sich die ausgewählten Studien jeweils für einen indirekten Vergleich gegenüber dem zu bewertenden Arzneimittel und damit für den Nachweis eines Zusatznutzens durch indirekten Vergleich eignen.

<< Nicht zutreffend >>

4.5.2 Begründung für die Vorlage nichtrandomisierter vergleichender Studien und weiterer Untersuchungen

Sofern mit dem Dossier nichtrandomisierte vergleichende Studien (Abschnitt 4.3.2.2) oder weitere Untersuchungen (Abschnitt 4.3.2.3) eingereicht wurden, nennen Sie die Gründe, nach denen es unmöglich oder unangemessen ist, zu den in diesen Studien bzw. Untersuchungen behandelten Fragestellungen Studien höchster Evidenzstufe (randomisierte klinische Studien) durchzuführen oder zu fordern.

<< 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.

<< 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 bzw. 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[13],

[13] *Burzykowski T (Ed.: The evaluation of surrogate endpoints. New York: Springer; 2005.*

Molenberghs2010[14]). Diese Studien müssen bei Patientenkollektiven und Interventionen durchgeführt worden sein, die Aussagen für das dem vorliegenden Antrag zugrunde liegende Anwendungsgebiet und das zu bewertende Arzneimittel sowie die Vergleichstherapie erlauben.

Eine Möglichkeit der Verwendung von Surrogatendpunkten ohne abschließende Validierung stellt die Anwendung des Konzepts eines sogenannten Surrogate-Threshold-Effekts (STE) (Burzykowski 2006[15]) 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[16]) soll ausreichend begründet werden, insbesondere dann, wenn als Datengrundlage nur eine einzige Studie verwendet werden soll.

Berichten Sie zu den Studien zur Validierung oder zur Begründung für die Verwendung von Surrogatendpunkten mindestens folgende Informationen:

– Patientenpopulation
– Intervention
– Kontrolle
– Datenherkunft
– verwendete Methodik
– entsprechende Ergebnisse (abhängig von der Methode)
– Untersuchungen zur Robustheit
– ggf. Untersuchungen zur Übertragbarkeit

Sofern Sie im Dossier Ergebnisse zu Surrogatendpunkten eingereicht haben, benennen Sie die Gründe für die Verwendung von Surrogatendpunkten. Beschreiben Sie, ob und warum die verwendeten Surrogatendpunkte im betrachteten Kontext valide Surrogatendpunkte darstellen bzw. Aussagen zu patientenrelevanten Endpunkten zulassen.

<< Nicht zutreffend >>

[14] *Molenberghs G, Burzykowski T, Alonso A, Assam P, Tilahun A, Buyse M: A unified framework for the evaluation of surrogate endpoints in mental-health clinical trials. Stat Methods Med Res 2010; 19(3): 205-236.*
[15] *Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. Pharm Stat 2006; 5(3): 173-186.*
[16] *Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. Stat Med 2006; 25(2): 183-203.*

4.6 Liste der eingeschlossenen Studien

Listen Sie alle für die Nutzenbewertung berücksichtigten Studien und Untersuchungen unter Angabe der im Dossier verwendeten Studienbezeichnung und der zugehörigen Datenquellen auf.

Studienberichte und zusätzliche Auswertungen zur CLEOPATRA-Studie:

CLEOPATRA (WO20698). Clinical Study Report –WO20698C/TOC4129g - A Phase III, Randomized, Double-Blind, Placebo-Controlled Clinical Trial to Evaluate the Efficacy and Safety of Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel in Previously Untreated HER2-Positive Metastatic Breast Cancer – Report No. 1046288, Roche, October 2011. [46]

CLEOPATRA (WO20698). Update Clinical Study Report - WO20698C/TOC4129g - A Phase III, Randomized, Double-Blind, Placebo-Controlled Clinical Trial to Evaluate the Efficacy and Safety of Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel in Previously Untreated HER2-Positive Metastatic Breast Cancer - Report No. 1053649, Roche, Dezember 2012. [47]

Zusätzliche Analysen zur CLEOPATRA-Studie (WO20698-TOC4129g). Roche, 2013. [55]

Publikationen zur CLEOPATRA-Studie:

(1) Baselga J, Swain SM. CLEOPATRA: a phase III evaluation of pertuzumab and trastuzumab for HER2-positive metastatic breast cancer. Clin Breast Cancer 2010; 10(6):489-491. [48]

(2) Baselga J, Cortes J, Kim SB, Im SA, Hegg R, Im YH et al. Pertuzumab plus trastuzumab plus docetaxel for metastatic breast cancer. N Engl J Med 2012; 366(2):109-119. [49]

4.7 Referenzliste

Benennen Sie nachfolgend alle Quellen (z. B. Publikationen, Studienberichte, Studienregistereinträge), die Sie in Modul 4 angegeben haben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard).

(1) IQWIG. Allgemeine Methoden Version 4.0 [online] 09.2011. URL: https://www.iqwig.de/download/IQWiG_Methoden_Version_4_0.pdf [Aufgerufen am 07.08.2012]. 2011.

(2) Roche. Perjeta® 420 mg Konzentrat zur Herstellung einer Infusionslösung Draft der Fachinformation Deutschland 02.2013. 2013.

(3) G-BA. Niederschrift zum Beratungsgespräch gemäß § 8 AM-NutzenV, Beratungsanforderung 2012-B-016 Pertuzumab zur Behandlung des HER2-positiven Brustkrebs. 2012 Jul 16.

(4) G-BA. Niederschrift zum Beratungsgespräch gemäß § 8 AM-NutzenV, Beratungsanforderung 2012-B-016 Pertuzumab zur Behandlung des HER2-positiven Brustkrebs: Nachtrag ZVT zu lokal rezidiviert, inoperabel. 2012 Nov 27.

(5) AGO. Diagnostik und Therapie primärer und metastasierter Mammakarzinome: Chemotherapie des metastasierten Mammakarzinoms Version 2012. 1D [online] 03.2012. URL: http://www.ago-online.de/fileadmin/downloads/leitlinien/mamma/maerz2012/21_2012D_Chemotherapie_bei_metastasiertem_Mammakarzinom.pdf [Aufgerufen am 17.07.2012]. 2012.

(6) AGO. Diagnostik und Therapie primärer und metastasierter Mammakarzinome: Zielgerichtete Substanzen Version 2012. 1D [online] 03.2012. URL: http://www.ago-online.de/fileadmin/downloads/leitlinien/mamma/maerz2012/25_2012D_Zielgerichtete_Therapie.pdf [Aufgerufen am 17.07.2012]. 2012.

(7) AGO. Diagnostik und Therapie primärer und metastasierter Mammakarzinome: Endokrine Therapie des metastasierten Mammakarzinoms Version 2012. 1D [online] 03.2012. URL: http://www.ago-online.de/fileadmin/downloads/leitlinien/mamma/maerz2012/20_2012D_Endokrine_Therapie_des_metastasierten_Mammakarzinoms.pdf [Aufgerufen am 17.07.2012]. 2012.

(8) DKG. Interdisziplinäre S3-Leitlinie für die Diagnostik, Therapie und Nachsorge des Mammakarzinoms Langversion 3.0, Aktualisierung 2012 AWMF-Register-Nummer: 032 . 045OL [online] 07.2012. URL: http://www.krebsgesellschaft.de/download/S3_Brustkrebs_Update_2012_OL_Langversion.pdf [Aufgerufen am 17.07.2012]. 2012.

(9) NCCN. NCCN Clinical Practice Guidelines in Oncology - Breast Cancer, Version 3.2012 [online] 2012. URL: http://www.nccn.org/professionals/physician_gls/pdf/breast.pdf [Aufgerufen am: 12.12.2012]. 2012.

(10) Schneeweiss A. Stellungnahme zur Bestimmung der Strahlentherapie als zulässige Vergleichstherapie für die Kombination Pertuzumab, Trastuzumab und Docetaxel beim lokal rezidivierenden, inoperablen Brustkrebs durch den G-BA. Gutachten 2012.

(11) G-BA. Verfahrensordnung des Gemeinsamen Bundesausschusses (Fassung Okt. 2012) [online] Stand 06.12.2012. URL: http://www.g-ba.de/downloads/62-492-654/VerfO_2012-10-18.pdf [Aufgerufen am 12.12.2012]. 2012.

(12) EMA. Points to consider on application with 1. metaanalyses; 2. one pivotal study [online] 05.2001. URL:http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500003657.pdf [Aufgerufen am 12.12.2012]. 2001.

(13) Schwartz FW. Kurzgutachten zur verhinderten oder die aufgeschobenen krankheitsspezifischen Mortalität als klinische bzw. als patientenrelevante Endpunkte in der Onkologie im Rahmen des GKV-Systems. Kurzutachten im Auftrag für Roche 2011.

(14) IQWIG. Aussagekraft von Surrogatendpunkten in der Onkologie [online] 11.2011. URL: https://www.iqwig.de/download/A10-05_Rapid_Report_Version_1-1_Surrogatendpunkte_in_der_Onkologie.pdf 2011;[Aufgerufen am: 3.12.2012].

(15) Montazeri A. Health-related quality of life in breast cancer patients: A bibliographic review of the literature from 1974 to 2007. Journal of Experimental & Clinical Cancer Research 2008;27(1):32.

(16) FACIT. FACT-B Administration and scoring guidelines. 2003. Report No.: URL: www.facit.org.

(17) Beslija S, Bonneterre J, Burstein HJ, Cocquyt V, Gnant M, Heinemann V, et al. Third consensus on medical treatment of metastatic breast cancer. Ann.Oncol 2009 Nov;20(11):1771-85.

(18) DGHO, Wörmann B, Overkamp F, Rick O, Possinger K. Mammakarzinom der Frau Leitlinie [online] 03.2011. URL: http://www.dgho-onkopedia.de/onkopedia/leitlinien/mammakarzinom-der-frau [Aufgerufen am 17.07.2012]. 2011.

(19) Cardoso F, Harbeck N, Fallowfield L, Kyriakides S, Senkus E. Locally recurrent or metastatic breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann.Oncol. 2012 Oct;23 Suppl 7:vii11-vii19.

(20) EMA. Note for guidance on studies in support of special populations: geriatrics (CPMP/ICH/379/95) [online] 03.1994. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002875.pdf 1994;[Aufgerufen am 12.12.2012].

(21) EMA. Adequacy of guidance on the Elderly regarding Medicinal Products for human use,Doc.Ref.EMEA/498920/ [online] 12.2006. URL:

http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/01/WC500049541.pdf 2006;[Aufgerufen am 12.12.2012].

(22) FDA. Guidance for Industry E7 Studies in Support of Special Populations: Geriatrics, Questions ans Answers [online] 02.2012. URL: http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM189544.pdf 2012;[Aufgerufen am 12.12.2012].

(23) ICH. Final Concept Paper E7(R1): Studies in Support of Special Populations: Geriatrics (Revision of the ICH E7 Guideline) [online] 10.2008. URL: http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E7/Q_As/E7R1_Final_Concept_Paper.pdf 2008;[Aufgerufen am 12.12.2012].

(24) EMA. Note for guidance on structure and content of clinical study reports (CPMP/ICH/137/95) [online] 07.1996. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002832.pdf 1996;[Aufgerufen am 12.12.2012].

(25) EMA. Note for guidance on ethnic factors in the acceptability of foreign clinical data CPMP/ICH/289/95 [online] 09.1998. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC500002842.pdf 1998;[Aufgerufen am 12.12.2012].

(26) Exner DV, Dries DL, Domanski MJ, Cohn JN. Lesser response to angiotensin-converting-enzyme inhibitor therapy in black as compared with white patients with left ventricular dysfunction. N.Engl.J.Med. 2001 May 3;344(18):1351-7.

(27) FDA. Guidance for Industry Collection of Race and Ethnicity Data in Clinical Trials [online] 09.2005. URL: http://www.fda.gov/downloads/RegulatoryInformation/Guidances/ucm126396.pdf 2005;[Aufgerufen am 12.12.2012].

(28) Reddy KR, Hoofnagle JH, Tong MJ, Lee WM, Pockros P, Heathcote EJ, et al. Racial differences in responses to therapy with interferon in chronic hepatitis C. Consensus Interferon Study Group. Hepatology 1999 Sep;30(3):787-93.

(29) Xie HG, Kim RB, Wood AJ, Stein CM. Molecular basis of ethnic differences in drug disposition and response. Annu.Rev.Pharmacol.Toxicol. 2001;41:815-50.

(30) Clark GM, Sledge GW, Jr., Osborne CK, McGuire WL. Survival from first recurrence: relative importance of prognostic factors in 1,015 breast cancer patients. J.Clin.Oncol. 1987 Jan;5(1):55-61.

(31) Goldhirsch A, Gelber RD, Castiglione M. Relapse of breast cancer after adjuvant treatment in premenopausal and perimenopausal women: patterns and prognoses. J.Clin.Oncol. 1988 Jan;6(1):89-97.

(32) Sherry MM, Greco FA, Johnson DH, Hainsworth JD. Metastatic breast cancer confined to the skeletal system. An indolent disease. Am.J.Med. 1986 Sep;81(3):381-6.

(33) Swenerton KD, Legha SS, Smith T, Hortobagyi GN, Gehan EA, Yap HY, et al. Prognostic factors in metastatic breast cancer treated with combination chemotherapy. Cancer Res. 1979 May;39(5):1552-62.

(34) Roche. Herceptin® 150 mg Fachinformation [online] 02.2012. URL: http://www.fachinfo.de [Aufgerufen am 29.08.2012]. 2012.

(35) Slamon DJ, Leyland-Jones B, Shak S, Fuchs H, Paton V, Bajamonde A, et al. Use of Chemotherapy plus a Monoclonal Antibody against HER2 for Metastatic Breast Cancer That Overexpresses HER2. N Engl J Med 2001 Mar 15;344(11):783-92.

(36) Oken MM, Creech RH, Tormey DC, Horton J, Davis TE, McFadden ET, et al. Toxicity and response criteria of the Eastern Cooperative Oncology Group. Am.J.Clin.Oncol. 1982 Dec;5(6):649-55.

(37) Cardoso F, Costa A, Norton L, Cameron D, Cufer T, Fallowfield L, et al. 1st International consensus guidelines for advanced breast cancer (ABC 1). The Breast 2012 Jun;21(3):242-52.

(38) RKI. Krebs in Deutschland 2007/2008 8. Ausgabe 2012 [online] 2012. URL: http://www.krebsdaten.de/Krebs/DE/Content/Publikationen/Krebs_in_Deutschland/krebs_in_deutschland_node.html;jsessionid=C6511404A56D0713CFAA559C2BDBC84E.2_cid030 [Aufgerufen am 05.03.2012]. 2012.

(39) F.Hoffmann-La Roche Ltd. CLINICAL STUDY PROTOCOL NUMBER BO22589 D / TDM4788G D-EU Stand 10.2011 (MARIANNE-Studie). 2011.

(40) F.Hoffmann-La Roche Ltd. CLINICAL STUDY PROTOCOL PROTOCOL NUMBER MO27775 Stand 08.2012 (PERTAIN-Studie). 2012.

(41) EORTC. EORTC protocol 75111-10114 Pertuzumab + trastuzumab (PH) versus PH plus metronomic chemotherapy (PHM) in the elderly HER2+ metastatic breast cancer population who may continue on T-DM1 alone following disease progression while on PH / PHM: an open-label multicentre randomized phase II selection trial of the EORTC Elderly Task Force and Breast Cancer Group Stand 05.2012 (NCT01597414). 2012.

(42) ClinicalTrials.gov. Registereintrag zur Studie NCT00567190 (CLEOPATRA) Stand 10.2012. URL: http://clinicaltrials.gov/ct2/show/NCT00567190 [Aufgerufen am 09.01.2013]. 2012.

(43) ICTRP. Registereintrag zur Studie NCT00567190 (CLEOPATRA) Stand 10.2012. URL: http://apps.who.int/trialsearch/Trial.aspx?TrialID=NCT00567190 [Aufgerufen am 09.01.2013]. 2012.

(44) ICTRP. Registereintrag zur Studie EUCTR2007-002997-72-FI (CLEOPATRA) Stand 06.2012. URL: http://apps.who.int/trialsearch/Trial.aspx?TrialID=EUCTR2007-002997-72-FI [Aufgerufen am 09.01.2013]. 2012.

(45) Roche Trials Database. Registereintrag zur CLEOPATRA-Studie (Protokoll No WO20698) Stand 04.2012. URL: http://www.roche-

trials.com/trialDetailsGet.action?studyNumber=WO20698&productGenericName=pertuzumab&productType=Drug&divisionName=PHA [Aufgerufen am 09.01.2013]. 2012.

(46) Roche. CLEOPATRA: Clinical Study Report -WO20698C/TOC4129g: A Phase III, Randomized, Double-Blind, Placebo-Controlled Clinical Trial to Evaluate the Efficacy and Safety of Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel in Previously Untreated HER2-Positive Metastatic Breast Cancer Report No. 1046288, October 2011. 2011.

(47) Roche. Update Clinical Study Report - WO20698C/TOC4129g - A Phase III, Randomized, Double-Blind, Placebo-Controlled Clinical Trial to Evaluate the Efficacy and Safety of Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel in Previously Untreated HER2-Positive Metastatic Breast Cancer - Report No. 1053649, Dezember 2012. 2012.

(48) Baselga J, Swain SM. CLEOPATRA: a phase III evaluation of pertuzumab and trastuzumab for HER2-positive metastatic breast cancer. Clin Breast Cancer 2010 Dec 1;10(6):489-91.

(49) Baselga J, Cortès J, Kim SB, Im SA, Hegg R, Im YH, et al. Pertuzumab plus Trastuzumab plus Docetaxel for Metastatic Breast Cancer. N Engl J Med 2012 Jan 12;366(2):109-19.

(50) Snapinn S, Jiang Q. On the clinical meaningfulness of a treatment's effect on a time-to-event variable. Stat.Med 2011 Aug 30;30(19):2341-8.

(51) Lange S, Kaiser T, Beatrice-Schüler Y, Skipka G, Vervölgyi V, Wieseler B. Bewertung der klinischen Relevanz bei der Nutzenbewertung - Ein formalisiertes Vorgehen zur Bewertung der Relevanz von Gruppenunterschieden auf Skalen [online] 06.2010. [Aufgerufen am 20.12.2012]. URL: https://www.iqwig.de/download/IQWiG_im_Dialog_2010_Stefan_Lange.pdf. 2010. IQWiG im Dialog, Köln.

(52) Victor N. On clinically relevant differences and shifted null hypotheses. Methods Inf.Med 1987 Jul;26(3):109-16.

(53) Pocock SJ, Clayton TC, Altman DG. Survival plots of time-to-event outcomes in clinical trials: good practice and pitfalls. Lancet 2002 May 11;359(9318):1686-9.

(54) Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J.Natl.Cancer Inst. 2000 Feb 2;92(3):205-16.

(55) Roche. Zusätzliche Analysen zur CLEOPATRA-Studie (WO20698-TOC4129g). 2013.

(56) FACIT. FACT-B Fragebogen: German. 2012.

(57) Eton DT, Cella D, Yost KJ, Yount SE, Peterman AH, Neuberg DS, et al. A combination of distribution- and anchor-based approaches determined minimally important differences (MIDs) for four endpoints in a breast cancer scale. Journal of clinical epidemiology 57[9], 898-910. 1-9-2004.

(58) Brady MJ, Cella DF, Mo F, Bonomi AE, Tulsky DS, Lloyd SR, et al. Reliability and validity of the Functional Assessment of Cancer Therapy-Breast quality-of-life instrument. Journal of Clinical Oncology 1997 Mar 1;15(3):974-86.

(59) Webster K, Odom L, Peterman A, Lent L, Cella D. The functional assessment of chronic illness therapy (FACIT) Measurement system: Validation of version 4 of the core questionnaire: Abstract 216 page 604. Quality of Life Research 1999 8, 555-665. 1999. Kluwer Academic Publishers.

(60) Blackwell KL, Burstein HJ, Storniolo AM, Rugo H, Sledge G, Koehler M, et al. Randomized study of Lapatinib alone or in combination with trastuzumab in women with ErbB2-positive, trastuzumab-refractory metastatic breast cancer. J Clin Oncol 2010 Mar 1;28(7):1124-30.

(61) Geyer CE, Forster J, Lindquist D, Chan S, Romieu CG, Pienkowski T, et al. Lapatinib plus capecitabine for HER2-positive advanced breast cancer. N Engl J Med 2006 Dec 28;355(26):2733-43.

(62) Johnston S, Pippen J, Jr., Pivot X, Lichinitser M, Sadeghi S, Dieras V, et al. Lapatinib combined with letrozole versus letrozole and placebo as first-line therapy for postmenopausal hormone receptor-positive metastatic breast cancer. J Clin Oncol 2009 Nov 20;27(33):5538-46.

(63) DKG, Hartmann JT, Dörr W, Steingräber M, Grötz KA. Kurzgefasste interdisziplinäre Leitlinien 2008: Kapitel O9 Schleimhauttoxizität [online] 2008. URL: http://www.krebsgesellschaft.de/download/ll_o_09.pdf [Aufgerufen am 19.12.2012], 381-385. 2008. W. Zuckschwerdt Verlag GmbH.

(64) Lacouture ME. Mechanisms of cutaneous toxicities to EGFR inhibitors. Nat Rev Cancer 2006 Oct;6(10):803-12.

(65) Onkovis. Doce onkovis 20 mg/ml Konzentrat zur Herstellung einer Infusionslösung Fachinformation [online] 06.2011. URL: http://www.fachinfo.de [Aufgerufen am 28.08.2012]. 2011.

(66) Blijlevens NM, Logan RM, Netea MG. The changing face of febrile neutropenia-from monotherapy to moulds to mucositis. Mucositis: from febrile neutropenia to febrile mucositis. J.Antimicrob.Chemother. 2009 May;63 Suppl 1:i36-i40.

(67) Pico JL, Avila-Garavito A, Naccache P. Mucositis: Its Occurrence, Consequences, and Treatment in the Oncology Setting. Oncologist. 1998;3(6):446-51.

(68) Baselga J, Bradbury I, Eidtmann H, Di CS, de AE, Aura C, et al. Lapatinib with trastuzumab for HER2-positive early breast cancer (NeoALTTO): a randomised, open-label, multicentre, phase 3 trial. Lancet 2012 Feb 18;379(9816):633-40.

(69) Marty M, Cognetti F, Maraninchi D, Snyder R, Mauriac L, Tubiana-Hulin M, et al. Randomized Phase II Trial of the Efficacy and Safety of Trastuzumab Combined With Docetaxel in Patients With Human Epidermal Growth Factor Receptor 2- Positive Metastatic Breast Cancer Administered As First-Line Treatment: The M77001 Study Group. Journal of Clinical Oncology 2005 Jul 1;23(19):4265-74.

(70) Gianni L, Pienkowski T, Im YH, Roman L, Tseng LM, Liu MC, et al. Efficacy and safety of neoadjuvant pertuzumab and trastuzumab in women with locally advanced, inflammatory, or early HER2-positive breast cancer (NeoSphere): a randomised multicentre, open-label, phase 2 trial. Lancet Oncol 2012 Jan;13(1):25-32.

(71) EMA. Draft EPAR Perjeta Dezember 2012. 2012.

(72) Windeler J. Externe Validität. Evid.Fortild.Qual.Gesundh.wesen 2008;102:253-60.

(73) Iomedico (Hrsg.). Tumorregister Mammakarzinom. Auswertung August 2012.

(74) Valero V, Forbes J, Pegram MD, Pienkowski T, Eiermann W, von Minckwitz G, et al. Multicenter Phase III Randomized Trial Comparing Docetaxel and Trastuzumab With Docetaxel, Carboplatin, and Trastuzumab As First-Line Chemotherapy for Patients With HER2-Gene-Amplified Metastatic Breast Cancer (BCIRG 007 Study): Two Highly Active Therapeutic Regimens. Journal of Clinical Oncology 2011 Jan 10;29(2):149-56.

(75) Wardley AM, Pivot X, Morales-Vasquez F, Zetina LM, de Fatima Dias GM, Reyes DO, et al. Randomized phase II trial of first-line trastuzumab plus docetaxel and capecitabine compared with trastuzumab plus docetaxel in HER2-positive metastatic breast cancer. J Clin Oncol 2010 Feb 20;28(6):976-83.

(76) ClinicalTrials.gov. Registereintrag zur Studie NCT01120184 (MARIANNE) Stand 02.2013. URL: http://www.clinicaltrials.gov/ct2/show/NCT01120184?term=NCT01120184&rank=1 [Aufgerufen am 06.02.2013]. 2013.

(77) ICTRP. Registereintrag zur Studie EUCTR2009-017905-13-AT (MARIANNE) Stand 04.2012. URL: http://apps.who.int/trialsearch/Trial.aspx?TrialID=EUCTR2009-017905-13-AT [Aufgerufen am 28.02.2013]. 2012.

(78) Roche Trials Database. Registereintrag zur MARIANNE-Studie (BO22589) Stand 01.2013. URL: http://www.roche-trials.com/trialDetailsGet.action?studyNumber=BO22589 [Aufgerufen am 06.02.2013]. 2013.

(79) ClinicalTrials.gov. Registereintrag zur Studie NCT01491737 (PERTAIN) Stand 02.2013. URL: http://www.clinicaltrials.gov/ct2/show/NCT01491737?term=NCT01491737&rank=1 [Aufgerufen am 06.02.2013]. 2013.

(80) ICTRP. Registereintrag zur Studie EUCTR2011-002132-10-ES (PERTAIN) Stand 09.2012. URL: http://apps.who.int/trialsearch/Trial.aspx?TrialID=EUCTR2011-002132-10-ES [Aufgerufen am 06.02.2013]. 2012.

(81) ICTRP. Registereintrag zur Studie NCT01491737 (PERTAIN) Stand 02.2013. URL: http://apps.who.int/trialsearch/Trial.aspx?TrialID=NCT01491737 [Aufgerufen am 06.02.2013]. 2013.

(82) Roche Trials Database. Registereintrag zur PERTAIN-Studie (MO27775) Stand 01.2013. URL: http://www.roche-trials.com/trialDetailsGet.action?studyNumber=MO27775 [Aufgerufen am 06.02.2013]. 2013.

(83) ClinicalTrials.gov. Registereintrag zur Studie NCT01597414 (Elderly Metastatic Breast Cancer: Pertuzumab-Herceptin vs Pertuzumab-Herceptin-Metronomic Chemotherapy, Followed by T-DM1) Stand 06.2012. URL: http://www.clinicaltrials.gov/ct2/show/NCT01597414?term=NCT01597414&rank=1 [Aufgerufen am 06.02.2013]. 2012.

(84) ICTRP. Registereintrag zur Studie NCT01597414 (Elderly Metastatic Breast Cancer: Pertuzumab-Herceptin vs Pertuzumab-Herceptin-Metronomic Chemotherapy, Followed by T-DM1) Stand 10.2012. URL: http://apps.who.int/trialsearch/Trial.aspx?TrialID=NCT01597414 [Aufgerufen am 06.02.2013]. 2012.

Anhang 4-A: Suchstrategien – bibliografische Literaturrecherche

Geben Sie nachfolgend die Suchstrategien für die in Abschnitt 4.2.3.2 beschriebene bibliografische Literaturrecherche an, und zwar einzeln für jede Datenbank. Geben Sie dabei zunächst jeweils den Namen der durchsuchten Datenbank (z. B. EMBASE), die verwendete Suchoberfläche (z. B. DIMDI, Ovid etc.), das Datum der Suche, das Zeitsegment (z. B.: „1980 to 2010 week 50“) und die gegebenenfalls verwendeten Suchfilter (mit Angabe einer Quelle) an. Listen Sie danach die Suchstrategie einschließlich der resultierenden Trefferzahlen auf. Geben Sie zusätzlich an, für welchen Abschnitt von Modul 4 die Suche durchgeführt wurde.

Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel:

Datenbankname	EMBASE
Suchoberfläche	Ovid
Datum der Suche	08.12.2010
Zeitsegment	1980 to 2010 week 50
Suchfilter	Filter für randomisierte kontrollierte Studien nach Wong 2006 [Quelle[17]] – Strategy minimizing difference between sensitivity and specificity

#	Suchbegriffe	Ergebnis
1	Meglitinide/	848
2	Nateglinide/	1686
3	Repaglinide/	2118
4	(glinid* or meglitinid* or nateglinid* or repaglinid*).ab,ti.	1069
5	(starlix or novonorm or novo norm or prandin).ab,ti.	32
6	(105816-04-4 or 135062-02-1).rn.	2854
7	or/1-6	3467
8	Diabetes mellitus/	224164
9	Non Insulin dependent Diabetes mellitus/	91081
10	(diabet* or niddm or t2dm).ab,ti.	379777
11	or/8-10	454517
12	(random* or double-blind*).tw.	650136
13	Placebo*.mp.	243550
14	or/12-13	773621
15	and/7,11,14	719

[17] Das Zitat zu dem hier beispielhaft angegebenen Suchfilter lautet wie folgt: Wong SSL, Wilczynski NL, Haynes RB. Comparison of top-performing search strategies for detecting clinically sound treatment studies and systematic reviews in MEDLINE and EMBASE. J Med Libr Assoc 2006; 94(4): 451-455.

Die Durchführung einer bibliografischen Literaturrecherche ist für den Wirkstoff Pertuzumab nicht erforderlich:

1. Die weltweit erste Zulassung für Pertuzumab wurde weniger als 12 Monate (FDA Zulassung am 08. Juni 2012) vor dem Zeitpunkt der Dossiereinreichung erteilt.

2. Zur Darstellung des (Zusatz-)Nutzens wurden keine indirekten Vergleiche herangezogen.

Anhang 4-B: Suchstrategien – Suche in Studienregistern

Geben Sie nachfolgend die Suchstrategien für die in Abschnitt 4.2.3.3 beschriebene Suche in Studienregistern an, und zwar einzeln für jedes Studienregister. Geben Sie dabei jeweils den Namen des durchsuchten Studienregisters (z. B. clinicaltrials.gov), die Internetadresse, unter der das Studienregister erreichbar ist (z. B. http://www.clinicaltrials.gov), das Datum der Suche, die verwendete Suchstrategie und die resultierenden Treffer an. Geben Sie zusätzlich an, für welchen Abschnitt von Modul 4 die Suche durchgeführt wurde.

Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel:

Studienregister	clinicaltrials.gov
Internetadresse	http://www.clinicaltrials.gov
Datum der Suche	08.12.2010
Suchstrategie	(Starlix OR Novonorm OR Prandin OR Nateglinid OR Repaglinid) [ALL-FIELDS] AND ("Phase II" OR "Phase III" OR "Phase IV") [PHASE]
Treffer	23

Studienregister	clinicaltrials.gov
Internetadresse	http://www.clinicaltrials.gov
Datum der Suche	10.12.2012
Suchstrategie	(Pertuzumab OR Omnitarg OR Perjeta OR RhuMab 2C4 OR RO4368451)
Treffer	39 Studien
Offensichtliche Verletzung der Einschlusskriterien (Vorselektion*)	38 Studien
Einschluss	1 Studie (CLEOPATRA)

Studienregister	ICTRP
Internetadresse	http://www.who.int/ictrp/search/en/
Datum der Suche	11.12.2012
Suchstrategie	(Pertuzumab OR Omnitarg OR Perjeta OR RhuMab 2C4 OR RO4368451)
Treffer	139 Treffer für 43 Studien
Offensichtliche Verletzung der Einschlusskriterien (Vorselektion*)	41 Studien
Einschluss	2 Treffer für 1 Studie (CLEOPATRA)

*Vorselektion: Die angezeigten Treffer wurden zunächst anhand der elektronisch verfügbaren Informationen auf offensichtliche Verletzungen der vorab definierten Einschlusskriterien (siehe Tabelle 4-1) untersucht. Treffer, die bereits an dieser Stelle eine eindeutige Verletzung eines oder mehrerer der Einschlusskriterien erkennen liessen, wurden im ersten Selektionsschritt ohne weitere Dokumentation der Gründe ausgeschlossen (siehe 4.2.3.3).

Studienregister	Roche trials database: Clinical Trial Protocol Registry
Internetadresse	http://www.roche-trials.com/protocolRegistry.action
Datum der Suche	09.01.2013
Suchstrategie	Pertuzumab / Pertuzumab [Perjeta]
Treffer	14 Studien / 1 Studie
Offensichtliche Verletzung der Einschlusskriterien (Vorselektion*)	13 Studien / 1 Studie
Einschluss	1 Studie (CLEOPATRA)

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

Listen Sie die im Volltext gesichteten und ausgeschlossenen Studien unter Nennung des Ausschlussgrunds auf.

Es wurden keine Publikationen im Volltext ausgeschlossen.

Anhang 4-D: Liste der abgebrochenen Studien

Benennen Sie in der nachfolgenden Tabelle die abgebrochenen Studien aus dem durch die verschiedenen Suchschritte in Abschnitt 4.3.1.1identifizierten Studienpool.

Tabelle 4-86 (Anhang): Studienpool, abgebrochene Studien, zu bewertendes Arzneimittel

Studie	**Studienkategorie**			**verfügbare Datenquellen[a]**		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	**gesponserte Studie[b] (ja/nein)**	**Studie Dritter (ja/nein)**	**Studienbericht (ja/nein [Zitat])**	**Registereintrag[c] (ja/nein [Zitat])**	**Publikation (ja/nein [Zitat])**
ggf. Zwischenüberschrift zur Strukturierung des Studienpools						
placebokontrolliert						
aktivkontrolliert, zweckmäßige Vergleichstherapie(n)						

a: Bei Angabe „ja“ sind jeweils die Zitate der Datenquelle(n) mit anzugeben.
b: Studie, für die der Unternehmer Sponsor war oder an der der Unternehmer anderweitig finanziell beteiligt war.
c: Zitat der Studienregistereinträge sowie, falls vorhanden, der in den Studienregistern aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.

Es wurden im Rahmen der bibliographischen sowie der Registersuche keine abgebrochenen Studien identifiziert. Auch in der Liste des pU wurden keine abgebrochenen Studien dokumentiert.

Anhang 4-E: Liste der laufenden Studien

Benennen Sie in der nachfolgenden Tabelle die laufenden Studien aus dem durch die verschiedenen Suchschritte in Abschnitt 4.3.1.1 identifizierten Studienpool.

Tabelle 4-87 (Anhang): Studienpool, laufende Studien, zu bewertendes Arzneimittel

Studie	**Studienkategorie**			**verfügbare Datenquellen[a]**		
	Studie zur Zulassung des zu bewert-enden Arznei-mittels (ja/nein)	**gesponserte Studie[b] (ja/nein)**	**Studie Dritter (ja/nein)**	**Studien-bericht (ja/nein [Zitat])**	**Registereintrag[c] (ja/nein [Zitat])**	**Publikation (ja/nein [Zitat])**
CLEOPATRA (WO20698)	Ja	Ja	Nein	Ja [46;47]	Ja [42-45]	Ja [48;49]
MARIANNE (BO22589) NCT01120184 EUCTR2009-017905-13-AT	Nein	Ja	Nein	Ja [39]	Ja [76-78]	Nein
PERTAIN (MO27775) NCT01491737 EUCTR2011-002132-10-ES	Nein	Ja	Nein	Ja [40]	Ja [79-82]	Nein
NCT01597414 EORTC-75111-10114	Nein	Ja	Ja	Ja [41]	Ja [83;84]	Nein

a: Bei Angabe „ja“ sind jeweils die Zitate der Datenquelle(n) mit anzugeben.

b: Studie, für die der Unternehmer Sponsor ist oder an der der Unternehmer anderweitig finanziell beteiligt ist.

c: Zitat der Studienregistereinträge sowie, falls vorhanden, der in den Studienregistern aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.

Anhang 4-F: Methodik der eingeschlossenen Studien – RCT

Beschreiben Sie nachfolgend die Methodik jeder eingeschlossenen, in Abschnitt 4.3.1.1.4 genannten Studie. Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle 4-88 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 inklusive eines Flow-Charts für den Patientenfluss.

Tabelle 4-88 (Anhang): Studiendesign und -methodik für die CLEOPATRA-Studie (WO20698C/TOC4129g) (Datenquellen: CLEOPATRA (WO20698). Clinical Study Report – WO20698C/TOC4129g - A Phase III, Randomized, Double-Blind, Placebo-Controlled Clinical Trial to Evaluate the Efficacy and Safety of Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel in Previously Untreated HER2-Positive Metastatic Breast Cancer – Report No. 1046288, Roche, October 2011. **Und** CLEOPATRA (WO20698). Update Clinical Study Report - WO20698C/TOC4129g - A Phase III, Randomized, Double-Blind, Placebo-Controlled Clinical Trial to Evaluate the Efficacy and Safety of Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel in Previously Untreated HER2-Positive Metastatic Breast Cancer - Report No. 1053649, Roche, Dezember 2012.)

Item[a]	**Charakteristikum**	**Studieninformation**
-	**Studienziel**	
2 b	Genaue Ziele, Fragestellung und Hypothesen	Primäres Ziel der Studie war der Vergleich des progressionsfreien Überlebens (*progression-free survival* (PFS)) zwischen den beiden Behandlungsgruppen, basierend auf der Bewertung durch ein unabhängiges Review-Kommittee (*independent review facility* (IRF)). Sekundäre Ziele waren die Vergleiche zwischen den Behandlungsarmen für folgende Endpunkte: § Gesamtüberleben (*overall survival* (OS)). § Progressionsfreies Überleben (*progression-free survival* (PFS)) auf Basis der Bewertung der Progression durch den Investigator. § Gesamtansprechrate (*overall response rate*). § Dauer des objektiven Ansprechens (*objective response)*. § Sicherheitsprofil. § Zeit bis zur Symptomprogression, basierend auf dem *Functional Assessment of Cancer Therapy (FACT) Trial Outcome Index—Physical/Functional/Breast* (TOI-PFB) Fragebogen. § Zeit bis zur Verschlechterung des BCS-Score (FACT-B BCS) **Statistische Hypothesen** Die Differenz für den primären Endpunkt für den Vergleich zwischen den beiden Therapiearmen, das IRF-bewertete PFS, wurde mittels dem zwei-seitigen Log-Rank Test auf dem 5% Signifikanz-Niveau berechnet, stratifiziert nach dem vorherigen Behandlungsstatus (de novo und vorherige adjuvante oder neoadjuvante Therapie) und Region (Europa, Nordamerika, Südamerika, und Asien). Die Nullhypothese (H_0) lautete,

Item[a]	Charakteristikum	Studieninformation
		dass die Überlebensverteilung des PFS (S) in den beiden Behandlungsgruppen gleich ist. Die alternative Hypothese (H_1) lautet, dass die Überlebensfunktion für die Verteilung des PFS in den beiden Behandlungsarmen unterschiedlich ist: H_0: $S_{<pertuzumab>} = S_{<placebo>}$ vs H_1: $S_{<pertuzumab>} \neq S_{<placebo>}$ Zusätzliche Tests wurden für den Vergleich, ob die Verteilung oder die Summenstatistik für die sekundären Endpunkte zwischen den beiden Behandlungsarmen gleich waren, mittels zwei-seitigem Test auf dem Alpha-Level von 5% durchgeführt. Die Typ I Fehlerrate für die Analyse des primären Endpunkts PFS, das OS, und das objektive Ansprechen (ORR) wurde bei 5% unter Verwendung einer *fixed-sequence testing procedure* kontrolliert. Die drei Variablen wurden jeweils mit dem zwei-seitigen 5% Signifikanzlevel in der angegebenen Reihenfolge analysiert.
-	**Methoden**	
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	CLEOPATRA ist eine Phase III, randomisierte, doppelt verblindete, placebokontrollierte, parallele, internationale, multizentrische klinische Studie.
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	Die geplante Studie wurde am 17.April 2007 mit der FDA diskutiert sowie am 2. Juli 2007 mit der European Medicines Agency (EMA), woraus sich die folgenden wesentlichen Änderungen ergaben, die im September 2007 in das Studienprotokoll integriert wurden: § Die Fallzahl wurde von 600 auf 800 Patienten erhöht. § Die Erhöhung der Fallzahl basierte auf zwei Aspekten: 1) eine Verbesserung des Gesamtüberlebens (OS) von 33 % mit einer Power von 80 % bei einem zweiseitigen 5 % Signifikanzniveau aufzudecken, und 2) die Folgezeit zwischen finalem PFS und finalen OS Daten zu reduzieren, so dass ungefähr 50 % der benötigten Überlebensrate zum Zeitpunkt der primären Effektivitätsanalyse des PFS aufgetreten sein würden. § Die folgenden Ausschlusskriterien wurden zusätzlich aufgenommen, um zu gewährleisten, dass die Prognose der aufgenommenen Patienten vergleichbar ist: Systemische Brustkrebsbehandlung im neoadjuvanten oder adjuvanten Setting in der Vorgeschichte mit einem Krankheits-freien Intervall zwischen Abschluss der systemischen Behandlung (exklusive Hormonbehandlung) bis zur Diagnose von Metastasen von über 12 Monaten. § Der kardiale Sicherheitsplan wurde folgendermaßen angepasst: Wenn verfügbar, wurden alle Werte, die vor Studienbeginn zur linksventrikulären Ejektionsfraktion (*left ventricular ejection fraction* (LVEF)) auf den Patienenfragebögen (*case report forms* (CRFs)) erfasst. Ein Minimum von 50 % für die LVEF war für die Aufnahme in die Studie erforderlich. Patienten, in deren Vorgeschichte bereits eine Verminderung der LVEF vor der Behandlung mit adjuvanter Trastuzumab-Behandlung dokumentiert war, wurden nicht in die Studie aufgenommen. Daten zur Analyse der Risiken für eine kardiale Dysfunktion sollten erfasst werden. Diese Daten umfassen die vorherige Behandlung mit Anthrazyklinen, vorherige Bestrahlung, vorherige Behandlung mit Trastuzumab, Alter, Raucherstatus und Anamnese sowie

Item[a]	Charakteristikum	Studieninformation
		begleitende kardiale, medizinische Konditionen. Im Fall einer symptomatischen KHK wurden die Daten zur LVEF weiterhin erhoben. § Die Dosis von Docetaxel konnte nach Einschätzung des behandelnden Arztes hinsichtlich der Tolerabilität angepasst werden. Die Docetaxeldosis konnte um 25 % aufgrund von Toxizität vermindert werden, oder auf 100 mg/m^2 erhöht werden, bei Patienten, die diese Dosis tolerierten, um eine optimale Behandlung zu ermöglichen. § Die Definition des primären Endpunkts PFS wurde revidiert: PFS war als die Zeit von der Randomisierung bis zum ersten radiografisch dokumentierten PD definiert, basierend auf der Bewertung des IRF an Hand der RECIST Kriterien (*Response Evaluation Criteria in Solid Tumors*) oder Tod jedweder Ursache – das zuerst auftretende Ereignis wurde gezählt. Eine karzinomatöse Meningitis, diagnostiziert auf Basis einer zytologischen Untersuchung des Liquors, war ebenfalls als PD definiert. Medizinische Fotografie war zur Dokumentation von Rückfällen von subkutanen Läsionen an der Thoraxwand vorgesehen. Bei Patienten, die nicht innerhalb von 18 Wochen nach der letzten Tumorbewertung verstorben waren und keine Dokumentation einer PD hatten, wurde der Zeitpunkt der letzten verfügbaren Tumorbewertung durch das IRF zensiert. Wenn keine Tumorbewertung nach Studienbeginn vorlag, wurde eine Zensierung für den Zeitpunkt der Randomisierung plus einen Tag vorgenommen. Für Patienten, die die Studienbehandlung aufgrund anderer Ursachen als Tod oder IRF-bewertetes PD abbrachen, wurde das Protokoll revidiert, um bis zum Tod bzw. einer IRF-bewerteten PD (was zuerst auftrat) radiografische Tumorbewertung erheben zu können. Um Verzerrungen bei der Bestimmung der Dauer des PFS zu minimieren, wurden Tumorbewertungen alle 9 Wochen nach dem Datum der Randomisierung durchgeführt. § Eine Interim-Analyse für OS wurde für den Zeitpunkt der finalen primären Effektivitätsanalyse für PFS zusätzlich geplant. Die finale Analyse für OS war für den Zeitpunkt geplant, an dem 385 Todesfälle aufgetreten sein würden. § Die *clinical benefit rate* wurde als sekundärer Endpunkt fallen gelassen. § Eine Interim-Sicherheitsanalyse wurde zusätzlich für den Zeitpunkt geplant, an dem 100 Patienten in die Studie aufgenommen sein würden und für mindesten 4 Monate teilgenommen haben würden. Basierend auf Empfehlungen der EMA wurde der ursprüngliche Plan für eine Interim-Effektivitätsanalyse des PFS fallen gelassen. § Aufgrund von EMA-Empfehlungen wurde anstatt der ursprünglich geplanten dynamischen Randomisierung eine Block-Randomisierung mit zwei Stratifikationsfaktoren vorgesehen. § Zwei Amendements zum Original Studienprotokoll wurden erstellt, die die folgenden Änderungen enthielten. **Protokoll Version B (Dezember 2007)** Das Original Studienprotokoll wurde durch die FDA im November 5, 2007 revidiert. Das Amendment B wurde als Antwort auf die FDA

Item[a]	Charakteristikum	Studieninformation
		Empfehlungen erstellt. Die wesentlichen Änderungen umfassen folgende Punkte: § Einschlusskriterium 4 wurde für Patienten, die adjuvante Therapie vor Aufnahme in die Studie erhielten, verändert, um Informationen über LEVF in der Anamnese während und nach adjuvanter Behandlung mit Trastuzumab zu dokumentieren. § LVEF Bewertungen wurden zusätzlich während der Folgebeobachtungszeit vorgesehen, um eine langfristige Erfassung der Herzfunktion zu ermöglichen. Die Zeitpunkte für die LVEF Bewertung für die Patienten, die die Studienbehandlung aufgrund einer Verschlechterung des LEVF unterbrachen, wurden ebenfalls aktualisiert. § Reporting und Grading von symptomatischer linksventrikulärer Dysfunktion (LVSD) wurde in Übereinstimmung mit NCI-CTCAE Version 3.0 gebracht. § Der statistische Analysenplan wurde für die *objective response* angepasst. § Scans zur Tumor-Bewertung, die zu Studienbeginn und während der Studie erforderlich waren, wurden genauer beschrieben. Zur Baseline wurden Patienten mindestens mittels eines CTs oder MTIs für Brustkorb und Abdomen untersucht, PET Scans waren zur Bewertung der Effektivität während der Studie erlaubt. **Protokoll Version C (Juni 2009)** Die Beschreibung der erforderlichen Verhütungsmaßnahmen für die Studie wurde genauer formuliert, um unerwünschte Schwangerschaften zu vermeiden.
4	Probanden/Patienten	
4a	Ein-/Ausschlusskriterien der Probanden/Patienten	**Einschlusskriterien** Die folgenden Kriterien mussten für einen Einschluss in die Studie erfüllt sein: <u>Krankheitsspezifische Einschlusskriterien</u>: § Histologisch oder zytologisch gesichertes Adenokarzinom der Brust mit lokalem Rezidiv oder metastatischer Krankheit, und Eignung für eine Chemotherapie. Patienten mit messbarer und/oder nicht messbarer Krankheit konnten aufgenommen werden. § Patienten, die nur Knochenmetastasen hatten, konnten aufgenommen werden, wenn einige der Metastasen vor dem Studienbeginn nicht bestrahlt worden waren, und Tumorgewebeproben des primären Tumors für einen zentralen HER2 Test sowie eine Biomarker-Analyse verfügbar waren. Lokale Rezidive mussten nicht für eine Resektion mit kurablem Ziel geeignet sein. Patienten mit de-novo Stadium i.v. Krankheit konnten aufgenommen werden. § HER2-positiver (definiert als 3+ IHC oder FISH Amplifikations-Ratio ≥ 2.0) MBC wurde durch ein vom Sponsor eingesetztes zentrales Labor bestätigt. Es wurde empfohlen, dass ein Formalin-fixierter Paraffin-eingelegter (FFPE) Gewebsblock des Primärtumors (oder metastatisch, wenn primär nicht verfügbar) an das zentrale Labor zur Bestätigung der HER2-Rezeptorstatus eingesendet wurde. Wenn dies nicht möglich war, wurden 25 unbehandelte frische Gewebsschnitte eingereicht.

Item[a]	Charakteristikum	Studieninformation
		Generelle Einschlusskriterien: § Alter > 18 Jahre § LVEF > 50 % zu Studienbeginn (innerhalb von 42 Tagen nach Randomisierung), durch Bestimmung mittels ECHO oder MUGA. Alle verfügbaren historischen LVEF-Werte während und nach einer adjuvanten Trastuzumab-Behandlung vor Studienbeginn wurden erfasst. § Eastern Cooperative Oncology Group (ECOG) Leistungsstatus (PS) von 0 oder 1. § Frauen im gebärfähigen Alter und Männer mit einer Partnerin im gebärfähigen Alter, mussten eine hoch-effektive Methode zur Verhütung anwenden. Die Kontrazeption musste während der Studienbehandlung und für mindestens 6 Monat nach der letzten Dosis der Studienbehandlung durchgeführt werden. § Unterschriebene Einverständniserklärung. **Ausschlusskriterien** Patienten mit einem der folgenden Kriterien konnten nicht in die Studie aufgenommen werden: Krankheitsspezifische Ausschlusskriterien: § Anti-tumoröse Therapie für MBC in der Anamnese (unter Ausnahme einer vorherigen hormonalen Therapie für MBC, die vor der Randomisierung gestoppt werden musste). Anti-tumoröse Therapie für MBC umfasste jede EGFR oder anti-HER2 Substanzen oder Vakzine, zytotoxische Chemotherapie, oder mehr als eine vorherige hormonale Therapie für MBC. § Vorherige zugelassene oder experimentelle Therapien, wie Tyrosinkinase/HER Inhibitoren für Brustkrebs in jedem Behandlungssetting, ausgenommen Trastuzumab im neoadjuvanten oder adjuvanten Setting. § Systemische Brustkrebs-Behandlung im neoadjuvanten oder adjuvanten Setting in der Anamnese mit einem Krankheits-freien Intervall zwischen Beendigung der systemischen Behandlung (außer hormonale Therapie) bis zur Diagnose von Metastasen von < 12 Monaten. § Persistierende NCI-CTCAE, Version 3.0 Grad ≥ 2 haematologische Toxizität als Folge einer vorherigen adjuvanten Therapie. § Bestehende periphere Neuropathie Grad ≥ 3 bei Randomisierung. § Andere maligne Erkrankungen in den letzten 5 Jahren, außer *Carcinoma in situ* der Zervix, Basalzell-Karzinom oder squamöses Zellkarzinom der Haut, die mit kurativem Ziel behandelt wurden. § Bestehende klinische oder radiographische Evidenz von Metastasen im Zentralnervensystem (*central nervous system* (CNS)). CT oder MRI Scans des Gehirns waren obligatorisch (innerhalb von 28 Tagen nach der Randomisierung) im Fall des klinischen Verdachts auf Hirnmetastasen. § Erhalt der folgenden kumulativen Dosen von Anthrazyklinen in der Anamnese: – Doxorubicin oder liposomales Doxorubicin > 360 mg/m² – Epirubicin > 720 mg/m² – Mitoxantron > 120 mg/m² und Idarubicin > 90 mg/m² – andere (im Einzelnen liposomales Doxorubicin oder andere

Item[a]	Charakteristikum	Studieninformation
		Anthrazykline > die äquivalenten Dosierungen von 360 mg/m² von Doxorubicin) – wenn mehr als ein Anthrazyklin eingesetzt wurde, durfte die kumulative Dosis ein Äquivalent von 360 mg/m² Doxorubicin nicht übersteigen. <u>Ausschlusskriterien aufgrund von hämatologischen, biochemischen und Organfunktionparametern</u>: § Bestehende unkontrollierte Hypertension (systolisch > 150 mmHg und/oder diastolisch > 100 mmHg) oder unstabile Angina. § Chronisches Herzversagen (CHF) in der Anamnese nach New York Heart Association (NYHA) Kriterien (jede Kategorie), oder ernsthafte kardiale Arrhythmien mit der Notwendigkeit einer Behandlung (Ausnahme: Vorhofflimmern, paroxysmale supraventrikuläre Tachykardie). § Herzinfarkt innerhalb der letzten 6 Monate vor Randomisierung. § LVEF Verschlechterung unter 50 % während oder nach der neo-adjuvanten oder adjuvanten Therapie mit Trastuzumab. § Bestehende Dyspnoe in Ruhe aufgrund von Komplikationen bei fortgeschrittener Krebserkrankung oder andere Erkrankungen, die eine dauerhafte Sauerstofftherapie benötigen. <u>Generelle Ausschlusskriterien:</u> § Inadäquate Organfunktion, verbunden mit folgenden Laborwerten innerhalb von 28 Tagen nach Randomisierung: – Absolute Neutrophilen Zahl < 1.500 Zellen/mm³ – Thrombozytenzahl < 100.000 Zellen/mm³ – Hämoglobin < 9 g/dL – Gesamt Bilirubin > untere Grenze des Normalen (ULN) – AST (SGOT) oder ALT (SGPT) > 2,5 × ULN – AST (SGOT) oder ALT (SGPT) > 1,5 × ULN mit gleichzeitiger Serum alkalischer Phosphatase > 2,5 × ULN. Serum alkalische Phosphatase darf > 2,5 × ULN nur dann übersteigen, wenn Knochenmetastasen vorliegen und AST (SGOT) und ALT (SGPT) < 1,5 × ULN – Serum Kreatinin > 2,0 mg/dL oder 177 µmol/L – Internationale normalisierte Ratio (INR) und aktivierte partielle Thromboplastinzeit oder partielle Thromboplastinzeit (aPTT oder PTT) > 1,5 × ULN (außer, der Patient stand unter therapeutischer Antikoagulation). § Bestehende schwere, unkontrollierte systemische Krankheit (z.B. klinisch signifikante kardiovaskuläre, pulmonale oder metabolische Krankheit; Wundheilungsstörungen; Ulzera oder Knochenfrakturen). § Größere chirurgische Prozeduren oder signifikante traumatische Verletzung innerhalb der letzten 28 Tage nach Beginn der Studienbehandlung oder geplante und erforderliche große chirurgische Eingriffe während der Studienbehandlung. § Schwangerschaft oder Stillzeit. § Invasive Diagnostik innerhalb der letzten 28 Tage vor Randomisierung. § Bestehende bekannte Infektion mit HIV, HBV oder HCV.

Item[a]	Charakteristikum	Studieninformation
		§ Erhalt von i.v. Antibiotika für eine Infektion innerhalb von 14 Tagen nach der Randomisierung. § Bestehende chronische tägliche Behandlung mit Kortikosteroiden (Dosis von > 10 mg/Tag Methylprednisolon Äquivalent) (exkl. inhalative Steroide). § Bekannte Hypersensitivität auf eine der Studiensubstanzen. § Nach Einschätzung des Investigators unfähig oder unwillig, die Erfordernisse des Protokolls zu erfüllen. § Die Teilnahme an anderen interventionellen oder nicht-interventionellen Studien war nicht erlaubt.
4b	Studienorganisation und Ort der Studiendurchführung	Die Studie wurde gemeinsam von Genentech, Inc in den United States (Studiennummer TOC4129g) und von F. Hoffmann-La Roche Ltd für den Rest der Welt (Studiennummer WO20698) durchgeführt. Die Studienperiode liegt zwischen dem 12. Februar 2008 und dem Q3/4 2013. Die Aufnahme der ersten Patientin erfolgte am 12.02.2008. Erster Datenschnitt war am 13.05.2011, der zweite Datenschnitt am 14. Mai 2012. Insgesamt nehmen 204 Zentren in 25 Ländern teil (Brasilien, Kanada, China, Costa Rica, Kroatien, Ecuador, Frankreich, Finnland, Deutschland, United Kingdom, Guatemala, Italien, Japan, Lettland, Mazedonien, Mexiko, Polen, Argentinien, Korea, Philippinen, Russland, Singapore, Spanien, Thailand, USA).
5	Interventionen Präzise Angaben zu den geplanten Interventionen jeder Gruppe und zur Administration etc.	Es war geplant, insgesamt 800 Patienten in der Studie aufzunehmen, randomisiert im Verhältnis 1:1 auf einen der Behandlungsarme (A oder B): Arm A (Pla+T+D) § Pertuzumab Placebo: i.v. Infusion alle 3 Wochen (q3w) § Trastuzumab: Startdosis von 8 mg/kg i.v., gefolgt von 6 mg/kg i.v. q3w § Docetaxel 75 mg/m^2 i.v. q3w für mindestens sechs Zyklen Arm B (Ptz+T+D) § Pertuzumab: Startdosis von 840 mg/kg i.v., gefolgt von 420 mg/kg i.v. q3w § Trastuzumab: Startdosis von 8 mg/kg i.v., gefolgt von 6 mg/kg i.v. q3w § Docetaxel 75 mg/m^2 i.v. q3w für mindestens sechs Wochen. Nach Einschätzung des Investigators, konnte für Patienten, die diese Dosis ohne signifikante Toxizität tolerierten, die Dosis für Docetaxel auf 100 mg/m^2 erhöht werden. Die Behandlung wurde bis zum Auftreten einer durch den Investigator bewerteten radiografisch oder klinisch progressiven Krankheit, einer unakzeptablen Toxizität oder einem Widerruf der Einverständniserklärung des Patienten gegeben. Im Fall eines permanenten Therapieabbruchs oder einer Unterbrechung von mehr als zwei Zyklen, wurde der Patient von der Studie ausgeschlossen. Unterbrach ein Patient nur die Behandlung mit Docetaxel aufgrund einer Toxizität, konnte die Therapie mit Pertuzumab/Placebo und Trastuzumab fortgesetzt werden. **Formulierung und Verpackung**

Item[a]	Charakteristikum	Studieninformation
		Pertuzumab wurde als Einzel-Ampulle mit 30 mg/mL Pertuzumab enthalten in 20 mM L-Histidin (pH 6,0), 120 mM Sucrose, und 0,02 % Polysorbat 20 zur Verfügung gestellt. Jede 20-cc Ampulle enthielt etwa 420 mg Pertuzumab (14,0 mL/Ampulle). Kommerzielle Präparationen von Docetaxel wurden lokal von den Zentren beschafft. Wo durch die lokalen Behörden erlaubt, konnten auch kommerzielle Präparationen von Trastuzumab verwendet werden. Andernfalls wurde Trastuzumab durch den Sponsor als gefriergetrocknete Präparation mit einem nominalen Inhalt von entweder 440 mg oder 150 mg pro Ampulle zur Verfügung gestellt. Jede Lösung enthielt 21 mg/mL der aktiven Substanz bei einem pH von etwa 6,0. **Administration** Pertuzumab/Placebo wurde mit einer Startdosis von840 mg i.v. in Zyklus 1 und mit einer Dosis von 420 mg in den folgenden Zyklen gegeben. Die Dosis war nicht nach Körpergewicht adjustiert. Trastuzumab wurde mit einer Startdosis von 8 mg/kg i.v. in Zyklus 1 und mit einer Dosis von 6 mg/kg in folgenden Zyklen gegeben. Die Dosis von Trastuzumab wurde angepasst, wenn das Körpergewicht sich um mindestens ± 10 % nach Studienbeginn veränderte. Docetaxel wurde in einer Dosis von 75 mg/m^2 i.v. alle drei Wochen gegeben. Bei Patienten, die mindestens einen Zyklus ohne signifikante Toxizität erhalten hatten, konnte die Docetaxel-Dosis auf 100 mg/m^2, entsprechend der Bewertung durch den Investigator, erhöht werden. Anpassungen der Docetaxel-Dosis bei Änderungen des Körpergewichts waren nach Einschätzung und der üblichen Praxis des Investigators möglich. Behandlungszyklen während der Studie hatten eine Dauer von drei Wochen (21 Tage). Die erste Dosis Pertuzumab/Placebo (Zyklus 1, Tag 1) wurde innerhalb der ersten drei Tage nach Randomisierung verabreicht. Die erste Dosis Trastuzumab wurde 24 h Stunden (Zyklus 1, Tag 2) später gegeben, gefolgt von der ersten Dosis Docetaxel. Wurde die initiale Infusion der drei Substanzen nach Einschätzung des Investigators gut toleriert, wurden in den folgenden Zyklen alle drei Substanzen an Tag 1 des Zyklus in der folgenden Reihenfolge gegeben: Pertuzumab/Placebo; Trastuzumab; Docetaxel. Die erste Infusion von Trastuzumab wurde über 90 Minuten gegeben und die verblindete Gabe Pertuzumab/Placebo über 60 Minuten. Wurde die erste Infusion von Trastuzumab und die beiden ersten Infusionen der verblindeten Gabe von Pertuzumab/Placebo ohne Infusions-assoziierte unerwünschte Ereignisse (Fieber oder Schüttelfrost) vertragen, konnten die folgenden Infusionen über 30 Minuten erfolgen. Die Behandlung mit Pertuzumab/Placebo und Trastuzumab wurde bis zum Auftreten einer Investigator-bewerteten PD oder nicht-behandelbaren Toxizität fortgesetzt. Die Behandlung mit Docetaxel sollte für mindestens 6 Zyklen durchgeführt werden, außer der Patient erlitt nicht-behandelbare Toxizitäten oder PD. Nach sechs Zyklen wurde über die Fortsetzung einer Therapie mit Docetaxel nach Einschätzung des Investigators entschieden.
6	Zielkriterien	
6a	Klar definierte primäre und sekundäre Zielkriterien,	**Primärer Effektivitätsparameter** Der primäre Endpunkt war PFS, basierend auf der Tumorbewertung

Item[a]	Charakteristikum	Studieninformation
	Erhebungszeitpunkte, ggf. alle zur Optimierung der Ergebnisqualität verwendeten Erhebungs-methoden (z. B. Mehrfach-beobachtungen, Training der Prüfer) und ggf. Angaben zur Validierung von Erhebungsinstrumenten	durch ein IRF. PFS war als Zeit zwischen der Randomisierung und der ersten Dokumentation einer PD definiert, bestimmt durch das IRF auf der Basis der RECIST-Kriterien, Version 1.0 oder Tod jeglicher Ursache (innerhalb von 18 Wochen nach der letzten Tumorbewertung), je nachdem, welches Ereignis zuerst auftrat. Bewertung der PD basierte auf dem Review von radiografischen (NMR, CT, Knochenscan, Brust-Röntgen etc.), wie auch zytologischen (z.B. relevante zytologische Berichte zur Dokumentation von malignen pleuralen Effusionen, Knochenmark-Aspirationen, Liquor, etc.), und photographischer Daten, wenn verfügbar. **Sekundäre Effektivitätsparameter** § Gesamtüberleben (*overall survival* (OS)) war als Zeit zwischen dem Tag der Randomisierung bis zum Tod aus jeglicher Ursache definiert. § Investigator-bewertetes PFS war als Zeit zwischen dem Tag der Randomisierung bis zum ersten dokumentierten radiografischen PD, bestimmt durch den Investigator unter Verwendung der RECIST-Kriterien oder Tod jeglicher Ursache definiert, je nachdem welches Ereignis zuerst auftrat. § Das Gesamtansprechen, die *overall response rate* (ORR) war als komplettes Ansprechen (*complete response* [CR]), oder partielles Ansprechen (*partial response* [PR]) definiert, bewertet durch das IRF auf der Basis von RECIST-Kriterien an zwei aufeinanderfolgenden Untersuchungen mit einem Abstand von mindestens 4 Wochen (Patienten ohne messbare Krankheit oder mit einer nur auf die Knochen beschränkten Krankheit wurden nicht in die Analyse zur *objective response* eingeschlossen). § Die Dauer der *objective response* war als Periode zwischen dem Datum eines initialen gesicherten partiellen oder kompletten Ansprechens bis zum Datum einer PD oder Tod jeglicher Ursache definiert (Bewertung des Tumoransprechens erfolgte durch das IRF auf das Basis von RECIST). § Die Zeit bis zur Symptomprogression war als Zeit zwischen der Randomisierung und der ersten Symptomprogression definiert, gemessen unter Verwendung des FACT TOI-PFB - einer 24-Item Subskala, die sich aus drei Abschnitten aus dem FACT-B Fragebogen (*physical well-being, functional well-being, und additional concerns*) zusammensetzt. Eine Verminderung um fünf Punkte ist als klinisch relevant definiert und somit als Schwelle für eine Symptomprogression. **Gesundheitsbezogene Lebensqualität** (*health related quality of life* HRQoL) Die Zeit bis zur Symptomprogression wurde unter Verwendung des *Functional Assessment of Cancer Therapy-for Patientinnen mit Breast Cancer (FACT-B) questionnaire* mit der *Trial Outcomes Index-Physical/Functional/Breast (TOI-PFB)* Subskala bewertet. Weibliche Patientinnen füllten den Fragebogen alle 9 Wochen aus (innerhalb von drei Tagen vor jeder Tumorbewertung) bis zum IRF-bestimmten PD. Patientinnen sollten den Bogen immer zum geplanten Zeitpunkt ausfüllen, auch wenn die Studienmedikation aufgrund von Toxizität oder Investigator-bewertetem PD (Bewertungen wurden zum IRF-bewerteten PD gestoppt) nicht mehr gegeben wurde. Für die Bewertung wurden dieselben Kriterien angewendet, wie für die primäre Effektivitätsanalyse (*IRF-determined PFS*). Eine *Time to event* Analyse mit der Kaplan

Item[a]	Charakteristikum	Studieninformation
		Meier Methode wurde zur Testung der Hypothese, dass die Zugabe von Pertuzumab zum Behandlungsregime keinen bedeutenden Einfluss auf die HRQoL hat, durchgeführt. Die **Sicherheitsanalyse** umfasste: § Inzidenz und Schweregrad von UE und schwerwiegenden UE (SUE) (UE >= Grad 3, UE Grad 3, UE Grad 4 und UE Grad 5) § Behandlungsabbruch wegen UE § LVEF Messwerte über den Verlauf der Studie. § UE von besonderem Interesse (siehe **Fehler! Verweisquelle konnte nicht gefunden werden.**) Wurde die Progression durch das IRF bestätigt, wurde dem Investigator mitgeteilt, dass für die betroffene Patientin keine Tumorbewertungen für die Studie mehr erforderlich waren. Wurde die Progression nicht bestätigt, wurde der Investigator informiert, dass eine Fortsetzung der Tumorbewertung alle 9 Wochen im Rahmen der Studie erforderlich war. Unabhängig von diesem Vorgang konnte der Investigator frei entscheiden, ob der betroffene Patient mit einer alternativen antikanzerösen Behandlung beginnen sollte, entsprechend der klinischen Bewertung. **Erhebungszeitpunkte:** Plan siehe unterhalb dieser Tabelle.
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	Die geplante Studie wurde am 17.April 2007 mit der FDA diskutiert sowie am 2. Juli 2007 mit der European Medicines Agency (EMA), woraus sich die folgenden wesentlichen Änderungen ergaben, die im September 2007 in das Studienprotokoll integriert wurden: Die clinical benefit rate wurde als sekundärer Endpunkt fallen gelassen. Das primäre Endpunkt wurde genauer definiert, siehe auch unter Punkt 3b diese Tabelle. Andere Änderungen der Zielkriterien gab es nicht.
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	Die primäre Analyse des PFS war nach Auftreten von ungefähr 318 IRF-bewerteten PFS-Ereignissen geplant. Es wurde geschätzt, dass insgesamt 318 IRF-bewertete PFS-Ereignisse eine 80%ige Power haben, um eine 33%ige Verbesserung des Medians des PFS zu zeigen (Hazard Ratio (HR) von 0,75 bei einem zweiseitigen Signifikanzlevel von 5 %). Für das Design der Studie wurde geschätzt, dass der Median für das PFS 10,5 Monate beträgt und sich auf 14 Monate durch die Zugabe von Pertuzumab verbessert, unter der Annahmen, dass das PFS exponentiell verteilt ist. Die geplante Studie wurde am 17.April 2007 mit der FDA diskutiert sowie am 2. Juli 2007 mit der European Medicines Agency (EMA), woraus sich die folgenden wesentlichen Änderungen ergaben, die im September 2007 in das Studienprotokoll integriert wurden: § Die Fallzahl wurde von 600 auf 800 Patienten erhöht. § Die Erhöhung der Fallzahl basierte auf zwei Aspekten: 1) eine Verbesserung des Gesamtüberlebens (OS) von 33 % mit einer Power von 80 % bei einem zweiseitigen 5 % Signifikanzniveau aufzudecken, und 2) die Folgezeit zwischen finalem PFS und finalen OS Daten zu reduzieren, so dass ungefähr 50 % der benötigten Überlebensrate zum Zeitpunkt der primären

Item[a]	Charakteristikum	Studieninformation
		Effektivitätsanalyse des PFS aufgetreten sein würden.
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	Eine formale Sicherheits-Interim-Analyse war nach der Aufnahme von 100 Patienten in die Studie und einer Mindestfolgebeobachtungszeit von vier Monaten geplant (und wurde auch durchgeführt). Die Analyse wurden dem *independent Data Monitoring Committee* (DMC) vorgestellt, während der Sponsor verblindet blieb. Die primäre Analyse des PFS (basierend auf der Bewertung durch das IRF) war nach Auftreten von ungefähr 318 IRF-bewerteten PFS-Ereignissen geplant (korrespondierend zu etwa 448 Investigator-bewerteten Ereignissen). Eine Interimanalyse für das OS war für den Zeitpunkt der primären Analyse des PFS geplant. Um diese Interimanalyse des OS durchzuführen, wurden eine Lan deMets – *spending function* mit der O'Brien Fleming Grenzfunktion angewendet. Laut Protokoll waren bis zu diesem Zeitpunkt etwa 50 % der insgesamt benötigten beobachteten Todesfälle (193 Todesfälle) aufgetreten (unter der Annahme, dass das Alpha-Level für die erste OS-Analyse 0,0031 betragen würde). **Statistische Power für die finale PFS-Analyse:** Effect Size: Power for Log-Rank Test of PFS 40% improvement in PFS: 90% 33% improvement in PFS: 80% Die finale Analyse des OS sollte nach dem Auftreten von 385 Todesfällen erfolgen, was einer 80%igen Power zur Detektion einer 33%igen Verbesserung des OS (Median OS für die Kontrollgruppe wurde mit 36 Monaten angenommen) entspricht.
8	Randomisierung, Erzeugung der Behandlungsfolge	
8a	Methode zur Generierung der zufälligen Zuteilung	Um die Ausgeglichenheit zwischen den Behandlungsarmen zu gewährleisen, wurde innerhalb jeden Stratums ein komplettes Block-Randomisierungsschema angewendet.
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	Um die Ausgeglichenheit zwischen den Behandlungsarmen zu gewährleisen, wurde innerhalb jeden Stratums ein komplettes Block-Randomisierungsschema angewendet. Strata waren: vorheriger Behandlungsstatus (de novo vs vorherige adjuvante oder neoadjuvante Therapie) und Region (Europa, Nordamerika, Süd-Amerika und Asien).
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	Ein *Interactive Voice Response System* (IVRS) wurde zur Sammlung der Screening-Informationen und zur Randomisierung der geeigneten Patienten im Verhältnis 1:1 auf einen der beiden Behandlungsarme verwendet.
10	Randomisierung,	Ein *Interactive Voice Response System* (IVRS) wurde zur Sammlung der

Item[a]	Charakteristikum	Studieninformation
	Durchführung Wer führte die Zuteilung durch, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	Screening-Informationen und zur Randomisierung der geeigneten Patienten im Verhältnis 1:1 auf einen der beiden Behandlungsarme verwendet.
11	Verblindung	
11a	Waren a) die Probanden/Patienten und/oder b) diejenigen, die die Intervention/Behandlung durchführten, und/oder c) diejenigen, die die Zielgrößen bewerteten, verblindet oder nicht verblindet, wie wurde die Verblindung vorgenommen?	Patienten und Ärzte waren in dieser Studie verblindet. Eine Entblindung war während der Studie nicht erlaubt, außer bei Sicherheitsbedenken, die während der Studienbehandlung auftraten. Eine Absicherung jeder Entblindung durch den medizinischen Monitor des Sponsors war dafür erforderlich.
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Die Formulierung von Pertuzumab als Placebo war der von Pertuzumab äquivalent, ohne die aktive Substanz.
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	**Effektivitätsanalyse** Effektivitätsanalysen basieren auf der ITT-Population mit den Patienten in dem Behandlungsarm, auf den sie randomisiert waren. Die statistischen Tests, die für jeden einzelnen Endpunkt verwendet wurden, werden in der folgenden Tabelle gezeigt: **Table 5 Statistical Analyses of Efficacy Endpoints** **Variable** / **Test** / **Stratification*** **Primary endpoint:** IRF-assessed PFS — Log-rank — prior treatment status, region **Secondary endpoints:** time-to-event: investigator-assessed PFS, OS, time to symptom progression, duration of response — Cox regression — prior treatment status, region objective response rate — Mantel–Haenszel χ^2 — prior treatment status, region objective response rate — Fisher's exact — unadjusted (sensitivity) * Strata: Prior treatment status: de novo vs prior (neo)adjuvant therapy; Region: Europe, North America, South America, and Asia. Die folgende fixierte Sequenz der Testhierarchie wurde für die primäre PFS Analyse und Adjustierung für multiples statistisches Testen für das IRF-bewertete PFS, OS und ORR zum Zweck des konfirmatorischen statistischen Testens festgelegt: 1. Test für den primären Endpunkt, IRF-bewertetes PFS, auf dem zweiseitigen 5 % Signifikanzlevel. Im Falle eines positiven Ergebnisses, wurde Schritt 2 durchgeführt; andernfalls wurde der Test gestoppt. 2. Test des OS mit einem zweiseitigen 5 % Signifikanzlevel. Bei positivem Ergebnis wurde mit Schritt 3 fortgefahren, andernfalls wurde der Test gestoppt. 3. Test des ORR auf dem zweiseitigen 5 % Signifikanzlevel.

Item[a]	Charakteristikum	Studieninformation
		Primärer Endpunkt Der primäre Endpunkt PFS war das IRF-bewertete PFs. Der Log-Rank Test, stratifiziert nach dem vorherigen Behandlungsstatus (de novo und vorherige adjuvante oder neoadjuvante Therapie) und Region (Europa, Nordamerika, Süd-Amerika, und Asien), wurde für den Vergleich des PFS zwischen den Behandlungsarmen angewendet. Die Ergebnisse des unstratifizierten Log-Rank Tests wurden auch für eine Sensitivitätsanalyse verwendet. Die Kaplan-Meier Methode wurde zur Bewertung des Medians für das PFS für jeden Behandlungsarm und das Cox Proportional Hazard Model, stratifiziert nach vorherigem Behandlungsstatus und Region, wurde zur Berechnung des HR zwischen den beiden Behandlungsarmen (die Größe des Behandlungseffekts) und des 95% Konfidenzintervalls (KI) verwendet. Die Analysen wurden in prädefinierten demografischen Subgruppen durchgeführt, vorausgesetzt, es ergab sich eine sinnvolle Gruppengröße. Univariate und multivariate Cox-Regressionsanalysen wurden ebenfalls durchgeführt, um die Beziehung zwischen der prädefinierten Stratifikation und der prognostischen Kovariate zu Studienbeginn zu analysieren. **Sekundäre Endpunkte** Gesamtüberleben (*Overall survival*): Die Analysemethoden waren die gleichen, wie für den primären Endpunkt. PFS Investigator-bewertet: Daten von Patienten ohne dokumentiertes PD bzw. von Patienten, die nicht innerhalb von 18 Wochen nach der letzten Tumorbewertung verstarben, wurden zum Zeitpunkt der letzten Investigator-Beurteilung zensiert (oder, wenn nach Studienbeginn keine Tumorbewertung stattgefunden hatte, mit den Ergebnissen der Visite an Tag 1). Als Analysemethoden wurden die gleichen wie für den primären Endpunkt verwendet. Gesamtansprechrate (*overall response rate*): Nur Patienten, die eine messbare Krankheit (bestimmt durch das IRF) zu Studienbeginn hatten, wurden in die Analyse für das Gesamtansprechen eingeschlossen. Gesamtansprechen basierte auf dem besten Gesamtansprechen, das nach Studienbeginn auftrat bis zur IRF-Feststellung eines PD, eines Todesfalles oder der ersten Gabe einer NACT, je nachdem, welches Ereignis zuerst auftrat. Patienten, die nach Studienbeginn keine Tumorbewertung hatten, wurden als *non-responders* eingestuft. Eine Schätzung für die *overall response rate* und dessen 95% KI wurde für jeden Behandlungsarm kalkuliert. Der Mantel-Haenszel Test wurde nach dem vorherigen Behandlungsstatus und Region stratifiziert und für den Vergleich der *objective response rate* zwischen den beiden Behandlungsarms angewendet. Ein nicht-adjustierter exakter Fisher's Test wurde als Sensitivitätsanalyse durchgeführt. Eine weitere Sensitivitätsanalyse für das Investigator-bewertete objektive Ansprechen basierte auf Patienten mit Investigator-bewerteter messbarer Krankheit zu Studienbeginn. Dauer der *objective response*: Dauer der *objective response* basierte auf der IRF-Bewertung. Eine formale Hypothesen-Testung wurde nicht durchgeführt, da die Subgruppe der Patienten mit *objective response* kein randomisiertes Subset ist. Die mediane Dauer für die *objective response* für jeden Behandlungsarm wurde mittels Kaplan-Meier Analyse ermittelt. Die Hazard Ratio für den Vergleich zwischen den beiden Behandlungsarmen wurde mittels Cox Regression ermittelt. Im Rahmen einer Sensitivitätsanalyse, wurde die Dauer des objektiven Ansprechens, basierend auf der Bewertung des Investigators analysiert.

Item[a]	Charakteristikum	Studieninformation
		Zeit bis zur Symptomprogression: siehe Analyse QoL **Gesundheitsbezogene Lebensqualität** Die Zeit bis zur Symptomprogression wurde unter Verwendung des Fragebogens *Functional Assessment of Cancer Therapy-f*ür Patienten mit *Breast Cancer* (FACT-B) mit der Subskala *Trial Outcomes Index-Physical/Functional/Breast* (TOI-PFB) bewertet. Weibliche Patientinnen füllten den Fragebogen alle 9 Wochen (innerhalb von drei Tagen vor der jeweilig anstehenden Tumorbewertung) bis zu einer vom IRF-bewerteten PD aus. Die Patientinnen füllten den Bogen auch dann aus, wenn die Studientherapie aufgrund von Toxizität oder vom Investigator-bewertetem PD (Bewertungen wurden gestoppt, sobald ein IRF-bewertetes PD vorlag) nicht mehr verabreicht wurde. Die Zeit bis zum Ereignis wurde mittels Kaplan Meier Methode getestet. Die Hypothese lautete, dass die Zugabe von Pertuzumab zum Behandlung regime keinen bedeutenden Einfluss auf die HRQoL hat. **Sicherheitsanalyse** Patientinnen, die in irgendeiner Form etwas von der Studienmedikation erhalten hatten, wurden in die Sicherheitsanalyse eingeschlossen (*safety analysis population* - SAP). Sicherheitsergebnisse werden für den Behandlungsarm ausgewiesen, in dem die Patientinnen die Behandlung erhielten. **Unerwünschte Ereignisse und schwerwiegende unerwünschte Ereignisse (inkl. UE>=Grad 3, UE=Grad3, 4 und 5)** Die Inzidenz von UE, inklusive SUE, wurde pro Behandlungsarm dargestellt. Nur UE, die bis zu 42 Tage nach der letzten Dosis der Studienbehandlung auftraten, wurden in den Tabellen dargestellt. In den Listen aller UE (*listing*) sind alle UE dargestellt. Separate zusammenfassende Tabellen wurden für schwerwiegende unerwünschte Ereignisse, UE mit Todesfolge, UE mit der Folge des Studienabbruchs, UE mit Studienabbruch (Pertuzumab/Placebo, Trastuzumab oder Docetaxel) und UE-bedingte Dosisreduktion oder Unterbrechungen erstellt Falls ein Patient mehrere UE hatte, wurde nur die schwerste Episode oder die Episode mit dem stärksten kausalen Zusammenhang gezählt. Alle UE wurden kodiert und in einem System nach Organklasse eingeordnet und nach absteigender Häufigkeit dargestellt. Die Inzidenz der UE wurde ebenfalls nach Schweregrad (NCI-CTCAE Grad) und der Beziehung zur Studienmedikation tabellarisch aufgelistet. **ECOG Score** Die Veränderung des ECOG Scores über die Zeit wurde pro Behandlungsarm dokumentiert. Die Veränderungen wurden einer der folgenden Kategorien zugewiesen: § Verbessert: der Score verminderte sich nach Studienbeginn § Unverändert: der Score behielt denselben Wert, wie zu Studienbeginn § Verschlechtert: der Score erhöhte sich nach Studienbeginn
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	Im CSR präspezifiziert waren folgende Subgruppen: § vorausgehender Behandlungsstatus (De novo; adjuv. od. neoadjuvante Behandlung) § Region (Europa; Nordamerika, Südamerika; Asien) § Altersgruppe (<65; ≥65), Altersgruppe (<75; ≥75) § Rasse (Kaukasisch; Afro-amerikanisch; Asiatisch; Weitere) § Art der Erkrankung (Viszeral; Nicht-viszeral)

Item[a]	Charakteristikum	Studieninformation
		§ IHC-HER2-Status (3+) § FISH-Status (positiv) § ER/PgR-Status (ER und/oder PgR Positiv; ER und PgR Negativ, Unbekannt) Zusätzlich wurden folgende Subgruppen-Analysen durchgeführt: § IHC-HER2-Status (2+) § Baseline ECOG (0; 1+) § Dosiseskalation mit Docetaxel auf 100 mg/m2 (Ja; Nein) Im Rahmen des G-BA Beratungsgesprächs wurden folgende weitere Subgruppenanalysen bestimmt: § Vorherige (neoajduvante oder adjuvante) Behandlung mit Trastuzumab (Ja; Nein) § Vorherige (neo-)adjuvante Taxanbehandlung (Ja; Nein) § Lokalrezidiv (Ja; Nein) Für das PFS waren sechs **Sensitivitätsanalysen** geplant, um einen möglichen Einfluss der folgenden Faktoren einzuschätzen: § Ein möglicher Unterschied zwischen den verschiedenen Bewertungsverfahren durch Investigator und IRF wurde analysiert. Das früheste Datum der Dokumentation einer PD durch das IRF bzw. den Investigator wurde als Datum für das Auftreten eines PFS-Ereignisses zu Grunde gelegt. § Zensierung zum Zeitpunkt des Beginns der nächsten Linie der anti-tumorösen Therapie (next-line anti-cancer therapy (NACT)): Patientinnen, die mit einer NACT vor der IRF-Bewertung einer Progression (innerhalb von 18 Wochen nach der letzten Tumorbewertung) bzw. der letzten IRF-verfügbaren Tumorbewertung begannen, wurden zum Datum der letzten verfügbaren IRF-Tumorbewertung vor dem Start einer NACT zensiert. § Einfluss der Bewertung des PFS durch das IRF auf die Behandlung. Nur IRF-bewertete PFS Ereignisse, die nicht später als 42 Tage nach der letzten Administration jeglicher Studienmedikation stattfanden, wurden in die Analyse einbezogen. § Potentieller Bias, der durch unterschiedliche Zeitintervalle zwischen den Tumorbewertungen aufgrund verpasster Visiten auftreten könnte: Fehlende Bewertungen (oder die früheste fehlende Bewertung im Fall von mehreren fehlenden Bewertungen) wurde durch eine Bewertung des PD ersetzt und die Zeit bis zum Ereignis wurde auf den Tag der erwarteten fehlenden Bewertung gesetzt. Fehlte keine Visite, wurde das erste PD-Datum, das vom IRF bewertet wurde, als Datum für das Ereignis gewertet. Im Fall eines Todes vor einer PD innerhalb von 18 Wochen nach der letzten Tumorbewertung, wo vor dem Todesfall die letzte Visite fehlte, wurde der Zeitpunkt dieser fehlenden Visite als Zeitpunkt für das Ereignis gewertet. § Zeitpunkt des Todes, inklusive aller Todesfälle, die mehr als 18 Wochen nach der letzten Tumorbewertung auftraten. § Kontrolle für Behandlungsabbruch aufgrund von Toxizität: Patienten, die sämtliche Studienbehandlung aufgrund Toxizität abbrachen, wurden vom Zeitpunkt der letzten IRF-Bewertung an oder vor dem Beginn des Behandlungsabbruchs zensiert. Alle Sensitivitätsanalysen basierten auf der ITT-Population und waren

Item[a]	Charakteristikum	Studieninformation
		nach vorherigem Behandlungsstatus und Region stratifiziert. Zusätzlich wurde für das Gesamtüberleben ebenfalls eine Sensitivitätsanalyse durchgeführt: § Im Rahmen einer Sensitivitätsanalyse wurden die Gesamtüberlebenskurven unter Verwendung des Wilcoxon Tests berechnet.
-	**Resultate**	
13	Patientenfluss (inklusive Flow-Chart zur Veranschaulichung im Anschluss an die Tabelle)	
13a	Anzahl der Studienteilnehmer für jede durch Randomisierung gebildete Behandlungsgruppe, die a) randomisiert wurden, b) tatsächlich die geplante Behandlung/Intervention erhalten haben, c) in der Analyse des primären Zielkriteriums berücksichtigt wurden	Insgesamt wurden 1196 Patienten für die Studie gescreent, Patienten wurden von 204 Zentren in 25 Ländern aufgenommen. Insgesamt wurden 808 Patienten auf einen der beiden Behandlungsarme randomisiert – Placebo und Trastuzumab mit Docetaxel (Pla+T+D, 406 Patienten) oder Pertuzumab und Trastuzumab mit Docetaxel (Ptz+T+D, 402 Patienten). Von den randomisierten Patienten, zwei im Pla+T+D Behandlungsarm und zwei im Ptz+T+D Behandlungsarm, erhielten vier keine Studienbehandlung. Drei von diesen Patienten hatten ALT und AST Werte die größer als 2,5 x ULN (ein Bruch von Ausschlusskriterium Nr.14). Der vierte Patient zog seine Einverständniserklärung vor der ersten Gabe der Studienmedikation zurück. Zum Zeitpunkt des klinischen Datenschnitts *(Mai 2011)*, waren 121 Patienten (29,8 %) in dem Pla+T+D Behandlungsarm und 171 Patienten (42,5 %) in dem Ptz+T+D Behandlungsarm noch lebend und unter Behandlung. Weitere 166 Patienten (40,9 %) im Pla+T+D und 144 Patienten (35,8 %) im Ptz+T+D Behandlungsarm lebten und befanden sich in der Folgebeobachtungsperiode für das Überleben (*survival follow-up period)*. 41 Patienten hatten entweder das Einverständnis zurückgezogen oder waren während dieser Periode *lost to follow-up*; 23 Patienten (5,7 %) im Pla+T+D Behandlungsarm vs 18 Patienten (4,5 %) im Ptz+T+D Behandlungsarm. Die verbleibenden 165 Patienten waren verstorben (96 Patienten [23,6 %] im Pla+T+D Behandlungsarm vs 69 Patienten [17,2 %] im Ptz+T+D Behandlungsarm. Zum Zeitpunkt des klinischen Datenschnitts, betrug der Median für die Zeit der Studienteilnahme, inklusive der nachfolgenden Behandlung, 73,1 Wochen im Pla+T+D Behandlungsarm und 77,1 Wochen im Ptz+T+D Behandlungsarm. Zum Zeitpunkt des 2. Datenschnitts (Mai 2012) wurden für die Patienten folgende Daten angegeben:

Item[a]	Charakteristikum	Studieninformation
		Table 5 Duration of Patient Time on Study Pla+T+D / Ptz+T+D Cutoff Date: 13 May 2011 / 14 May 2012 / 13 May 2011 / 14 May 2012 n[a]: 397 / 396 / 407 / 408 Overall time on study treatment (weeks) Median: 47.0 / 49.3 / 57.1 / 75.7 Range: 0.3–150.3 / 0.3–197.0 / 0.6–165.1 / 0.6–201.1 Overall time on study including post-treatment follow-up (weeks) Median: 73.1 / 105.9 / 77.1 / 117.1 Range: 0.4–165.3 / 0.4–207.9 / 0.7–165.3 / 0.7–207.9 Overall time on study treatment is calculated from date of first dose to last known date in study treatment period. Overall time on study, including post-treatment follow-up, is calculated from date of randomization to last known date. [a] One additional patient randomized to placebo received one dose of pertuzumab in error since the first analysis (see Section 3.2 for details). Thus, the number of patients included in the safety analysis population has changed. Derived from: t_ufup_s (13 May 2011) and page 298 (14 May 2012) Für die Auswertung der Studie wurden zwei Analysepopulationen definiert: die Intent-to-Treat (ITT)-Population und die Safety-Analysis-Population (SAP): Die ITT Population beinhaltet alle 808 randomisierten Patienten (406 Pla+T+D, 402 Ptz+T+D). Die SAP bezieht sich nur auf Patienten, welche mindestens eine Dosis von einer der Komponenten der Studienbehandlung erhalten hatten und beinhaltet 804 Patienten von denen zwei im jeweiligen Behandlungsarm keine Studienmedikation nach der Randomisierung erhalten hatten. Acht der randomisierten Patienten, welche Placebo erhalten sollten, hatten tatsächlich mindestens eine Dosis Pertuzumab während der Studie erhalten. Für die Safety Analyse wurden diese Patienten in den Ptz+T+D Arm mit einbezogen. Zusätzlich erhielt ein randomisierter Patient, der auf den Behandlungsarm mit Pertuzumab randomisiert war, in jedem Zyklus Placebo. Dieser Patient wurde in den Pla+T+D Arm mit einbezogen. Folglich beinhaltet die SAP 397 Patienten im Pla+T+D Arm und 407 Patienten im Ptz+T+D Arm. Seit der Durchführung der 1. Analyse (13.5.2011), erhielt ein weiterer Patient aus der Placebo-Gruppe versehentlich Pertuzumab, und wurde gemäß. Analyseplan der Safety-Population zugeordnet, so dass die Anzahl Patienten in der Safety Analyse sich änderte: Der betroffene Patient erhielt im Zyklus 33 (April 2012) Pertuzumab, im folgenden Zyklus 34 (kurz vor dem Datenschnitt) wurde wieder Placebo verabreicht. Daher wurde die Safety Analyse für den zweiten Datenschnitt mit 396 Patienten im Pla+T+D Arm und mit 408 Patienten im Ptz+T+D Arm gerechnet.
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	Auswertung Mai 2011: Insgesamt brachen 279 Patienten (70 %) im Pla+T+D Behandlungsarm und 233 Patienten (57 %) unter Ptz+T+D sämtliche Studienbehandlungen bis zum Zeitpunkt des klinischen Datenschnitts ab. Die meisten der vorzeitigen Behandlungsabbrüche erfolgten auf Grund einer PD (227 Patienten [57,2 %] im Pla+T+D Behandlungsarm, 180 Patienten [44,2 %] im Ptz+T+D Behandlungsarm). Die Inzidenz eines frühen Behandlungsabbruchs aus anderen Gründen als einer PD war niedrig, mit 13 % pro Behandlungsarm (52 Patienten im Pla+T+D Behandlungsarm vs. 53 Patienten im Ptz+T+D Behandlungsarm). Ein vergleichbarer Prozentsatz pro Behandlungsarm brach aufgrund Sicherheitsaspekten die Studienbehandlung ab (31

Item[a]	Charakteristikum	Studieninformation
		Patienten [7,8 %] im Pla+T+D vs. 30 Patienten (7,4 %) im Ptz+T+D Behandlungsarm). Zum Zeitpunkt des clinical Datenschnitts hatten die Patienten 47 Wochen (Median) (Pla+T+D Behandlungsarm) bzw. 57 Wochen (Ptz+T+D Behandlungsarm) die Studienbehandlung erhalten. Patienten sollten Docetaxel für minimal 6 Zyklen erhalten, wonach der Investigator die Option hatte, die Therapie zu beenden oder fortzusetzen, während die anderen Antikörper-Komponenten der Behandlung fortgesetzt werden konnten (Placebo/Pertuzumab und Trastuzumab). Die Docetaxeltherapie wurde bei 255 Patienten (64,2 %) im Pla+T+D Behandlungsarm und 298 Patienten (73,2 %) im Ptz+T+D Behandlungsarm beendet, während der Patient noch Pertuzumab/Placebo und Trastuzumab erhielt. Neunzig Patienten (23 %) im Pla+T+D Behandlungsarm und 97 Patienten (24 %) im Ptz+T+D Behandlungsarm stoppten Docetaxel frühzeitig aufgrund eines UE. Etwa 40 % der Patienten pro Behandlungsarm beendeten Docetaxel vor der Beendigung der Therapie mit anderen Studiensubstanzen aufgrund administrativer Gründe, wie z.B. Praxisstandards oder weil der Investigator eine adäquate Behandlung mit Docetaxel in Betracht gezogen hatte. <u>Auswertung Mai 2012:</u> Die Abbruchraten für die Studienbehandlung wurden mit den demografischen Daten und den Baseline Charakteristika verglichen, um feststellen, ob bestimmte Baseline-Charakteristika einen Vorhersagewert hatten. Es wurden keine Charakteristika festgestellt, die in einer klaren Korrelation zum Behandlungsabbruch standen. Trotz unterschiedlicher Toxizitätsprofile zwischen asiatischen und anderen Patienten, war die Abbruchrate unter Asiaten nicht erhöht.
14	Aufnahme/Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	Aufnahme des ersten Patienten: 12.02.2008 Studienphase 1: 2 / 2008 – 5 / 2011 Studienphase 2: 6 / 2011 – 5 / 2012 Studienphase 3: 06 / 2012 – Q3 / 4 / 2013 Abschluss der Datenbank für den ersten Datenschnitt: 13.05.2011 Abschluss der Datenbank für den zweiten Datenschnitt: 05.2012
14b	Informationen, warum die Studie endete oder beendet wurde	Die Studie wird wie geplant durchgeführt und befindet sich seit 6 / 2012 derzeit in der Studienperiode 3, welche bis zum Auftreten von 385 Ereignissen (Gesamtüberleben) – voraussichtlich Q3 / 4 / 2013 – weitergeführt wird. Die konformatorischen finalen Analysen für den primären Endpunkt und die Gesamtansprechrate (05 / 2011) sowie das Gesamtüberleben (05 / 2012) wurden bereits durchgeführt.

Stellen Sie den Patientenfluss in einem Flow-Chart je Studie dar.

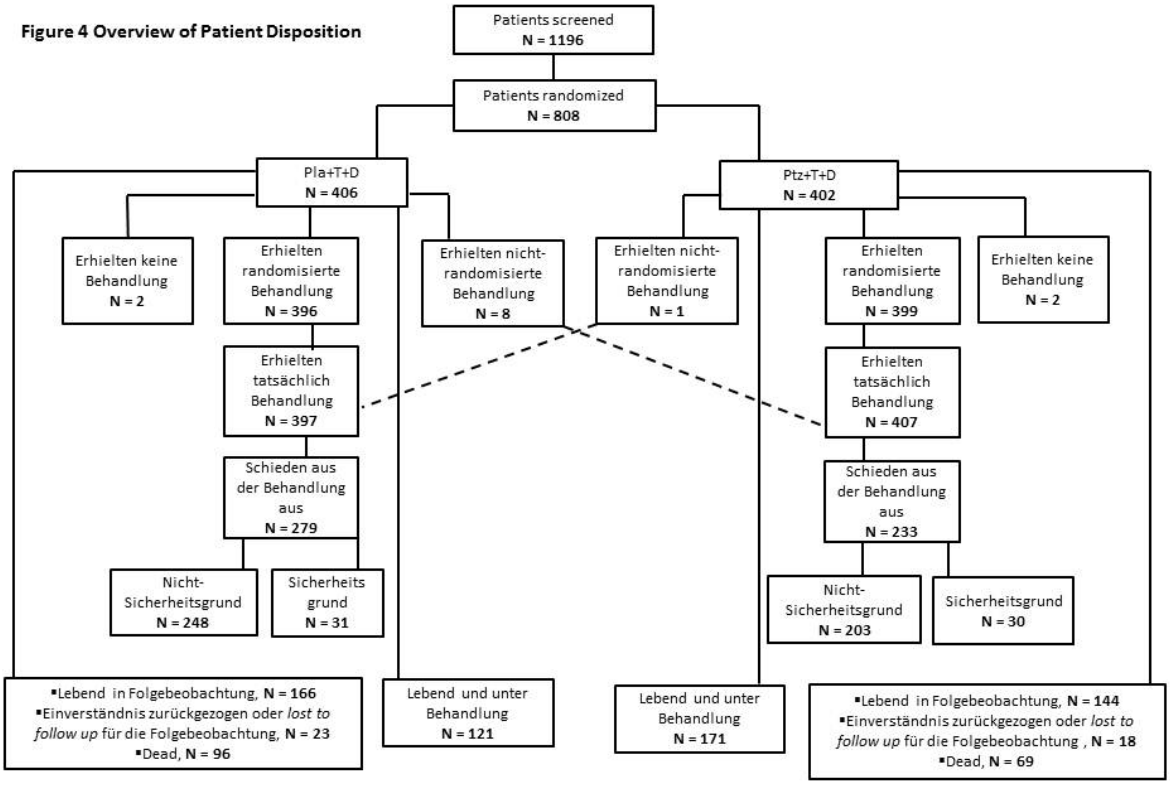


Abbildung 4: Patientenfluss in der Studie, zweiter Datenschnitt

Table 6 Patient Withdrawal from Treatment

Snapshot Date: 01JUN2012 Clinical Cut-Off Date: 14MAY2012

Reason for Withdrawal	Placebo + Trastuzumab + Docetaxel N = 396 No. (%)	Pertuzumab + Trastuzumab + Docetaxel N = 408 No. (%)
Safety	36 (9)	37 (9)
Abnormality of Laboratory Test	0	0
Adverse Event(a)	23	30
Death	13	7
Non-Safety	296 (75)	263 (64)
Insufficient Therapeutic Response	271	235
Early Improvement	0	0
Violation of Selection Criteria at Entry	1	2
Other Protocol Violation	1	0
Refused Treatment(b)	21	18
Failure to Return	1	3
Other	1	5
Total	332 (84)	300 (74)

(a)=Including intercurrent illness (b)=Including 'did not co-operate', 'withdrew consent'
Percentages are based on N.
'Abnormality of Laboratory Test' and 'Early Improvement' not included as eCRF options, so not applicable in this study.
EX11 17JUN2012:11:05:11 (1 of 1)

Source: t_ex11 page 299

Abbildung 5: Angabe von Gründen für Patienten, die die Studie beendet hatten bis Mai 2012

Tabelle 4-89 (Anhang): Erhebungszeitpunkte für die einzelnen Befunde während des Verlaufs der Studie CLEOPATRA (Studienbericht der Studie CLEOPATRA, 2011)

	Screening / Baseline		Behandlungsperiode			Nachbeobachtungsphase			
			Jeder Zyklus (Zyklus=21 Tag		Jeder 3 Zyklus	Behandlung Visite bei Studienabbruch	Week 18 Nach Behandlung Visite bei Studienabbruch	Jede 18. Woche nach Behandlung Visite bei Studien-abbruch	Bis zu 3 Jahre nach Behandlung Visite bei Studien-abbruch
Tag	D-28 to -1	D-7 to -1	D1	D8		Tag 28-42 nach Behandlung	Tag 126 nach Behandlung Visite bei Studienabbruch	Jeder 126 Tags nach Behandlung Visite bei Studien-abbruch	
Einverständniserklärung	x								
Komplette medizinische Anamnese, inkl. Demografie	x								
Review von Ein- und Ausschlusskriterien	x								
Komplette physikalische Untersuchung und Vitale Zeichen	x								
Symptom- bezogene physikalische Untersuchung und vitale Zeichen			x			x			
Elektrokardiogramm (ECG)	x		Alle 9 Wochen zum Zeitpunkt des LVEF			x			
Thorax-Röntgen	x		Wenn klinisch indiziert			x	Wenn klinisch indiziert		
ECOG Leistungs-Status	x		x			x	Alle 9 Wochen zum Zeitpunkt jeder Tumorbewertung		
Fact-B- Quality of Life (nur Frauen)		x	Alle 9 Wochen innerhalb von 3 Tagen vor jeder Tumorbewertung						
Tumor Beurteilung	x		Alle 9 Wochen nach Randomisierung bis zur IRF-bestätigten PD.						
LVEF (ECHO oder MUGA)	x		Alle 9 Wochen nach Randomisierung			x	Alle 6 Monate im ersten Jahr anschließend jährlich für 3 Jahre		
Knochenscan	x		Wenn klinisch indiziert			x	Wenn klinisch indiziert bis zum IRF-bestätigten PD		
Unerwünschte Ereignisse	x		Fortlaufend				Fortlaufend		
Begleitmedikation und Krebs-bezogene Chirurgie			Fortlaufend			Fortlaufend			

	Screening / Baseline		Behandlungsperiode			Nachbeobachtungsphase			
			Jeder Zyklus (Zyklus=21 Tag		Jeder 3 Zyklus	Behandlung Visite bei Studienabbruch	Week 18 Nach Behandlung Visite bei Studienabbruch	Jede 18. Woche nach Behandlung Visite bei Studien-abbruch	Bis zu 3 Jahre nach Behandlung Visite bei Studien-abbruch
Tag	D-28 to -1	D-7 to -1	D1	D8		Tag 28-42 nach Behandlung	Tag 126 nach Behandlung Visite bei Studienabbruch	Jeder 126 Tags nach Behandlung Visite bei Studien-abbruch	
Prozeduren									
Pertuzumab / Placebo Gabe			x						
Trastuzumab Gabe			x						
Docetaxel Gabe			x						
Proben									
Tumor für HER2 Eligibility & Biomarkers, zentral Labor	x								
Hämatologie, lokales Labor		x	x	x		x			
Biochemie, lokales Labor		x	x			x			
INR und aPTT or PTT, lokales Labor		x	x						
Schwangerschaftstest, lokales Labor (wenn zutreffend)		x			x	x	3 und 6 Monate nach Behandlung Visite bei Studienabbruch		
Serum für Trastuzumab PK, zentrales Labor		x							
1. Serum für Antikörper auf Pertuzumab, zentrales Labor			*4.* Perform every 9 weeks at the time of the TA[V]			5.			
Proben für HER2 ECD& HER Ligands, zentrales Labor		x	Alle 9 Wochen zum Zeitpunkt der Tumorbewertung						
Vollblut für FCy Polymorphism (klinischer Genotyp), zentrales Labor									
Proben, die separate Einverständniserklärung benötigen									
Metastatic Tumor for Biomarkers, zentrales Labor		x							

	Screening / Baseline		Behandlungsperiode			Nachbeobachtungsphase			
			Jeder Zyklus (Zyklus=21 Tag		Jeder 3 Zyklus	Behandlung Visite bei Studienabbruch	Week 18 Nach Behandlung Visite bei Studienabbruch	Jede 18. Woche nach Behandlung Visite bei Studien-abbruch	Bis zu 3 Jahre nach Behandlung Visite bei Studien-abbruch
Tag	D-28 to -1	D-7 to -1	D1	D8		Tag 28-42 nach Behandlung	Tag 126 nach Behandlung Visite bei Studienabbruch	Jeder 126 Tags nach Behandlung Visite bei Studien-abbruch	
Serum & Plasma Biomarker Sample Repository (BSR), zentrales Labor		x	Every 9 weeks at the time of each tumor assessment (until at least 18 weeks post-treatment)						
Dokumentation der nach-Studien Behandlung Krebs-bezogene medizinische oder chirurgische Prozeduren und Therapien								x	
Überlebensinformation							x	x	

Anhang 4-G: Bewertungsbögen zur Einschätzung von Verzerrungsaspekten

Beschreiben Sie nachfolgend die Verzerrungsaspekte jeder eingeschlossenen Studie (einschließlich der Beschreibung für jeden berücksichtigten Endpunkt). Erstellen Sie hierfür je Studie eine separate Version des nachfolgend dargestellten Bewertungsbogens.

Tabelle 4-90 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie <CLEOPATRA (WO20698)>

Grundsätzliche Hinweise zum Ausfüllen des Bewertungsbogens

Dieser Bewertungsbogen dient der Dokumentation der Einstufung des Potenzials der Ergebnisse für Verzerrungen (Bias). Für jede Studie soll aus diesem Bogen nachvollziehbar hervorgehen, inwieweit die Ergebnisse für die einzelnen Endpunkte als möglicherweise verzerrt bewertet wurden, was die Gründe für die Bewertung waren und welche Informationen aus den Quellen dafür Berücksichtigung fanden.

Der Bogen gliedert sich in zwei Teile:

- **Kriterien zur Bewertung des Verzerrungspotenzials auf Studienebene**. In diesem Teil sind die endpunktübergreifenden Kriterien aufgelistet.
- **Kriterien zur Bewertung des endpunktspezifischen Verzerrungspotenzials.** In diesem Teil sind die Kriterien aufgelistet, die für jeden Endpunkt separat zu prüfen sind.

Für jedes Kriterium sind unter „Angaben zum Kriterium“ alle relevanten Angaben aus den Quellen zur Bewertung einzutragen (Stichworte reichen ggf., auf sehr umfangreiche Informationen in den Quellen kann verwiesen werden).

Grundsätzlich werden die Bögen studienbezogen ausgefüllt. Wenn mehrere Quellen zu einer Studie vorhanden sind, müssen die herangezogenen Quellen in der folgenden Tabelle genannt und jeweils mit Kürzeln (z. B. A, B, C, …) versehen werden. Quellenspezifische Angaben im weiteren Verlauf sind mit dem jeweiligen Kürzel zu kennzeichnen.

Studie: CLEOPATRA-Studie (WO20698)

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genaue Benennung der Quelle	**Kürzel**
CLEOPATRA (WO20698). Clinical Study Report –WO20698C/TOC4129g - A Phase III, Randomized, Double-Blind, Placebo-Controlled Clinical Trial to Evaluate the Efficacy and Safety of Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel in Previously Untreated HER2-Positive Metastatic Breast Cancer – Report No. 1046288, Roche, October 2011. Sponsor’s Signatory: Hans Ulrich Burger, PhD [46]	CSR 10/2011
CLEOPATRA (WO20698). Update Clinical Study Report - WO20698C/TOC4129g - A Phase III, Randomized, Double-Blind, Placebo-Controlled Clinical Trial to Evaluate the Efficacy and Safety of Pertuzumab + Trastuzumab + Docetaxel vs. Placebo + Trastuzumab + Docetaxel in Previously Untreated HER2-Positive Metastatic Breast Cancer - Report No. 1053649, Roche, Dezember 2012. [47]	CSR 12/2012
Zusätzliche Analysen zur CLEOPATRA-Studie (WO20698-TOC4129g). 2013 [55]	Roche 2013
Baselga J et al: Pertuzumab plus Trastuzumab plus Docetaxel for Metastatic Breast Cancer, New England Journal of Medicine, 2012 [49]	Baselga 2012

Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse:

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A: Kriterien des Verzerrungspotenzials der Ergebnisse auf Studienebene:

Einstufung als randomisierte Studie

☒**ja** à Bewertung der Punkte 1 und 2 für randomisierte Studien

☐**nein**: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anhaltspunkte vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.
à Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

In der randomisierten, kontrollierten, doppelt verblindeten CLEOPATRA-Studie wurden die Patienten im Verhältnis 1:1 auf die Studienarme randomisiert. als Strata fanden der vorherige Behandlungsstatus (de novo vs. vorherige adjuvante oder neoadjuvante Therapie) und die Region (Europa, Nordamerika, Süd-Amerika und Asien) Eingang in die Studie. Patienten und Ärzte waren in dieser Studie verblindet. Um die Ausgeglichenheit zwischen den Behandlungsarmen zu gewährleisten, wurde innerhalb jeden Stratums ein komplettes Block-Randomisierungsschema angewendet. Die Screening-Informationen wurden mittels eines Interactive Voice Response Systems (IVRS) erfasst und die Patienten entsprechend auf die Behandlungsarme randomisiert.

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 Anhaltspunkte dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
· Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

☐**unklar**: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau.

☐**nein**: Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

3. Verblindung von Patienten und Behandlern

Patient

☒**ja:** Die Patienten waren verblindet.

☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

☐**nein:** Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

Patienten und Ärzte waren in dieser Studie verblindet. Die Verblindung der Ärzte und Patienten in den Studienperioden 1 und 2 durfte nur bei Sicherheitsbedenken, die während der Studienbehandlung auftraten, aufgehoben werden.
Die Formulierung von Pertuzumab als Placebo war der von Pertuzumab äquivalent, ohne die aktive Substanz.

Behandler bzw. Weiterbehandler

☒**ja:** Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, den primären Behandler (Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

☐**nein:** Aus den Angaben geht hervor, dass die Behandler nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

---

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:
- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter
- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:
- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*

- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*
  Zulässige Gründe sind:
  - *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
  - *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*
  - *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
- *Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.*
- *Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.*

Anzumerken ist, dass Anhaltspunkte für eine ergebnisgesteuerte Darstellung eines Endpunkts das Verzerrungspotenzial für die Ergebnisse der übrigen Endpunkte erhöhen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anhaltspunkten dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, ist das Verzerrungspotenzial für die anderen Endpunkte u. U. erhöht. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Erhöhung des Verzerrungspotenzials der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen.
Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und das Verzerrungspotenzial anderer Endpunkte nicht berührt.

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

5. Fehlen sonstiger (endpunktübergreifender) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien*
- *intransparenter Patientenfluss*
- *Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:*
  - *Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).*
  - *Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (ansonsten ggf.im Nachhinein von der Biometrie durchzuführen).*
  - *Eine Adjustierung sollte auch dann erfolgen, wenn die maximale Fallzahl erreicht wurde.*
  - *Sind weitere Endpunkte korreliert mit dem Endpunkt, dessentwegen die Studie abgebrochen wurde, so sollten diese ebenfalls adäquat adjustiert werden.*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

_____

_____

Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials der Ergebnisse erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen Punkte A.1 bis A.5. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse durch diese endpunktübergreifenden Aspekte relevant verzerrt sind.

☐ **hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

_____

_____

B: Kriterien des Verzerrungspotenzials der Ergebnisse pro Endpunkt

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: Gesamtüberleben

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

☒ **ja:** Der Endpunkt wurde verblindet erhoben.

☐ **unklar:** Es finden sich keine diesbezüglichen Angaben.

☐ **nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Endpunkterheber waren verblindet. Die Gruppenzuteilung erfolgte verblindet über ein IVRS. Die Interventionen waren vergleichbar mit Ausnahme der Inhaltsstoffe.

2. Adäquate Umsetzung des ITT-Prinzips

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten

ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set"). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja" genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.
- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Progressionsfreies Überleben bewertet durch IRF

1. **Verblindung der Endpunkterheber**

   Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
   In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

   ☒**ja:** Der Endpunkt wurde verblindet erhoben.

   ☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

   ☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

   Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Endpunkterheber waren verblindet. Die Gruppenzuteilung erfolgte verblindet über ein IVRS. Die Interventionen waren vergleichbar mit Ausnahme der Inhaltsstoffe.

2. **Adäquate Umsetzung des ITT-Prinzips**

   Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Gesamtansprechrate

1. **Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

☒**ja:** Der Endpunkt wurde verblindet erhoben.

☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Endpunkterheber waren verblindet. Die Gruppenzuteilung erfolgte verblindet über ein IVRS. Die Interventionen waren vergleichbar mit Ausnahme der Inhaltsstoffe. Alle Befunde wurden durch das IRF beurteilt, was auch die verblindete Erhebung der Endpunkte sicherstellte. Die Bewertung einer Krankheitsprogression erfolgte mittels RECIST-Kriterien.

2. **Adäquate Umsetzung des ITT-Prinzips**

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr

genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.
- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Zeit bis Symptomprogression

1. **Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

☒**ja:** Der Endpunkt wurde verblindet erhoben.

☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Endpunkterheber waren verblindet. Die Gruppenzuteilung erfolgte verblindet über ein IVRS. Die Interventionen waren vergleichbar mit Ausnahme der Inhaltsstoffe. Die Patientinnen, die den Fragebogen ausfüllten, waren verblindet, die Auswertung erfolgte ebenfalls verblindet.

2. **Adäquate Umsetzung des ITT-Prinzips**

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr

genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.
- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Zeit bis zur Verschlechterung des BSC-Scores

1. **Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

☒**ja:** Der Endpunkt wurde verblindet erhoben.

☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Endpunkterheber waren verblindet. Die Gruppenzuteilung erfolgte verblindet über ein IVRS. Die Interventionen waren vergleichbar mit Ausnahme der Inhaltsstoffe. Die Patientinnen, die den Fragebogen ausfüllten, waren verblindet, die Auswertung erfolgte ebenfalls verblindet.

2. **Adäquate Umsetzung des ITT-Prinzips**

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr

genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.
- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Patienten mit unerwünschtem Ereignis

1. **Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

☒**ja:** Der Endpunkt wurde verblindet erhoben.

☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Befragung der verblindeten Patienten sowie die Beurteilung von Krankheitszeichen und Symptomen erfolgten durch den ebenfalls verblindeten Investigator.

2. **Adäquate Umsetzung des ITT-Prinzips**

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Patienten mit schwerwiegendem unerwünschtem Ereignis (SUE)

1. **Verblindung der Endpunkterheber**

   Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
   In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

   ☒**ja:** Der Endpunkt wurde verblindet erhoben.

   ☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

   ☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

   Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Befragung der verblindeten Patienten sowie die Beurteilung von Krankheitszeichen und Symptomen erfolgten durch den ebenfalls verblindeten Investigator.

2. **Adäquate Umsetzung des ITT-Prinzips**

   Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Patienten mit Behandlungsabbruch wegen UE

1. **Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

☒**ja:** Der Endpunkt wurde verblindet erhoben.

☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Befragung der verblindeten Patienten sowie die Beurteilung von Krankheitszeichen und Symptomen erfolgten durch den ebenfalls verblindeten Investigator.

2. **Adäquate Umsetzung des ITT-Prinzips**

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Patienten mit UE ≥ Grad 3

1. **Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

☒**ja:** Der Endpunkt wurde verblindet erhoben.

☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Befragung der verblindeten Patienten sowie die Beurteilung von Krankheitszeichen und Symptomen erfolgten durch den ebenfalls verblindeten Investigator.

2. **Adäquate Umsetzung des ITT-Prinzips**

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Patienten mit UE Grad 3

1. **Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

☒**ja:** Der Endpunkt wurde verblindet erhoben.

☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Befragung der verblindeten Patienten sowie die Beurteilung von Krankheitszeichen und Symptomen erfolgten durch den ebenfalls verblindeten Investigator.

2. **Adäquate Umsetzung des ITT-Prinzips**

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Patienten mit UE Grad 4

1. **Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

☒**ja:** Der Endpunkt wurde verblindet erhoben.

☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Befragung der verblindeten Patienten sowie die Beurteilung von Krankheitszeichen und Symptomen erfolgten durch den ebenfalls verblindeten Investigator.

2. **Adäquate Umsetzung des ITT-Prinzips**

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

_____

_____

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Patienten mit UE Grad 5

1. **Verblindung der Endpunkterheber**

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

☒**ja:** Der Endpunkt wurde verblindet erhoben.

☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Befragung der verblindeten Patienten sowie die Beurteilung von Krankheitszeichen und Symptomen erfolgten durch den ebenfalls verblindeten Investigator.

2. **Adäquate Umsetzung des ITT-Prinzips**

Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung:

Endpunkt: Patienten mit UE von besonderem Interesse

1. **Verblindung der Endpunkterheber**

   Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.
   In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

   ☒**ja:** Der Endpunkt wurde verblindet erhoben.

   ☐**unklar:** Es finden sich keine diesbezüglichen Angaben.

   ☐**nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

   Angaben zum Kriterium; obligate Begründung für die Einstufung

Die Befragung der verblindeten Patienten sowie die Beurteilung von Krankheitszeichen und Symptomen erfolgten durch den ebenfalls verblindeten Investigator.

2. **Adäquate Umsetzung des ITT-Prinzips**

   Lost-to-follow-up-Patienten sind Patienten, bei denen die Zielkriterien nicht vollständig bis zum Studienende erhoben werden können (z. B. weil ein Patient seine Einwilligung zurückzieht). Protokollverletzer sind u. a. Patienten, die die zugeteilte Therapie nicht protokollgemäß durchlaufen (z. B. Therapieabbrecher oder –wechsler, Einnahme unerlaubter Begleitmedikationen). Zu beachten ist, dass Begriffe wie Lost-to-follow-up und Protokollverletzer in Publikationen jedoch zum Teil sehr unterschiedlich definiert sind. Des Weiteren finden sich auch Begriffe wie Drop-outs, Withdrawals etc., die in diesem Bogen möglichst zu vermeiden bzw. genau zu definieren sind. Kommen in einer Studie solche Patienten vor, ist es notwendig, dass diese ausreichend genau beschrieben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen berücksichtigt (i. d. R. ITT-Analyse, siehe Äquivalenzstudien) wurden. Bei einer ITT(„intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen bei Lost-to-follow-up-Patienten fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

☒**ja:** Eines der folgenden Merkmale trifft zu:

- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5 %) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante Patientenzahlen in Flussdiagramm und Ergebnistabelle), die dies bezweifeln lassen.
- Die Protokollverletzer und Lost-to-follow-up-Patienten sind so genau beschrieben (Art, Häufigkeit und Charakteristika pro Gruppe), dass deren möglicher Einfluss auf die Ergebnisse abschätzbar ist (eigenständige Analyse möglich).
- Die Strategie zur Berücksichtigung von Protokollverletzern und Lost-to-follow-up-Patienten (u. a. Ersetzen von fehlenden Werten, Wahl der Zielkriterien, statistische Verfahren) ist sinnvoll angelegt worden (verzerrt die Effekte nicht zugunsten der zu evaluierenden Behandlung).

☐**unklar**: Aufgrund unzureichender Darstellung ist der adäquate Umgang mit Protokollverletzern und Lost-to-follow-up-Patienten nicht einschätzbar.

☐**nein**: Keines der unter „ja“ genannten drei Merkmale trifft zu.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

Beachte die Hinweise zu Punkt A.4!

☒**ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.

☐**unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.

☐**nein:** Es liegen Anhaltspunkte für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

---

4. Fehlen sonstiger (endpunktspezifischer) Aspekte, die das Verzerrungspotenzial beeinflussen

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*

- *Anwendung inadäquater statistischer Verfahren*

☒**ja**

☐**nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

☒**niedrig**: Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

☐**hoch**: Die Ergebnisse sind möglicherweise relevant verzerrt.

falls hoch, obligate Begründung für die Einstufung: