

Dossier zur Nutzenbewertung gemäß § 35a SGB V

Baricitinib (Olumiant®)

Lilly Deutschland GmbH

Modul 4 A

Zur Behandlung erwachsener Patienten mit mittelschwerer bis schwerer aktiver rheumatoider Arthritis, die auf eine vorangegangene Behandlung mit einem oder mehreren konventionellen krankheitsmodifizierenden Antirheumatika (cDMARDs) unzureichend angesprochen oder diese nicht vertragen haben.

Medizinischer Nutzen und
medizinischer Zusatznutzen,
Patientengruppen mit therapeutisch
bedeutsamem Zusatznutzen

Inhaltsverzeichnis

	Seite
Inhaltsverzeichnis	1
Tabellenverzeichnis	4
Abbildungsverzeichnis	13
Abkürzungsverzeichnis	14
4 Modul 4 – allgemeine Informationen	21
4.1 Zusammenfassung der Inhalte von Modul 4.....	22
4.2 Methodik.....	60
4.2.1 Fragestellung.....	60
4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung.....	64
4.2.3 Informationsbeschaffung.....	67
4.2.3.1 Studien des pharmazeutischen Unternehmers.....	67
4.2.3.2 Bibliografische Literaturrecherche.....	67
4.2.3.3 Suche in Studienregistern.....	69
4.2.3.4 Selektion relevanter Studien.....	70
4.2.4 Bewertung der Aussagekraft der Nachweise.....	71
4.2.5 Informationssynthese und -analyse.....	73
4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien.....	73
4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien.....	74
4.2.5.3 Meta-Analysen.....	97
4.2.5.4 Sensitivitätsanalysen.....	98
4.2.5.5 Subgruppenmerkmale und andere Effektmodifikatoren.....	103
4.2.5.6 Indirekte Vergleiche.....	110
4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen.....	112
4.3.1 Ergebnisse randomisierter kontrollierter Studien mit dem zu bewertenden Arzneimittel.....	112
4.3.1.1 Ergebnis der Informationsbeschaffung – RCT mit dem zu bewertenden Arzneimittel.....	112
4.3.1.1.1 Studien des pharmazeutischen Unternehmers.....	112
4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche.....	114
4.3.1.1.3 Studien aus der Suche in Studienregistern.....	117
4.3.1.1.4 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel.....	118
4.3.1.2 Charakteristika der in die Bewertung eingeschlossenen Studien – RCT mit dem zu bewertenden Arzneimittel.....	120
4.3.1.2.1 Studiendesign und Studienpopulationen.....	120
4.3.1.2.2 Verzerrungspotenzial auf Studienebene.....	139
4.3.1.3 Ergebnisse aus randomisierten kontrollierten Studien.....	142
4.3.1.3.1 Mortalität – RCT.....	145
4.3.1.3.2 Morbidität – RCT.....	148
4.3.1.3.2.1 Druckschmerzhaftes Gelenke und geschwollene Gelenke – RCT..	148
4.3.1.3.2.2 <i>Disease Activity Score 28 High Sensitivity C-Reactive Protein</i> (DAS28-hsCRP) – RCT.....	155
4.3.1.3.2.3 <i>Simplified Disease Activity Index (SDAI)</i> – RCT.....	169

4.3.1.3.2.4	<i>Clinical Disease Activity Index (CDAI) – RCT</i>	177
4.3.1.3.2.5	EULAR Responsekriterien („ <i>good Response</i> “, DAS28-basiert) – RCT	185
4.3.1.3.2.6	<i>Patient Assessment of Pain (PAP) Visual Analog Scale (VAS) – RCT</i>	190
4.3.1.3.2.7	<i>Patient’s Global Assessment of Disease Activity (PtGADA) (VAS) – RCT</i>	200
4.3.1.3.2.8	<i>Health Assessment Questionnaire-Disability Index (HAQ-DI) – RCT</i>	210
4.3.1.3.2.9	Dauer der Morgensteifigkeit (min) – RCT	223
4.3.1.3.2.10	<i>European Quality of Life-5 Dimensions (EQ-5D) (VAS) – RCT</i>	229
4.3.1.3.2.11	Schwerste Müdigkeit (<i>Worst Tiredness</i>) – RCT	237
4.3.1.3.3	<i>Lebensqualität - RCT</i>	242
4.3.1.3.3.1	<i>Short Form (36 item) Health Survey Version 2 akut (SF-36 v2 akut) – RCT</i>	242
4.3.1.3.4	Sicherheit – RCT	270
4.3.1.3.5	Subgruppenanalysen – RCT	292
4.3.1.3.5.1	Druckschmerzhafte Gelenke und geschwollene Gelenke – RCT ..	342
4.3.1.3.5.2	<i>Disease Activity Score 28 High Sensitivity C-Reactive Protein (DAS28-hsCRP) – RCT</i>	344
4.3.1.3.5.3	<i>Simplified Disease Activity Index (SDAI) – RCT</i>	345
4.3.1.3.5.4	<i>Clinical Disease Activity Index (CDAI) – RCT</i>	346
4.3.1.3.5.5	EULAR Responsekriterien („ <i>good Response</i> “, DAS28-basiert) – RCT	347
4.3.1.3.5.6	<i>Patient Assessment of Pain (PAP) (VAS) – RCT</i>	348
4.3.1.3.5.7	<i>Patient’s Global Assessment of Disease Activity (PtGADA) (VAS) – RCT</i>	360
4.3.1.3.5.8	<i>Health Assessment Questionnaire-Disability Index (HAQ-DI) – RCT</i>	372
4.3.1.3.5.9	<i>European Quality of Life-5 Dimensions (EQ-5D) VAS – RCT</i>	373
4.3.1.3.5.9.1	Schwerste Müdigkeit (<i>Worst Tiredness</i>) - RCT	376
4.3.1.3.5.10	<i>Short Form (36 item) Health Survey Version 2 akut (SF-36 v2 akut) – RCT</i>	378
4.3.1.3.5.11	Sicherheit – RCT	389
4.3.1.3.6	Zusammenfassung der Ergebnisse aus randomisierten kontrollierten Studien.....	399
4.3.2	Weitere Unterlagen.....	427
4.3.2.1	Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien	427
4.3.2.1.1	Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche	427
4.3.2.1.2	Charakteristika der Studien für indirekte Vergleiche.....	427
4.3.2.1.3	Ergebnisse aus indirekten Vergleichen	427
4.3.2.1.3.1	Endpunkt xxx> – indirekte Vergleiche aus RCT.....	428
4.3.2.2	Nicht randomisierte vergleichende Studien	430
4.3.2.2.1	Ergebnis der Informationsbeschaffung – nicht randomisierte vergleichende Studien	430
4.3.2.2.2	Charakteristika der nicht randomisierten vergleichenden Studien.....	431
4.3.2.2.3	Ergebnisse aus nicht randomisierten vergleichenden Studien	431

4.3.2.2.3.1	<Endpunkt xxx> – nicht randomisierte vergleichende Studien.....	431
4.3.2.2.3.2	Subgruppenanalysen – nicht randomisierte vergleichende Studien	432
4.3.2.3	Weitere Untersuchungen.....	433
4.3.2.3.1	Ergebnis der Informationsbeschaffung – weitere Untersuchungen	433
4.3.2.3.2	Charakteristika der weiteren Untersuchungen	433
4.3.2.3.3	Ergebnisse aus weiteren Untersuchungen	434
4.3.2.3.3.1	<Endpunkt xxx> – weitere Untersuchungen	434
4.3.2.3.3.2	Subgruppenanalysen – weitere Untersuchungen	434
4.3.2.4	Zusammenfassung der Ergebnisse aus weiteren Unterlagen	434
4.4	Abschließende Bewertung der Unterlagen zum Nachweis des Zusatznutzens.....	435
4.4.1	Beurteilung der Aussagekraft der Nachweise	435
4.4.2	Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß.....	457
4.4.3	Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht	470
4.5	Begründung für die Vorlage weiterer Unterlagen und Surrogatendpunkte	471
4.5.1	Begründung für die Vorlage indirekter Vergleiche.....	471
4.5.2	Begründung für die Vorlage nicht randomisierter vergleichender Studien und weiterer Untersuchungen.....	471
4.5.3	Begründung für die Bewertung auf Grundlage der verfügbaren Evidenz, da valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen	471
4.5.4	Verwendung von Surrogatendpunkten	472
4.6	Liste der eingeschlossenen Studien.....	474
4.7	Referenzliste.....	475
Anhang 4-A : Suchstrategien – bibliografische Literaturrecherche		489
Anhang 4-B : Suchstrategien – Suche in Studienregistern.....		498
Anhang 4-C : Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Literaturrecherche).....		500
Anhang 4-D : Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern).....		501
Anhang 4-E : Methodik der eingeschlossenen Studien – RCT		512
Anhang 4-F : Bewertungsbögen zur Einschätzung von Verzerrungsaspekten		533

Tabellenverzeichnis

	Seite
Tabelle 4-1: Zusammenfassende Darstellung zum Ausmaß des Zusatznutzens auf Endpunktebene für die mITT Population.....	50
Tabelle 4-2: Gesamtaussage zum medizinischen Zusatznutzen für Subpopulation A2 und A3 (basierend auf den Ergebnissen der mITT Population).....	57
Tabelle 4-3: Einschlusskriterien für die Studienbewertung	64
Tabelle 4-4: Einteilung der Krankheitsaktivität der Rheumatoiden Arthritis	78
Tabelle 4-5: Kategorisierung der Patienten ohne Ansprechen (<i>Non-Responder</i>), mit <i>moderatem Ansprechen</i> oder mit <i>gutem Ansprechen</i> mittels EULAR Responsekriterien (DAS28-basiert)	83
Tabelle 4-6: Übersicht der Analyseansätze für dichotome Endpunkte der Nutzendimension „Sicherheit“ (unadjustierte Analysen)	96
Tabelle 4-7: Übersicht statistischer Analysemethoden	101
Tabelle 4-8: Liste der Studien des pharmazeutischen Unternehmers – RCT mit dem zu bewertenden Arzneimittel	113
Tabelle 4-9: Studien des pharmazeutischen Unternehmers, die nicht für die Nutzenbewertung herangezogen wurden – RCT mit dem zu bewertenden Arzneimittel.....	114
Tabelle 4-10: Relevante Studien (auch laufende Studien) aus der Suche in Studienregistern – RCT mit dem zu bewertenden Arzneimittel	118
Tabelle 4-11: Studienpool – RCT mit dem zu bewertenden Arzneimittel.....	119
Tabelle 4-12: Charakterisierung der eingeschlossenen Studien - RCT mit dem zu bewertenden Arzneimittel	121
Tabelle 4-13: Charakterisierung der Interventionen - RCT mit dem zu bewertenden Arzneimittel.....	124
Tabelle 4-14: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel – Demographische Charakteristika der Patienten	125
Tabelle 4-15: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel - Krankheitsgeschichte der Patienten.....	128
Tabelle 4-16: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel - Klinische Charakteristika der Patienten.....	131
Tabelle 4-17: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel.....	139
Tabelle 4-18: Verzerrungspotenzial auf Studienebene – Studienabbrecher/Patienten mit Therapieanpassung in RCT	141
Tabelle 4-19: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel	142
Tabelle 4-20: Operationalisierung der Mortalität.....	146

Tabelle 4-21: Bewertung des Verzerrungspotenzials für Mortalität in RCT mit dem zu bewertenden Arzneimittel	146
Tabelle 4-22: Ergebnisse für Mortalität aus RCT mit dem zu bewertenden Arzneimittel	147
Tabelle 4-23: Operationalisierung der Anzahl druckschmerzhafter Gelenke und geschwollener Gelenke	149
Tabelle 4-24: Bewertung des Verzerrungspotenzials für Anzahl druckschmerzhafter Gelenke und geschwollener Gelenke	149
Tabelle 4-25: Ergebnisse für Veränderung der Anzahl der druckschmerzhaften Gelenke (basierend auf 28 Gelenken) aus RCT mit dem zu bewertenden Arzneimittel.....	150
Tabelle 4-26: Ergebnisse für Veränderung der Anzahl der geschwollenen Gelenke (basierend auf 28 Gelenken) aus RCT mit dem zu bewertenden Arzneimittel.....	152
Tabelle 4-27: Operationalisierung des DAS28-hsCRP	156
Tabelle 4-28: Bewertung des Verzerrungspotenzials für DAS28-hsCRP in RCT mit dem zu bewertenden Arzneimittel	157
Tabelle 4-29: Ergebnisse für Veränderung des DAS28-hsCRP aus RCT mit dem zu bewertenden Arzneimittel (mBOCF)	158
Tabelle 4-30: Ergebnisse für Veränderung des DAS28-hsCRP aus RCT mit dem zu bewertenden Arzneimittel (Sensitivitätsanalyse - mLOCF)	161
Tabelle 4-31: Ergebnisse Veränderung des DAS28-hsCRP aus RCT mit dem zu bewertenden Arzneimittel (Sensitivitätsanalyse - MMRM)	162
Tabelle 4-32: Ergebnisse für Anteil Patienten mit DAS28-hsCRP $\leq 3,2$ aus RCT mit dem zu bewertenden Arzneimittel	164
Tabelle 4-33: Ergebnisse für Anteil Patienten mit DAS28-hsCRP $< 2,6$ aus RCT mit dem zu bewertenden Arzneimittel	166
Tabelle 4-34: Operationalisierung des SDAI	170
Tabelle 4-35: Bewertung des Verzerrungspotenzials für SDAI in RCT mit dem zu bewertenden Arzneimittel	170
Tabelle 4-36: Ergebnisse für Veränderung des SDAI aus RCT mit dem zu bewertenden Arzneimittel.....	171
Tabelle 4-37: Ergebnisse für Anteil Patienten mit SDAI $\leq 3,3$ aus RCT mit dem zu bewertenden Arzneimittel	174
Tabelle 4-38: Operationalisierung des CDAI.....	178
Tabelle 4-39: Bewertung des Verzerrungspotenzials für CDAI in RCT mit dem zu bewertenden Arzneimittel	178
Tabelle 4-40: Ergebnisse für Veränderung des CDAI aus RCT mit dem zu bewertenden Arzneimittel.....	179
Tabelle 4-41: Ergebnisse für Anteil Patienten mit CDAI ≤ 10 aus RCT mit dem zu bewertenden Arzneimittel	182
Tabelle 4-42: Operationalisierung der EULAR Responsekriterien („good Response“, DAS28-basiert).....	186

Tabelle 4-43: Bewertung des Verzerrungspotenzials für EULAR Responsekriterien („good Response“, DAS28-basiert) in RCT mit dem zu bewertenden Arzneimittel	186
Tabelle 4-44: Ergebnisse für Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert) aus RCT mit dem zu bewertenden Arzneimittel.....	187
Tabelle 4-45: Operationalisierung von PAP (VAS).....	191
Tabelle 4-46: Bewertung des Verzerrungspotenzials für PAP (VAS) in RCT mit dem zu bewertenden Arzneimittel	191
Tabelle 4-47: Ergebnisse für Veränderung des PAP (VAS) aus RCT mit dem zu bewertenden Arzneimittel	192
Tabelle 4-48: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel	195
Tabelle 4-49: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel	197
Tabelle 4-50: Operationalisierung des PtGADA (VAS).....	201
Tabelle 4-51: Bewertung des Verzerrungspotenzials für PtGADA (VAS) in RCT mit dem zu bewertenden Arzneimittel	201
Tabelle 4-52: Ergebnisse für Veränderung des PtGADA (VAS) aus RCT mit dem zu bewertenden Arzneimittel	202
Tabelle 4-53: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel	205
Tabelle 4-54: Ergebnisse für Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel	207
Tabelle 4-55: Operationalisierung des HAQ-DI	211
Tabelle 4-56: Bewertung des Verzerrungspotenzials HAQ-DI in RCT mit dem zu bewertenden Arzneimittel	211
Tabelle 4-57: Ergebnisse für Veränderung des HAQ-DI aus RCT mit dem zu bewertenden Arzneimittel (mBOCF)	212
Tabelle 4-58: Ergebnisse für Veränderung des HAQ-DI aus RCT mit dem zu bewertenden Arzneimittel (Sensitivitätsanalyse - mLOCF)	215
Tabelle 4-59: Ergebnisse für Veränderung des HAQ-DI aus RCT mit dem zu bewertenden Arzneimittel (Sensitivitätsanalyse - MMRM)	216
Tabelle 4-60: Ergebnisse für Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$ aus RCT mit dem zu bewertenden Arzneimittel	218
Tabelle 4-61: Ergebnisse für Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$ aus RCT mit dem zu bewertenden Arzneimittel	220
Tabelle 4-62: Operationalisierung der Dauer der Morgensteifigkeit (min)	224
Tabelle 4-63: Bewertung des Verzerrungspotenzials für Dauer der Morgensteifigkeit (min) in RCT mit dem zu bewertenden Arzneimittel	224
Tabelle 4-64: Ergebnisse für Veränderung der Dauer der Morgensteifigkeit aus RCT mit dem zu bewertenden Arzneimittel (mLOCF).....	225

Tabelle 4-65: Operationalisierung der EQ-5D VAS	230
Tabelle 4-66: Bewertung des Verzerrungspotenzials für EQ-5D VAS in RCT mit dem zu bewertenden Arzneimittel	230
Tabelle 4-67: Ergebnisse für Veränderung der EQ-5D VAS aus RCT mit dem zu bewertenden Arzneimittel (mLOCF)	231
Tabelle 4-68: Ergebnisse für Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel.....	234
Tabelle 4-69: Operationalisierung der schwersten Müdigkeit	238
Tabelle 4-70: Bewertung des Verzerrungspotenzials für schwerste Müdigkeit in RCT mit dem zu bewertenden Arzneimittel.....	238
Tabelle 4-71: Ergebnisse für Veränderung der schwersten Müdigkeit aus RCT mit dem zu bewertenden Arzneimittel (mLOCF)	239
Tabelle 4-72: Operationalisierung des SF-36 v2 akut.....	243
Tabelle 4-73: Bewertung des Verzerrungspotenzials für SF-36 v2 akut in RCT mit dem zu bewertenden Arzneimittel	244
Tabelle 4-74: Ergebnisse für Veränderung des SF-36 v2 akut PCS aus RCT mit dem zu bewertenden Arzneimittel	245
Tabelle 4-75: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel.....	247
Tabelle 4-76: Ergebnisse für Veränderung des SF-36 v2 akut MCS aus RCT mit dem zu bewertenden Arzneimittel	249
Tabelle 4-77: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel	251
Tabelle 4-78: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Körperliche Funktionsfähigkeit aus RCT mit dem zu bewertenden Arzneimittel.....	253
Tabelle 4-79: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Körperliche Rollenfunktion aus RCT mit dem zu bewertenden Arzneimittel.....	255
Tabelle 4-80: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Körperliche Schmerzen aus RCT mit dem zu bewertenden Arzneimittel	257
Tabelle 4-81: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Allgemeiner Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel	259
Tabelle 4-82: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Vitalität aus RCT mit dem zu bewertenden Arzneimittel	261
Tabelle 4-83: Ergebnisse für Veränderung der SF-36 v2 akut Domäne– Soziale Funktionsfähigkeit aus RCT mit dem zu bewertenden Arzneimittel.....	263
Tabelle 4-84: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Emotionale Rollenfunktion aus RCT mit dem zu bewertenden Arzneimittel.....	265
Tabelle 4-85: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Mentaler Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel	267
Tabelle 4-86: Operationalisierung der <i>Sicherheit</i>	271

Tabelle 4-87: Bewertung des Verzerrungspotenzials für Sicherheit in RCT mit dem zu bewertenden Arzneimittel	272
Tabelle 4-88: Ergebnisse für Todesfälle aus RCT mit dem zu bewertenden Arzneimittel....	273
Tabelle 4-89: Ergebnisse für SUE aus RCT mit dem zu bewertenden Arzneimittel.....	274
Tabelle 4-90: Ergebnisse für UE aus RCT mit dem zu bewertenden Arzneimittel	276
Tabelle 4-91: Ergebnisse für schwere UE (ohne ausschließliche Laborereignisse) aus RCT mit dem zu bewertenden Arzneimittel	278
Tabelle 4-92: Ergebnisse für Studienabbrüche aufgrund von UE aus RCT mit dem zu bewertenden Arzneimittel	280
Tabelle 4-93: Ergebnisse für Therapieabbrüche aufgrund von UE aus RCT mit dem zu bewertenden Arzneimittel	282
Tabelle 4-94: Ergebnisse für UE von besonderem Interesse aus RCT mit dem zu bewertenden Arzneimittel	284
Tabelle 4-95: Ergebnisse für Depressionen QIDS-SR16 aus RCT mit dem zu bewertenden Arzneimittel.....	289
Tabelle 4-96: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A2 von Baseline bis Woche 52 (1)	294
Tabelle 4-97: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A2 von Baseline bis Woche 52 (2)	298
Tabelle 4-98: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A2 von Baseline bis Woche 52 (3)	302
Tabelle 4-99: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A2 von Baseline bis Woche 52 (4)	306
Tabelle 4-100: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A3 von Baseline bis Woche 52 (1)	310
Tabelle 4-101: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A3 von Baseline bis Woche 52 (2)	314
Tabelle 4-102: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A3 von Baseline bis Woche 52 (3)	318
Tabelle 4-103: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A3 von Baseline bis Woche 52 (4)	322
Tabelle 4-104: Übersicht über alle durchgeführten Interaktionstests in der mITT Population von Baseline bis Woche 52 (1)	326
Tabelle 4-105: Übersicht über alle durchgeführten Interaktionstests in der mITT Population von Baseline bis Woche 52 (2)	330
Tabelle 4-106: Übersicht über alle durchgeführten Interaktionstests in der mITT Population von Baseline bis Woche 52 (3)	334
Tabelle 4-107: Übersicht über alle durchgeführten Interaktionstests in der mITT Population von Baseline bis Woche 52 (4)	338
Tabelle 4-108: Ergebnisse für Veränderung der Anzahl der geschwollenen Gelenke aus RCT mit dem zu bewertenden Arzneimittel nach BMI zu Baseline.....	342

Tabelle 4-109: Ergebnisse für Anteil Patienten mit DAS28-hsCRP < 2,6 aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus.....	344
Tabelle 4-110: Ergebnisse für Anteil Patienten mit SDAI ≤ 3,3 aus RCT mit dem zu bewertenden Arzneimittel nach spezifischer Region I.....	345
Tabelle 4-111: Ergebnisse für Anteil Patienten mit CDAI ≤ 10 aus RCT mit dem zu bewertenden Arzneimittel nach Ethnie	346
Tabelle 4-112: Ergebnisse für Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert) aus RCT mit dem zu bewertenden Arzneimittel nach Ethnie	347
Tabelle 4-113: Ergebnisse für Veränderung des PAP (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach Geschlecht	348
Tabelle 4-114: Ergebnisse für Veränderung des PAP (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach spezifischer Region II	349
Tabelle 4-115: Ergebnisse für Veränderung des PAP (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität (ACPA).....	350
Tabelle 4-116: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Geschlecht.....	351
Tabelle 4-117: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus	352
Tabelle 4-118: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus	353
Tabelle 4-119: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Geschlecht	354
Tabelle 4-120: Ergebnisse für Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Alter.....	355
Tabelle 4-121: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus	356
Tabelle 4-122: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus.....	358
Tabelle 4-123: Ergebnisse für Veränderung des PtGADA (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach Geschlecht	360
Tabelle 4-124: Ergebnisse für Veränderung des PtGADA (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach spezifischer Region I.....	361
Tabelle 4-125: Ergebnisse für Veränderung des PtGADA (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus.....	362
Tabelle 4-126: Ergebnisse für Veränderung des PtGADA (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus	363

Tabelle 4-127: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus	364
Tabelle 4-128: Ergebnisse für Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus.....	365
Tabelle 4-129: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline	366
Tabelle 4-130: Ergebnisse für Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 aus RCT mit dem zu bewertenden Arzneimittel nach Alter.....	367
Tabelle 4-131: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus	368
Tabelle 4-132: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus	369
Tabelle 4-133: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline ..	371
Tabelle 4-134: Ergebnisse für Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$ aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität (ACPA)	372
Tabelle 4-135: Ergebnisse für Veränderung der EQ-5D VAS aus RCT mit dem zu bewertenden Arzneimittel nach Alter	373
Tabelle 4-136: Ergebnisse für Veränderung der EQ-5D VAS aus RCT mit dem zu bewertenden Arzneimittel nach Therapie mit Kortikosteroiden zu Baseline	374
Tabelle 4-137: Ergebnisse für Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität (ACPA)	375
Tabelle 4-138: Ergebnisse für Veränderung der schwersten Müdigkeit aus RCT mit dem zu bewertenden Arzneimittel nach Geschlecht	376
Tabelle 4-139: Ergebnisse für Veränderung der schwersten Müdigkeit aus RCT mit dem zu bewertenden Arzneimittel nach Therapie mit Kortikosteroiden zu Baseline.....	377
Tabelle 4-140: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel nach spezifischer Region II.	378
Tabelle 4-141: Ergebnisse für Veränderung der SF-36 v2 akut MCS aus RCT mit dem zu bewertenden Arzneimittel nach Therapie mit Kortikosteroiden zu Baseline	379
Tabelle 4-142: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel nach Erkrankungsdauer.....	380
Tabelle 4-143: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus.	381
Tabelle 4-144: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus	382

Tabelle 4-145: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel nach Art der Begleitmedikation	383
Tabelle 4-146: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline	384
Tabelle 4-147: Ergebnisse für Veränderung der SF-36 v2 akut Domäne - Körperliche Funktionsfähigkeit aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline	385
Tabelle 4-148: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Allgemeiner Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel nach Gewicht zu Baseline	386
Tabelle 4-149: Ergebnisse für Veränderung der SF-36 v2 akut Domäne - Emotionale Funktionsfähigkeit aus RCT mit dem zu bewertenden Arzneimittel nach Therapie mit Kortikosteroiden zu Baseline	387
Tabelle 4-150: Ergebnisse für UE aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus	389
Tabelle 4-151: Ergebnisse für UE aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus.....	390
Tabelle 4-152: Ergebnisse für SUE aus RCT mit dem zu bewertenden Arzneimittel nach BMI zu Baseline.....	391
Tabelle 4-153: Ergebnisse für SUE aus RCT mit dem zu bewertenden Arzneimittel nach Erkrankungsdauer.....	392
Tabelle 4-154: Ergebnisse für SUE aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus.....	393
Tabelle 4-155: Ergebnisse für schwere UE (ohne ausschließliche Laborereignisse) aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline	394
Tabelle 4-156: Ergebnisse für Therapieabbruch aufgrund von UE aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline	395
Tabelle 4-157: Ergebnisse für UE von besonderem Interesse – QIDS-SR16 aus RCT mit dem zu bewertenden Arzneimittel nach Gewicht zu Baseline	396
Tabelle 4-158: Ergebnisse für UE von besonderem Interesse – QIDS-SR16 aus RCT mit dem zu bewertenden Arzneimittel nach Krankheitsaktivität zu Baseline	397
Tabelle 4-159: Ergebnisse für UE von besonderem Interesse – QIDS-SR16 aus RCT mit dem zu bewertenden Arzneimittel nach Therapie mit Kortikosteroiden zu Baseline.....	398
Tabelle 4-160: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche	427
Tabelle 4-161: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden.....	428
Tabelle 4-162: Operationalisierung von <Endpunkt xxx>.....	428
Tabelle 4-163: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche	429

Tabelle 4-164: Ergebnisse für <Endpunkt xxx> aus RCT für indirekte Vergleiche.....	429
Tabelle 4-165: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien	431
Tabelle 4-166: Operationalisierung von <Endpunkt xxx>.....	431
Tabelle 4-167: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien	432
Tabelle 4-168: Operationalisierung von <Endpunkt xxx> – weitere Untersuchungen.....	434
Tabelle 4-169: Anforderungen an die Aussage zur Beleglage/Aussagesicherheit für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit	436
Tabelle 4-170: Inferenzstatistische Schwellenwerte (Hypothesengrenzen) für relative Effektmaße gemäß IQWiG-Methodenpapier	438
Tabelle 4-171: Methodik zur Feststellung des Ausmaßes des Zusatznutzens – hierarchisierte Kriterien gemäß AM-NutzenV mit Ergänzungen ^a	441
Tabelle 4-172: Kategorisierung der im Dossier zu Baricitinib dargestellten Endpunkte in die Zielgrößenkategorien schwerwiegende und nicht-schwerwiegende Symptome (bzw. Folgekomplikationen) und Nebenwirkungen mit Ergänzungen ^a	442
Tabelle 4-173: Nutzendimensionen und deren Operationalisierung in Studien zur Bewertung des medizinischen Zusatznutzens von Baricitinib und Ergebnissicherheit, Einstufung der Beleglage und Zielgrößenkategorie	454
Tabelle 4-174: Zusammenfassende Darstellung zum Ausmaß des Zusatznutzens auf Endpunktebene für die mITT Population.....	461
Tabelle 4-175: Gesamtaussage zum medizinischen Zusatznutzen für die mITT Population	468
Tabelle 4-176: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens.....	470
Tabelle 4-177: (Anhang) Studiendesign und -methodik für Studie RA-BEAM (JADV).....	512
Tabelle 4-178 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie RA-BEAM (JADV).....	534

Abbildungsverzeichnis

	Seite
Abbildung 1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel	115
Abbildung 2: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel (für die Subpopulationen A1, A2 und A3).....	116
Abbildung 3: Studiendesign der BEAM-Studie (RA-JADV).....	137
Abbildung 4: Zeitlicher Verlauf des DAS28-hsCRP (mBOCF) über 52 Wochen (mITT) ...	160
Abbildung 5: Zeitlicher Verlauf des SDAI über 52 Wochen (mITT).....	173
Abbildung 6: Zeitlicher Verlauf des CDAI über 52 Wochen (mITT)	181
Abbildung 7: Zeitlicher Verlauf des PAP (VAS) über 52 Wochen (mITT)	194
Abbildung 8: Zeitlicher Verlauf des PtGADA (VAS) über 52 Wochen (mITT).....	204
Abbildung 9: Zeitlicher Verlauf des HAQ-DI (mBOCF) über 52 Wochen (mITT).....	214
Abbildung 10: Zeitlicher Verlauf der Dauer der Morgensteifigkeit über 52 Wochen (mITT).....	227

Abkürzungsverzeichnis

Abkürzung	Bedeutung
ACPA	<i>Anti Citrullinated Peptide/Protein Antibodies</i> Antikörper gegen citrullinierte Peptide
ACR	<i>American College of Rheumatology</i> Amerikanische Rheumatologie-Gesellschaft
ACR20	20%ige Verbesserung der ACR-Kriterien
ACR50	50%ige Verbesserung der ACR-Kriterien
ACR70	70%ige Verbesserung der ACR-Kriterien
AE	<i>Adverse Event</i> Unerwünschtes Ereignis
AESI	<i>Adverse Event of Special Interest</i> Unerwünschtes Ereignis von besonderem Interesse
AMNOG	Arzneimittelmarktneuordnungsgesetz
ALT	Alanin-Aminotransferase
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
ANCOVA	<i>Analysis of Covariance</i> Kovarianzanalyse
ARD	<i>Absolute Risk Difference</i> Absolute Risikodifferenz
AST	Aspartat-Aminotransferase
bDMARD	<i>Biological Disease Modifying Anti-Rheumatic Drug</i> Biologisches Basistherapeutikum
BMI	<i>Body Mass Index</i>
boDMARDs	<i>Biological Originator Disease Modifying Anti-Rheumatic Drug</i> Biologisches Basistherapeutikum (Originator)
bsDMARDs	<i>Biosimilar Disease Modifying Anti-Rheumatic Drug</i> Biologisches Basistherapeutikum (Biosimilar)
BSG	Blutsenkungsgeschwindigkeit
bzw.	Beziehungsweise
ca.	Cirka
CDAI	<i>Clinical Disease Activity Index</i>
cDMARD	<i>Conventional Disease Modifying Anti-Rheumatic Drug</i> Konventionelles Basistherapeutikum
CONSORT	<i>Consolidated Standards of Reporting Trials</i>

Abkürzung	Bedeutung
CPK	<i>Creatine Phosphokinase</i> Kreatinphosphokinase
CRP	<i>C-Reactive Protein</i> C-reaktives Protein
csDMARD	<i>Conventional Synthetic Disease Modifying Anti-Rheumatic Drug</i> Konventionelles synthetisches Basistherapeutikum
CSR	<i>Clinical Study Report</i> Klinischer Studienbericht
CTD	<i>Common Technical Document</i>
DAS	<i>Disease Activity Score</i>
DGRh	Deutsche Gesellschaft für Rheumatologie e.V.
d. h.	Das heißt
DIMDI	Deutsches Institut für Medizinische Dokumentation
dl	Deziliter
DMARD	<i>Disease Modifying Anti-Rheumatic Drug</i> Krankheitsmodifizierende antirheumatische Medikamente, auch „Basistherapeutika“
DNA	<i>Deoxyribonucleic Acid</i> Desoxyribonukleinsäure
DSM-IV	<i>Diagnostic and Statistical Manual of Mental Disorders-IV</i>
eCRF	<i>Electronic Case Report Form</i>
EG	Europäische Gemeinschaft
eGFR	<i>Estimated Glomerular Filtration Rate</i> Geschätzte glomeruläre Filtrationsrate
EKG	Elektrokardiogramm
ESR	<i>Erythrocyte Sedimentation Rate</i> Erythrozytensedimentationsrate
etc.	Et cetera
EU-CTR	EU Clinical Trials Register
EULAR	<i>European League Against Rheumatism</i> Europäische Rheumatologie Gesellschaft
ePRO	<i>Electronic Patient Reported Outcome</i> Elektronisches <i>Patient Reported Outcome</i>
EQ-5D	<i>European Quality of Life-5 Dimensions</i>
EQ-5D-5L	<i>European Quality of Life-5 Dimensions-5 Levels</i>

Abkürzung	Bedeutung
FACIT-F	<i>Functional Assessment of Chronic Illness Therapy-Fatigue</i>
FSH	Follikelstimulierendes Hormon
G-BA	Gemeinsamer Bundesausschuss
GCP	<i>Good Clinical Practice</i> Gute klinische Praxis
gem.	Gemäß
ggf.	Gegebenenfalls
GmbH	Gesellschaft mit beschränkter Haftung
HAQ-DI	<i>Health Assessment Questionnaire-Disability Index</i>
HBV	Hepatitis-B-Virus
HCV	Hepatitis-C-Virus
HIV	Humanes Immundefizienz-Virus
HR	Hazard Ratio
hs	<i>High-Sensitivity</i> Hochsensitiv
i.m.	Intramuskulär
i. v.	Intravenös
ICH	<i>International Conference on Harmonisation</i>
ICTRP	<i>International Clinical Trials Registry Platform</i>
IL	Interleukin
Inf.	<i>Infeasible</i>
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
IR	<i>Inadequate Response</i> Unzureichendes Ansprechen
ITT	<i>Intention-To-Treat</i> Statistische Auswertung von Studiendaten basierend auf <i>a priori</i> definierten Patientenkollektiv
IU	<i>International Units</i> Internationale Einheiten
IVRS	<i>Interactive Voice Response System</i>
J.	Jahre
JAK	Januskinase
kg	Kilogramm

Abkürzung	Bedeutung
KI	Konfidenzintervall
l	Liter
LSM	<i>Least Square Mean</i> Mittelwert der kleinsten Quadrate
LSMD	<i>Least Square Mean Difference</i> Mittelwertdifferenz der kleinsten Quadrate
m	Männlich
MACE	<i>Major Adverse Cardiovascular Events</i>
mBOCF	<i>modified Baseline Observation Carried Forward</i> Fehlende Beobachtungen gehen mittels der zuerst erhobenen und modifizierten Beobachtung in die Auswertung ein
MCID	<i>Minimal Clinically Important Difference</i>
MCS	<i>Mental Component Score</i> Mentale Summenskala
MedDRA	<i>Medical Dictionary for Regulatory Activities</i>
mg	Milligramm
MI	Myokardinfarkt
MID	<i>Minimal Important Difference</i>
min	Minute
mITT	<i>Modified Intention-To-Treat</i> Statistische Auswertung von Studiendaten basierend auf modifiziertem, <i>a priori</i> definiertem Patientenkollektiv
MJS	<i>Morning Joint Stiffness</i> Morgensteifigkeit
ml	Milliliter
µl	Mikroliter
mLOCF	<i>Modified Last Observation Carried Forward</i> Fehlende Beobachtungen gehen mittels der zuletzt erhobenen und modifizierten Beobachtung in die Auswertung ein
mm	<i>Millimeter</i>
MMRM	<i>Mixed Effect Model Repeated Measurement</i> Statistische Imputationsmethode
MTC	<i>Mixed Treatment Comparisons</i>
mTSS	<i>Modified Total Sharp Score</i>
MTX	Methotrexat

Abkürzung	Bedeutung
MW	Mittelwert
MWD	Mittelwertdifferenz
N	Anzahl
n. b.	Nicht berechenbar
NE	<i>Not Estimated</i>
NRI	<i>Non-Responder Imputation</i> Statistische Imputationsmethode
NRS	<i>Numeric Rating Scale</i> Numerische Bewertungsskala
NSAID	<i>Nonsteroidal Anti-Inflammatory Drug</i> Nichtsteroidales antientzündliches Medikament
NYHA	<i>New York Heart Association</i>
OR	<i>Odds Ratio</i> Quotenverhältnis
PAP	<i>Patient Assessment of Pain</i>
Pbo	Placebo
PCS	<i>Physical Component Score</i> Physischer Summenskala
Peto OR	<i>Peto Odds Ratio</i> Peto-Quotenverhältnis
PhGADA	<i>Physician´s Global Assessment of Disease Activity</i>
p.o.	<i>Per oral</i>
PP	<i>Per Protocol</i>
PPD	<i>Purified Protein Derivate</i>
PRO	<i>Patient Reported Outcome</i> Messgrößen, die auf Patienteneinschätzungen basieren
PT	<i>Preferred Term</i>
PtGADA	<i>Patient´s Global Assessment of Disease Activity</i>
pU	Pharmazeutischer Unternehmer
Q.D.	<i>Quaque die</i> Einmal täglich
QIDS-SR ₁₆	<i>Quick Inventory of Depressive Symptomatology Self-Rated-16</i>
QUEST-RA	<i>Quantitative Standard Monitoring of Patients with Rheumatoid Arthritis</i>
RA	Rheumatoide Arthritis

Abkürzung	Bedeutung
RAID-Score	<i>Rheumatoid Arthritis Impact of Disease Score</i>
RF	<i>Rheumafaktor</i>
RCT	<i>Randomized Controlled Trial</i> Randomisierte kontrollierte Studie
RR	<i>Risk Ratio</i> Relatives Risiko
RS	Randomisierte Population
SAE	<i>Serious Adverse Event</i>
SAP	<i>Statistical Analysis Plan</i> Statistischer Analyseplan
SAS	<i>Statistical Analysis System</i> Software für statistische Analysen
s.c.	Subkutan
SD	<i>Standard Deviation</i> Standardabweichung
SDAI	<i>Simplified Disease Activity Index</i>
SE	<i>Standard Error</i> Standardfehler
SF-36 v2 akut	<i>Short Form (36 item) Health Survey Version 2 akut</i>
SGB	Sozialgesetzbuch
SJC	<i>Swollen Joint Count</i> Anzahl geschwollener Gelenke
SMQ	<i>Standardised MedDRA Queries</i> Standardisierte MedDRA-Abfrage
SOC	<i>System Organ Class</i>
STROBE	<i>Strengthening the Reporting of Observational Studies in Epidemiology</i>
SUE	Schwerwiegende Unerwünschte Ereignisse
TB	Tuberkulose
TEAE	<i>Treatment Emergent Adverse Event</i>
TJC	<i>Tender Joint Count</i> Anzahl druckschmerzhafter Gelenke
TNF	<i>Tumor Necrosis Factor</i> Tumornekrosefaktor
TREND	<i>Transparent Reporting of Evaluations with Non-Randomized Design</i>
tsDMARD	<i>Targeted Synthetic Disease Modifying Anti-Rheumatic Drug</i>

Abkürzung	Bedeutung
TSH	Thyreoidea-stimulierendes Hormon
U	<i>Unit</i>
u. a.	Unter Anderem
UE	Unerwünschte Ereignisse
ULN	<i>Upper Limit of Normal</i>
US	<i>United States</i> Vereinigte Staaten von Amerika
VAS	<i>Visual Analogue Scale</i> Visuelle Analogskala
VerfO	Verfahrensordnung
vs.	Versus
w	Weiblich
WBC	<i>White Blood Cell</i>
WHO	<i>World Health Organization</i>
WPAI-RA	<i>Work and Productivity and Activity Impairment-Rheumatoid Arthritis</i>
Wo.	Woche
zVT	Zweckmäßige Vergleichstherapie

4 Modul 4 – allgemeine Informationen

Modul 4 enthält folgende Angaben:

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

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

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

4.1 Zusammenfassung der Inhalte von Modul 4

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

Fragestellung

Gemäß Fachinformation ist:

„Olumiant angezeigt zur Behandlung von mittelschwerer bis schwerer aktiver rheumatoider Arthritis bei erwachsenen Patienten, die auf eine vorangegangene Behandlung mit einem oder mehreren krankheitsmodifizierenden Antirheumatika (DMARDs) unzureichend angesprochen oder diese nicht vertragen haben. Olumiant kann als Monotherapie oder in Kombination mit Methotrexat eingesetzt werden [1].“

Es sollen der medizinische Nutzen und der medizinische Zusatznutzen gemäß § 35a Sozialgesetzbuch (SGB) V von Baricitinib zur Behandlung erwachsener Patienten mit mittelschwerer bis schwerer aktiver rheumatoider Arthritis (RA), die auf eine vorangegangene Behandlung mit einem oder mehreren konventionellen krankheitsmodifizierenden Antirheumatika (*Conventional Disease Modifying Anti-Rheumatic Drugs*, cDMARDs) unzureichend angesprochen oder diese nicht vertragen haben, vergleichend zu einer vom Gemeinsamen Bundesausschuss (G-BA) festgelegten zweckmäßigen Vergleichstherapie (zVT) bewertet werden [2]. Baricitinib kann als Monotherapie oder in Kombination mit Methotrexat (MTX) eingesetzt werden.

Ursprünglich wurde eine Anfrage an den G-BA gestellt, die zVT für Baricitinib für den Fall zu bestimmen, dass eine Zulassung von Baricitinib neben der Zweit- und Drittlinien-Behandlung auch für die Erstlinien-Behandlung erfolgt. Eine Behandlung Therapie-naiver Patienten ohne vorherige Behandlung ist jedoch außerhalb der Zulassung von Baricitinib und somit für dieses Nutzendossier nicht von Relevanz.

Gemäß G-BA-Beratungsgespräch vom 10.12.2015 [2] sind im vorliegenden Anwendungsgebiet verschiedene Therapiesituationen zu unterscheiden, die in:

- Therapie-naive Patienten ohne vorherige Behandlung oder
- Patienten mit vorangegangenen Behandlungen mit einem oder mehreren konventionellen Basistherapeutika (cDMARDs), einschließlich MTX, oder (Anwendungsgebiet A)
- Patienten mit vorangegangenen Behandlungen mit einem oder mehreren biologischen Basistherapeutika (*Biological Disease Modifying Anti-Rheumatic Drugs*, bDMARDs) (Anwendungsgebiet B)

zu differenzieren sind.

Die Zielpopulation wurde aufgrund der verschiedenen, durch den G-BA festgelegten Therapiesituationen, in zwei Anwendungsgebiete aufgeteilt, die nachfolgend als Anwendungsgebiet A (Modul 4 A) und Anwendungsgebiet B (Modul 4 B) bezeichnet werden (siehe Module 2/3; Abschnitt 3.1).

Im Anwendungsgebiet A werden ausschließlich Informationen und Ergebnisse für Patienten mit vorangegangenen Behandlungen mit **einem** oder **mehreren** cDMARDs, inklusive MTX, abgebildet und der medizinische Nutzen und der medizinische Zusatznutzen vergleichend zu der vom G-BA genannten zVT im Anwendungsgebiet A bewertet.

Die Zielpopulation im Anwendungsgebiet A wurde vom G-BA in folgende Subpopulationen unterteilt:

- Subpopulation A1: Patienten, bei denen **keine ungünstigen Prognosefaktoren**¹ vorliegen und die unzureichend auf eine vorangegangene Behandlung mit **einem** cDMARD (inklusive MTX) ansprachen.
- Subpopulation A2: Patienten **mit ungünstigen Prognosefaktoren**¹, die unzureichend auf eine vorangegangene Behandlung mit **einem** cDMARD (inklusive MTX) ansprachen.
- Subpopulation A3: Patienten, die unzureichend auf eine vorangegangene Behandlung mit **mehreren** cDMARDs (inklusive MTX) ansprachen.

Für Patienten der Subpopulation A1 hat der G-BA die zVT alternative cDMARDs, sofern geeignet (z. B. MTX, Leflunomid) als Mono- oder Kombinationstherapie, bestimmt. Derzeit liegt keine direkte Evidenz zum Vergleich von Baricitinib mit der vom G-BA genannten zVT für diese Subpopulation vor.

Für die Subpopulationen A2 und A3 hat der G-BA die zVT bDMARDs in Kombination mit MTX (Adalimumab oder Etanercept oder Certolizumab-Pegol oder Golimumab oder Abatacept oder Tocilizumab); ggf. als Monotherapie unter Berücksichtigung des jeweiligen Zulassungsstatus bei MTX-Unverträglichkeit benannt.

Für die Subpopulationen A2 und A3 werden Ergebnisse von Baricitinib im Vergleich zu Adalimumab aus der randomisierten, aktiv-kontrollierten Phase-III-Studie RA-BEAM (JADV (NCT01710358)) zu Woche 24 und Woche 52 dargestellt. Die *Lilly Deutschland GmbH* folgt somit der vom G-BA genannten zVT.

¹ Beispiele für ungünstige Prognosefaktoren:

- Nachweis von Autoantikörpern (z. B. Rheumafaktoren, hohe Spiegel von Antikörpern gegen citrullinierte Peptid Antigene)
- Hohe Krankheitsaktivität (nachgewiesen durch DAS- bzw. DAS28-Bewertungssystem, geschwollene Gelenke, Parameter der Akute-Phase-Reaktion wie z. B. C-Reaktives Protein, Erythrozytensedimentationsrate)
- Frühes Auftreten von Gelenkerosionen.

Quelle: [2].

Datenquellen

Zur Identifizierung relevanter klinischer Studien von Baricitinib im Anwendungsgebiet A wurden verschiedene Datenquellen zur Informationsbeschaffung herangezogen. Informationsquellen waren interne Informationssysteme und Datenbanken der *Eli Lilly and Company*. Zusätzlich wurde eine systematische Literaturrecherche in den Datenbanken Cochrane (*Cochrane Central Register of Controlled Trials*), MEDLINE und EMBASE durchgeführt.

Zusätzlich wurde eine Studienregistersuche nach abgeschlossenen, abgebrochenen und laufenden Studien durchgeführt. Diese erfolgte in den öffentlich verfügbaren Studienregistern *ClinicalTrials.gov* (www.clinicaltrials.gov), *EU Clinical Trials Register* (www.clinicaltrialsregister.eu), der Datenbank zu in Deutschland genehmigten klinischen Prüfungen mit Arzneimitteln *PharmNet.Bund* (www.pharmnet-bund.de) und dem *International Clinical Trials Registry Platform* Suchportal (*ICTRP Search Portal*) der *World Health Organization* (WHO; www.who.int/ictrp).

Ein-/Ausschlusskriterien für Studien

Für die Auswahl in die Bewertung einzubeziehender Studien wurden Ein- und Ausschlusskriterien bezüglich der Studienpopulation, der Intervention, der Vergleichstherapie, der Endpunkte, des Studientyps, der Studiendauer und des Publikationstyps formuliert. Studien wurden nach den in Tabelle 4-3 genannten Kriterien für die Bewertung des medizinischen Nutzens und Zusatznutzens ein- bzw. ausgeschlossen.

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

Aussagekraft der Nachweise

Die Bewertung der Aussagekraft der Ergebnisse der eingeschlossenen Studie erfolgte in zwei Schritten.

Zuerst wurde die methodische Qualität der eingeschlossenen Studien endpunktübergreifend auf Studienebene untersucht (Anhang 4-E). In die Beurteilung sind Aspekte der Studienmethodik, die Erzeugung der Randomisierungssequenz, die Verdeckung der Gruppenzuteilung, die Verblindung der Patienten und Behandler, aber auch Verzerrungen bei der Berichterstattung und das Vorhandensein anderer verzerrender Aspekte eingeflossen. Anschließend wurde das Verzerrungspotenzial auf Studienebene als *niedrig* oder *hoch* eingestuft. Eine Einteilung als *niedrig* wurde vorgenommen, wenn eine Verzerrung der Ergebnisse unwahrscheinlich erschien und keine Anhaltspunkte für verzerrende Aspekte vorhanden waren, bei deren Behebung die Grundaussage der Ergebnisse in relevanter Weise verändert würde.

Im nachfolgenden Schritt wurden das Verzerrungspotenzial und damit die Aussagekraft der Ergebnisse auf Endpunktebene beurteilt. Hierbei wurden Aspekte u. a. zur Verblindung der Endpunkterheber, Analyse der Daten, Berichterstattung und weitere potenziell verzerrende Punkte berücksichtigt. Das Verzerrungspotenzial wurde als *niedrig* eingestuft, wenn keine Anhaltspunkte für Verzerrungen vorliegen, bei deren Behebung die Aussage der Ergebnisse in relevanter Weise verändert würde.

Die eingeschlossene Studie RA-BEAM (JADV) wurde anhand des CONSORT-Statements 2010 (Items 2b bis 14) beschrieben und in einem *Flow-Chart* dargestellt (Anhang 4-E). In diesem Rahmen erfolgte eine Beschreibung von Studienziel, Methodik (Design, Charakteristika der Studienteilnehmer, Intervention(en), Zielkriterien, Fallzahl, Erzeugung und Geheimhaltung der Randomisierungssequenz (*Allocation Concealment*), Durchführung der Randomisierung, Verblindung, statistische Methoden) und der Ergebnisdarstellung (Patientenfluss, Anzahl Studienteilnehmer, Aufnahme/Rekrutierung). Die Bewertungsgrundlage bildeten das Studienprotokoll, der statistische Analyseplan und der Studienbericht der Studie mit Baricitinib.

Die Ergebnisse der in die Bewertung einbezogenen patientenrelevanten Endpunkte wurden den in der Verfahrensordnung vorgegebenen Nutzendimensionen *Mortalität*, *Morbidität*, *Lebensqualität* und *Sicherheit* zugeordnet und zusammenfassend beschrieben. Ergänzend wurden Ergebnisse der patientenrelevanten Endpunkte aus Subgruppenanalysen dargestellt (Abschnitt 4.3.1.3.4).

Auf Endpunktebene erfolgte eine Bewertung der Aussagekraft der Nachweise, mit einer Einstufung in die Kategorien *Beleg*, *Hinweis* oder *Anhaltspunkt* (Abschnitt 4.4). Das Ausmaß des Zusatznutzens und die therapeutische Bedeutung des Zusatznutzens gegenüber dem Nutzen der zweckmäßigen Vergleichstherapie unter Berücksichtigung des Schweregrades der Erkrankung wurden gemäß Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) [3] quantifiziert.

Jeder in der Nutzen- bzw. Zusatznutzenbeurteilung betrachtete Endpunkt wurde als patientenrelevant eingestuft (Abschnitt 4.2.5.2).

Informationssynthese und Analyse

Die Ergebnisse der patientenrelevanten Endpunkte der Studie RA-BEAM (JADV) wurden zusammenfassend dargestellt. Hierzu wurden die Ergebnisse der in die Bewertung einbezogenen Endpunkte den in der Verfahrensordnung vorgegebenen Dimensionen *Mortalität*, *Morbidität*, *Lebensqualität* und *Sicherheit* zugeordnet und zusammenfassend beschrieben. Ergänzend wurden Ergebnisse der patientenrelevanten Endpunkte von Subgruppenanalysen dargestellt (Abschnitt 4.3.1.3.4). Auf Endpunktebene erfolgte eine Bewertung der Aussagekraft der Nachweise mit einer Einstufung in die Kategorien *Beleg*, *Hinweis* oder *Anhaltspunkt* (Abschnitt 4.4).

Ergebnisse zum medizinischen Nutzen und medizinischen Zusatznutzen

Die Zielpopulation im Anwendungsgebiet A wurde vom G-BA in die Subpopulationen A1, A2 und A3 unterteilt.

Für die Subpopulation A1, für Patienten, bei denen **keine ungünstigen Prognosefaktoren**¹ vorliegen und die unzureichend auf eine vorangegangene Behandlung mit **einem** cDMARD (inklusive MTX) ansprachen, konnte keine randomisierte kontrollierte Studie (*Randomized Controlled Trial*, RCT) identifiziert werden. Daher liegt keine direkte Evidenz zum Vergleich von Baricitinib mit der vom G-BA genannten zVT cDMARDs als Mono- oder Kombinationstherapie vor.

Für die Subpopulationen A2 und A3, d. h. Patienten **mit ungünstigen Prognosefaktoren**¹, die unzureichend auf eine vorangegangene Behandlung mit **einem** cDMARD (inklusive MTX) ansprachen (A2) bzw. für Patienten die unzureichend auf eine vorangegangene Behandlung mit **mehreren** cDMARDs (inklusive MTX) ansprachen (A3), wurde die RCT RA-BEAM (JADV) eingeschlossen, in welcher patientenrelevante Endpunkte zu *Mortalität, Morbidität, Lebensqualität* und *Sicherheit* erhoben wurden. Dabei wird die direkte Evidenz von Baricitinib gegenüber einer der vom G-BA genannten zVT, Adalimumab, bewertet. Die Aussagekraft der eingeschlossenen RCT ist hoch und damit für eine valide Beurteilung des medizinischen Nutzens und Zusatznutzens uneingeschränkt geeignet. Aufgrund der hohen Ergebnissicherheit ist die Studie RA-BEAM (JADV) dem Evidenzlevel Ib zuzuordnen. Die Studiendauer dieser abgeschlossenen Studie betrug 52 Wochen.

Aus Sicht der *Lilly Deutschland GmbH* besteht aufgrund der Leitlinienempfehlungen und der Bestimmung der zVT durch den G-BA die Notwendigkeit, für den Nachweis des Zusatznutzens von Baricitinib in den Subpopulationen A2 und A3 die gesamte Studienpopulation (*modified Intention-To-Treat* (mITT) Population) der Zulassungsstudie RA-BEAM (JADV) heranzuziehen.

Bei den Subpopulationen A2 und A3 handelt es sich um Patienten, die gemäß den Leitlinien für eine Zweitlinientherapie mit bDMARDs in Frage kommen. Die Leitlinien geben dabei eine Empfehlung, wann die Eskalation indiziert ist. So ist bei Patienten mit mehreren cDMARDs in der Vortherapie oder einem cDMARD in der Vortherapie und gleichzeitigem Vorliegen ungünstiger Prognosefaktoren eine Zweitlinientherapie mit bDMARDs indiziert. Konsequenterweise hat der G-BA für diese Patienten die gleiche zVT (bDMARDs in Kombination mit MTX (Adalimumab oder Etanercept oder Certolizumab-Pegol oder Golimumab oder Abatacept oder Tocilizumab); ggf. als Monotherapie unter Berücksichtigung des jeweiligen Zulassungsstatus bei MTX-Unverträglichkeit) festgelegt.

¹ Beispiele für ungünstige Prognosefaktoren:

- Nachweis von Autoantikörpern (z. B. Rheumafaktoren, hohe Spiegel von Antikörpern gegen citrullinierte Peptid Antigene)
- Hohe Krankheitsaktivität (nachgewiesen durch DAS- bzw. DAS28-Bewertungssystem, geschwollene Gelenke, Parameter der Akute-Phase-Reaktion wie z. B. C-Reaktives Protein, Erythrozytensedimentationsrate)
- Frühes Auftreten von Gelenkerosionen.

Quelle: [2].

Dabei empfiehlt die Leitlinie in jedem Fall ein individuelles Vorgehen, d. h. neben der aktuellen Therapiesituation und dem Vorliegen von ungünstigen Prognosefaktoren ist immer die jeweilige individuelle Patientensituation zu berücksichtigen [4]. Um das Therapieziel Remission oder niedrige Krankheitsaktivität zu erreichen, gibt die Leitlinie somit eine Entscheidungshilfe bei der Fragestellung des Beginns der Zweitlinientherapie.

Die Studie RA-BEAM (JADV) vergleicht Baricitinib mit der vom G-BA bestimmten zVT (in diesem Fall Adalimumab) bei jenen Patienten, für die gemäß der S1-Leitlinie eine erstmalige Behandlung mit Biologika vorgesehen ist [5]. Die Studie umfasst Eskalations-würdige Patienten und ist somit in ihrer Gesamtheit geeignet, die Fragestellung nach einem Zusatznutzen von Baricitinib gegenüber dem *Standard of Care* in der Zweitlinie zu beantworten. Ein Auftrennen der Studie generiert für die Nutzenbewertung keine zusätzliche Information für die Bewertung der Zweitlinientherapie.

Vor diesem Hintergrund wurde für jeden Endpunkt der RA-BEAM (JADV) untersucht, ob der Behandlungseffekt zwischen den verschiedenen Subpopulationen A2 und A3 heterogen, d. h. unterschiedlich ist (Abschnitt 4.3.1.3). Da diese Heterogenitätstests an Woche 52 unauffällig waren und demnach keine Heterogenität zeigten, gehen wir davon aus, dass die über die gesamte Studienpopulation zu beobachtenden Effekte auch auf die verschiedenen Subpopulationen übertragbar sind. Demnach können die Ergebnisse der mITT Population herangezogen werden, um den Zusatznutzen von Baricitinib in der Zweitlinientherapie abzuleiten.

Für die Subpopulationen A2 und A3 wurden lediglich die Patienten berücksichtigt, die gemäß Label Baricitinib in Kombination mit MTX erhalten hatten. Die mITT Population schließt neben diesen Patienten auch Patienten ein, die während der Studie neben MTX noch andere cDMARDs erhalten haben. Der prozentuale Anteil dieser Patienten in der Studienpopulation (15,5%) liegt unter dem vom Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) definierten Schwellenwert zum Einschluss von Studien (20%) (IQWiG Methodenpapier; Abschnitt 8.1.1 [6]) und erfüllt somit dieses Kriterium. Daher kann bei der Ableitung eines Zusatznutzens unter Heranziehung der mITT Population die vollständige Studienpopulation verwendet werden, ohne der Notwendigkeit Patienten herauszurechnen, die zusätzlich zu MTX noch weitere cDMARDs erhalten hatten.

Im Ergebnisteil (Abschnitt 4.3.1.3) werden neben der mITT der Vollständigkeit halber die Ergebnisse der Subpopulationen A2 und A3 ergänzend dargestellt.

Die Bewertung erfolgte anhand der patientenrelevanten Endpunkte der Nutzendimensionen *Mortalität*, *Morbidität*, *Lebensqualität* sowie *Sicherheit*, mit denen diese Kategorien in der zur Bewertung herangezogenen Studie operationalisiert wurden.

Die Daten der Ergebnisse der Subgruppenanalysen wurden dem Studienprotokoll, dem statistischen Analyseplan, dem Studienbericht und dem *post-hoc* Nachberechnungsdokument zur Studie RA-BEAM (JADV) entnommen [7-11].

Mortalität

Da *Mortalität* nicht als primärer oder sekundärer Endpunkt vorgesehen war und nicht im Rahmen einer Überlebenszeitanalyse untersucht wurde, wurde die Anzahl der Todesfälle im Rahmen der schwerwiegenden unerwünschten Ereignisse (SUE (*Serious Adverse Events*, SAE)) als Annäherung zur Beurteilung dieses patientenrelevanten Endpunktes ausgewertet (Abschnitt 4.3.1.3.4).

Morbidität

Um das primäre Ziel der Behandlung der RA, die Maximierung der langfristigen gesundheitsbezogenen Lebensqualität durch eine Kontrolle der Symptome, der Prävention von strukturellen Schäden, eine Normalisierung der körperlichen Funktion sowie der Teilhabe am sozialen Leben und an der Arbeit, zu erreichen, wird bei der Behandlung der RA die Remission (insbesondere in der frühen RA) oder das Erreichen einer niedrigen Krankheitsaktivität (hauptsächlich bei längerfristiger Erkrankung) angestrebt [12-13]. Die klinische Remission ist definiert als die Abwesenheit von Anzeichen und Symptomen signifikanter entzündlicher Krankheitsaktivität [14]. Das *American College of Rheumatology* (ACR) und die *European League Against Rheumatism* (EULAR) haben 2011 neue Remissionskriterien entwickelt (Tabelle 4-4) [15]. Weiterhin ist es von hoher Wichtigkeit, das strukturelle Voranschreiten der Erkrankung zu evaluieren. Die Behandlung der RA sollte strukturelle Veränderung verhindern oder aufhalten und dadurch körperliche Einschränkungen minimieren oder umkehren [15]. Zur Beurteilung der Morbidität bzw. Krankheitsaktivität, Symptomatik der RA und des körperlichen Funktionsstatus, werden daher die folgenden patientenrelevanten Endpunkte herangezogen.

Druckschmerzhaftige Gelenke und geschwollene Gelenke

Gelenkentzündungen und die damit einhergehende Schwellung (Synovitis) und Druckschmerzhaftigkeit der betroffenen Gelenke sind Hauptmerkmale der RA und zudem wichtige und für den Patienten unmittelbar spürbare Symptome. Sie sind verbunden mit körperlichen Beeinträchtigungen, die folglich mit Einschränkungen des alltäglichen Lebens einhergehen.

Veränderung der Anzahl der druckschmerzhaften Gelenke und geschwollenen Gelenke (basierend auf 28 Gelenken)

In der mITT Population nahm die Anzahl der druckschmerzhaften Gelenke und auch die Anzahl der geschwollenen Gelenke unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab stärker ab. Dieser Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung der Symptome Gelenkentzündung und der damit einhergehenden Schwellung.

In der mITT Population nahm die Anzahl der druckschmerzhaften und auch die Anzahl der geschwollenen Gelenke im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter ab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-25, Tabelle 4-26).

In der Subpopulation A2 nahm die Anzahl der druckschmerzhaften Gelenke und auch die Anzahl der geschwollenen Gelenke unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab stärker ab. In der Subpopulation A3 nahm die Anzahl der druckschmerzhaften Gelenke von Baseline bis Woche 24 unter Adalimumab und von Baseline bis Woche 52 unter Baricitinib stärker ab. Die Anzahl der geschwollenen Gelenke nahm in der Subpopulation A3 unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab stärker ab. Diese Unterschiede waren in den Subpopulationen A2 und A3 statistisch nicht signifikant.

In der Subpopulation A2 nahm die Anzahl der geschwollenen Gelenke im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter ab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-25, Tabelle 4-26).

Disease Activity Score 28 (DAS28) High Sensitivity C-Reactive Protein (hsCRP)

Die Krankheitsaktivität gemessen mittels DAS28 wird gemäß ACR und EULAR in verschiedene Stufen eingeteilt. Bei einem DAS28-Score von $< 2,6$ liegt eine Remission vor [16]. Des Weiteren erfolgt die Einteilung der Krankheitsaktivität mittels DAS28 in die Stufen gering ($\geq 2,6$ bis $< 3,2$), mittelschwer ($\geq 3,2$ bis $\leq 5,1$) und hoch ($> 5,1$) [15]. Besonders eine Remission aber auch das Erreichen einer niedrigen Krankheitsaktivität, die die Lebensqualität, soziale Teilhabe und die radiologisch nachweisbare Gelenkschädigung (radiologische Progression) jeweils positiv beeinflussen [17], stellen wichtige und anerkannte Therapieziele dar [16,18-31]. Auch im IQWiG Bericht „*Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis*“ (Kapitel 4.1.3) ist die DAS28-Remission als patientenrelevanter Endpunkt anerkannt [5].

Veränderung des DAS28-hsCRP

In der mITT Population konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Dieser Unterschied war sowohl von Baseline bis Woche 24 als auch bis Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Reduktion der Krankheitsaktivität (Tabelle 4-29).

In den Subpopulationen A2 und A3 konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant. (Tabelle 4-29).

Anteil Patienten mit DAS28-hsCRP $\leq 3,2$

In der mITT Population erreichte ein höherer Anteil der Patienten von Baseline bis Woche 24 und Woche 52 eine niedrige Krankheitsaktivität unter Baricitinib als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-32).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten von Baseline bis Woche 24 und Woche 52 eine niedrige Krankheitsaktivität unter Baricitinib als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-32).

Anteil Patienten mit DAS28-hsCRP $< 2,6$

In der mITT Population erreichte ein höherer Anteil der Patienten von Baseline bis Woche 24 und Woche 52 unter Baricitinib eine Remission als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-33).

In der Subpopulation A2 erreichte ein höherer Anteil der Patienten von Baseline bis Woche 24, jedoch von Baseline bis Woche 52 ein niedrigerer Anteil der Patienten unter Baricitinib eine Remission als unter Adalimumab. In der Subpopulation A3 erreichte ein höherer Anteil der Patienten von Baseline bis Woche 24 und Woche 52 unter Baricitinib eine Remission als unter Adalimumab. Der Unterschied war für die Subpopulationen A2 und A3 zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-33).

Simplified Disease Activity Index (SDAI)

Der SDAI ist ein zusammengesetztes und validiertes Konstrukt, das die Krankheitsaktivität der RA erfasst. Der SDAI besteht aus patientenrelevanten Einzelkomponenten (Anzahl druckschmerzhafter Gelenke (*Tender Joint Count*, TJC), Anzahl geschwollener Gelenke (*Swollen Joint Count*, SJC) sowie *Patient's Global Assessment of Disease Activity* (PtGADA)) und dem Akut-Phase-Protein hsCRP und bezieht zusätzlich Informationen zur Einschätzung der Krankheitsaktivität durch den Arzt (*Physician's Global Assessment of Disease Activity*, PhGADA) mit ein. Eine Remission liegt gemäß ACR und EULAR bei einem SDAI-Score von $\leq 3,3$ vor [15,20,27,32]. Des Weiteren wird die Krankheitsaktivität in die Stufen gering ($> 3,3$ bis ≤ 11), mittelschwer (> 11 bis ≤ 22) und hoch (> 26) eingeteilt [15].

Veränderung des SDAI

In der mITT Population konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Es lag ein statistisch signifikanter Unterschied zu beiden Zeitpunkten vor. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Reduktion der Krankheitsaktivität.

In der mITT Population nahm die Krankheitsaktivität im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter ab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-36).

In den Subpopulationen A2 und A3 konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Dieser Unterschied zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-36).

In den Subpopulationen A2 und A3 nahm die Krankheitsaktivität im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter ab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-36).

Anteil Patienten mit SDAI $\leq 3,3$

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine Remission als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-37).

In den Subpopulationen A2 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine Remission als unter Adalimumab. In der Subpopulation A3 erreichte ein niedrigerer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24, jedoch bis Woche 52 ein höherer Anteil Patienten unter Baricitinib eine Remission. Der Unterschied war für die Subpopulationen A2 und A3 zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-37).

Clinical Disease Activity Index (CDAI)

Ebenso wie der SDAI ist der CDAI ein validiertes Konstrukt, mit dem das komplexe Krankheitsgeschehen der RA abgebildet wird. Ein CDAI-Score von $\leq 2,8$ zeigt gemäß ACR und EULAR eine Remission an [15,20,27,32]. Des Weiteren wird die Krankheitsaktivität in die Stufen gering ($> 2,8$ bis ≤ 10), mittelschwer (> 10 bis ≤ 22) und hoch (> 22) eingeteilt [15]. Der CDAI besteht analog zu dem SDAI aus den patientenrelevanten Einzelkomponenten Anzahl TJC/SJC sowie PtGADA und PhGADA. Das Akute-Phase-Protein hsCRP wird im Gegensatz zum SDAI für die Berechnung des CDAI nicht berücksichtigt. Der CDAI besitzt wie der SDAI eine hohe Patientenrelevanz [15,33-34].

Veränderung des CDAI

In der mITT Population konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Es lag ein statistisch signifikanter Unterschied zu beiden Zeitpunkten vor. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Reduktion der Krankheitsaktivität.

In der mITT Population nahm die Krankheitsaktivität im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter ab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-40).

In den Subpopulationen A2 und A3 konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Dieser Unterschied war für die Subpopulationen A2 von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Für die Subpopulationen A3 war der Unterschied zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-40).

In den Subpopulationen A2 und A3 nahm die Krankheitsaktivität im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter ab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-40).

Anteil Patienten mit CDAI \leq 10

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine niedrige Krankheitsaktivität als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-41).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine niedrige Krankheitsaktivität als unter Adalimumab. Der Unterschied war für die Subpopulationen A2 und A3 zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-41).

EULAR Responsekriterien („good Response“, DAS28-basiert)

Mit Hilfe der validierten EULAR Responsekriterien [35] wird das Therapieansprechen eines Patienten auf Grundlage der relativen Veränderung und des erreichten Endwertes (Post-Baseline) des DAS28 beurteilt und klassifiziert einen Patienten somit als *Responder* oder *Non-Responder* (Abschnitt 4.2.5.2.)

Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 ein gutes Therapieansprechen gemäß EULAR-Kriterien als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-44).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 ein gutes Therapieansprechen gemäß EULAR-Kriterien als unter Adalimumab. Der Unterschied war in beiden Subpopulationen zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-44).

Patient Assessment of Pain (PAP) (Visual Analogue Scale, VAS)

Chronische Schmerzen sind eine der Hauptkomponenten der Krankheitslast für RA-Patienten. Die Patienten leiden unter stärkeren Schmerzen als die Allgemeinbevölkerung und erfahren ein Schmerzniveau, das mit dem von Schmerzpatienten vergleichbar ist. Höhere Schmerzniveaus korrelieren mit der Funktionseinschränkung sowie mit Depressionen, was sich negativ auf die Lebensqualität der Patienten auswirkt. Auch wenn eine Kontrolle der Schmerzen ein wichtiges Kriterium einer erfolgreichen Therapie darstellt, ist ein Großteil der Patienten trotz einer Therapie noch immer bedeutsamen Schmerzen ausgesetzt [36-40]. Schmerz ist ein unmittelbar vom Patienten wahrgenommener und somit patientenrelevanter Endpunkt. Schmerzen beeinträchtigen das tägliche Wohlempfinden und haben somit einen erheblichen Einfluss auf die Lebensqualität der Betroffenen [41-44].

Veränderung des PAP (VAS)

In der mITT Population wurden von den Patienten unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 weniger Schmerzen berichtet. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant.

In der mITT Population kam es im Studienverlauf, d. h. von Woche 24 bis Woche 52, zu einer weiteren Verringerung der Schmerzen unter Baricitinib im Vergleich zu Adalimumab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-47).

In den Subpopulationen A2 und A3 wurden von den Patienten unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 weniger Schmerzen berichtet. Für die Subpopulation A2 war der Unterschied zu beiden Zeitpunkten statistisch nicht signifikant. Für die Subpopulation A3 war dieser Unterschied von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen.

In den Subpopulationen A2 und A3 kam es im Studienverlauf, d. h. von Woche 24 bis Woche 52, zu einer weiteren Verringerung der Schmerzen unter Baricitinib im Vergleich zu Adalimumab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-47).

Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Dieser Unterschied war von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen (Tabelle 4-48).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Dieser Unterschied war in beiden Subpopulationen von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen (Tabelle 4-48).

Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Der Unterschied war in der mITT Population von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen (Tabelle 4-49).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. In der Subpopulation A2 war dieser Unterschied zu beiden Zeitpunkten statistisch nicht signifikant. In der Subpopulation A3 war der Unterschied von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen (Tabelle 4-49).

Patient's Global Assessment of Disease Activity (PtGADA) (VAS)

Der PtGADA spiegelt die globale Einschätzung der Krankheitsaktivität durch den Patienten wider und gibt Aufschluss über den Krankheitszustand bzw. der damit verbundenen Lebensqualität.

Veränderung des PtGADA (VAS)

In der mITT Population wurde aus Sicht des Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität berichtet als unter Adalimumab. In der mITT Population war dieser Unterschied zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität.

In der mITT Population kam es im Studienverlauf, d. h. von Woche 24 bis Woche 52, zu einer zunehmenden Verbesserung der Krankheitsaktivität aus Sicht des Patienten unter Baricitinib im Vergleich zu Adalimumab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-52).

In den Subpopulationen A2 und A3 wurde aus Sicht des Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität berichtet als unter Adalimumab. Für die Subpopulationen A2 und A3 war der Unterschied von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant.

In den Subpopulationen A2 und A3 kam es im Studienverlauf, d. h. von Woche 24 bis Woche 52, zu einer zunehmenden Verbesserung der Krankheitsaktivität aus Sicht des Patienten unter Baricitinib im Vergleich zu Adalimumab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-52).

Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-53).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Dieser Unterschied war in beiden Subpopulationen zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-53).

Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Dieser Unterschied war zu Woche 52, jedoch nicht zu Woche 24, statistisch signifikant (Tabelle 4-54).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Dieser Unterschied war in beiden Subpopulationen zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-54).

Health Assessment Questionnaire-Disability Index (HAQ-DI)

Der HAQ-DI gibt Auskunft über die krankheitsbedingten körperlichen Einschränkungen im Alltag, die vom Patienten selbst bewertet werden und steht somit im Zusammenhang mit der Krankheitslast als Maß für die Morbidität.

Veränderung des HAQ-DI

In der mITT Population waren die krankheitsassoziierten Einschränkungen der physischen Funktionalität unter Baricitinib von Baseline bis Woche 24 und Woche 52 niedriger als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung der physischen Funktionalität (Tabelle 4-57).

In den Subpopulationen A2 und A3 waren die krankheitsassoziierten Einschränkungen der physischen Funktionalität unter Baricitinib von Baseline bis Woche 24 und Woche 52 niedriger als unter Adalimumab. In den beiden Subpopulationen war dieser Unterschied zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-57).

Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine bedeutsame Verbesserung der physischen Funktionen als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung der physischen Funktionalität (Tabelle 4-60).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine bedeutsame Verbesserung der physischen Funktionen als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten für beide Subpopulationen statistisch nicht signifikant (Tabelle 4-60).

Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine bedeutsame Verbesserung der physischen Funktionen als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-61).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine bedeutsame Verbesserung der physischen Funktionen als unter Adalimumab. Dieser Unterschied war in beiden Subpopulationen zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-61).

Dauer der Morgensteifigkeit (min)

Morgensteifigkeit (*Morning Joint Stiffness*, MJS) beschreibt eine Einschränkung der Beweglichkeit der Gelenke, die direkt nach dem Aufwachen von den Patienten wahrgenommen wird. Sie führt zu einer starken Einschränkung der üblichen morgendlichen Tätigkeiten, wie beispielsweise dem Aufstehen und der Körperpflege. Bei der Morgensteifigkeit handelt es sich um ein typisches Begleitsymptom der RA und die Dauer der Morgensteifigkeit korreliert mit der körperlichen Einschränkung des Patienten und dessen Lebensqualität [45]. Es handelt sich dabei um einen patientenrelevanten Endpunkt.

Veränderung der Dauer der Morgensteifigkeit (min)

In der mITT Population verkürzte sich die Dauer der Morgensteifigkeit unter Baricitinib von Baseline bis Woche 24 um 55 min und bis Woche 52 um 61 min, hingegen unter Adalimumab nur um 34 min (beide Zeitpunkte). Der Unterschied zwischen den Behandlungsarmen Baricitinib und Adalimumab betrug 21 min zu Woche 24 und 27 min zu Woche 52 mit statistisch signifikantem Effekt zugunsten von Baricitinib zu Woche 52. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verkürzung der Dauer der Morgensteifigkeit.

In der mITT Population verkürzte sich die Dauer der Morgensteifigkeit im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-64).

In der Subpopulation A2 verkürzte sich die Dauer der Morgensteifigkeit unter Baricitinib von Baseline bis Woche 24 um 72 min und bis Woche 52 um 71 min, hingegen unter Adalimumab nur um 17 min bzw. um 29 min. Der Unterschied zwischen den Behandlungsarmen Baricitinib und Adalimumab betrug 54 min zu Woche 24 und 41 min zu Woche 52 mit statistisch signifikantem Effekt zugunsten von Baricitinib zu beiden Zeitpunkten. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verkürzung der Dauer der Morgensteifigkeit (Tabelle 4-64).

In der Subpopulation A3 verkürzte sich die Dauer der Morgensteifigkeit unter Baricitinib von Baseline bis Woche 24 um 55 min und bis Woche 52 um 68 min und unter Adalimumab um 50 min bzw. 44 min. Der Unterschied zwischen den Behandlungsarmen Baricitinib und Adalimumab betrug 6 min zu Woche 24 zugunsten von Adalimumab und 24 min zu Woche 52 zugunsten von Baricitinib. Diese Unterschiede waren statistisch nicht signifikant.

In der Subpopulation A3 verkürzte sich die Dauer der Morgensteifigkeit im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-64).

European Quality of Life-5 Dimensions-5 (EQ-5D) VAS

Mittels EQ-5D, einem krankheitsübergreifenden und validierten Fragebogen [46], kann die gesundheitsbezogene Lebensqualität eines Patienten ermittelt werden. Mittels EQ-5D VAS bewertet der Patient selbst seinen allgemeinen Gesundheitszustand und kategorisiert diesen auf einer Skala von null „schlechtester vorstellbarer Gesundheitszustand“ bis 100 „bester vorstellbarer Gesundheitszustand“. Eine Zunahme des EQ-5D VAS Wertes spiegelt eine Verbesserung des Gesundheitszustandes des Patienten wider.

Veränderung der EQ-5D VAS

In der mITT Population verbesserte sich der allgemeine Gesundheitszustand von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung des allgemeinen Gesundheitszustandes.

In der mITT Population nahm die Verbesserung der EQ-5D VAS im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-67).

In den Subpopulationen A2 und A3 verbesserte sich der allgemeine Gesundheitszustand von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung des allgemeinen Gesundheitszustandes.

In den Subpopulationen A2 und A3 nahm die Verbesserung der EQ-5D VAS im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-67).

Anteil Patienten mit einer Verbesserung des EQ-5D VAS ≥ 10 mm

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung des allgemeinen Gesundheitszustandes als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung des allgemeinen Gesundheitszustandes (Tabelle 4-68).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung des allgemeinen Gesundheitszustandes als unter Adalimumab. Für die Subpopulation A2 war dieser Unterschied von Baseline bis Woche 24, jedoch nicht bis Woche 52, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung des allgemeinen Gesundheitszustandes. Für die Subpopulation A3 waren die Unterschiede zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-68).

Schwerste Müdigkeit (Worst Tiredness)

Fatigue ist ein wiederkehrendes, stark beeinträchtigendes Problem für RA-Patienten [47] und wird von Patienten, nach dem Schmerz, als das belastendste Krankheitssymptom beschrieben [48], welches einen großen Einfluss auf die Lebensqualität der Patienten ausübt [49-50]. Während die Fatigue ein multi-dimensionales Symptomkonzept darstellt [51-54], ist die Müdigkeit eine Schlüsselkomponente dieses Symptomkonzeptes.

Veränderung der schwersten Müdigkeit

In der mITT Population verbesserte sich das Symptom Müdigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib im Vergleich zu Adalimumab. In der mITT Population lag ein statistisch signifikanter Unterschied von Baseline bis Woche 52, jedoch nicht bis Woche 24, vor. Dies spricht für eine Überlegenheit von Baricitinib in hinsichtlich der Verbesserung des Symptoms Müdigkeit.

In der mITT Population nahm die Verbesserung des Symptoms Müdigkeit im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-71).

In den Subpopulationen A2 und A3 verbesserte sich das Symptom Müdigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Für die Subpopulation A2 war dieser Unterschied zu beiden Zeitpunkten statistisch nicht signifikant. In der Subpopulation A3 lag ein statistisch signifikanter Unterschied von Baseline bis Woche 52, jedoch nicht bis Woche 24, vor. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Verbesserung des Symptoms Müdigkeit.

In den Subpopulationen A2 und A3 nahm die Verbesserung des Symptoms Müdigkeit im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-71).

Lebensqualität

Short Form (36 item) Health Survey Version 2 akut (SF-36 v2 akut)

Der SF-36 ist ein generischer und international anerkannter Fragebogen zur subjektiven Einschätzung der Lebensqualität durch den Patienten [55]. Dieses Messinstrument ist für die RA validiert und wurde in zahlreichen Studien verwendet [56-57].

Veränderung des SF-36 v2 akut Physical Component Score (PCS)

In der mITT Population verbesserte sich die physische Komponente des allgemeinen Gesundheitszustandes sowohl von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes und der Lebensqualität (Tabelle 4-74).

In den Subpopulationen A2 und A3 verbesserte sich die physische Komponente des allgemeinen Gesundheitszustandes sowohl von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. In der Subpopulation A2 war dieser Unterschied von Baseline bis Woche 52, jedoch bis Woche 24, statistisch signifikant. In der Subpopulation A3 lag ein statistisch signifikanter Unterschied von Baseline bis Woche 24, jedoch nicht bis Woche 52, vor. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes und der Lebensqualität (Tabelle 4-74).

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes (Tabelle 4-75).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. In beiden Subpopulationen war dieser Unterschied zu keinem Zeitpunkt statistisch signifikant (Tabelle 4-75).

Veränderung des SF-36 v2 akut Mental Component Score (MCS)

In der mITT Population verbesserte sich die mentale Komponente des allgemeinen Gesundheitszustandes sowohl von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant.

In der mITT Population nahm die Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-76).

In der Subpopulationen A2 verbesserte sich die mentale Komponente des allgemeinen Gesundheitszustandes von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Der Unterschied war für diese Population zu beiden Zeitpunkten statistisch nicht signifikant. In der Subpopulation A3 verbesserte sich die mentale Komponente des allgemeinen Gesundheitszustandes von Baseline bis Woche 24 zugunsten von Adalimumab. Dieser Unterschied war statistisch nicht signifikant. Zu Woche 52 gab es in der Subpopulation A3 keine Unterschiede zwischen den Behandlungsarmen.

In der Subpopulation A2 nahm die Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-76).

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-77).

In der Subpopulation A2 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Dieser Unterschied war zu Woche 24, jedoch nicht zu Woche 52, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine klinisch relevante Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes. In der Subpopulation A3 erreichte ein höherer Anteil der Patienten unter Adalimumab von Baseline bis Woche 24 und bis Woche 52 eine klinisch relevante Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Dieser Unterschied war zu keinem Zeitpunkt statistisch signifikant (Tabelle 4-77).

Veränderung der SF-36 v2 akut Domänen

SF-36 v2 akut Domäne – Körperliche Funktionsfähigkeit

In der mITT Population verbesserte sich die Domäne körperliche Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Funktionsfähigkeit und somit der Lebensqualität (Tabelle 4-78).

In den Subpopulationen A2 und A3 verbesserte sich die Domäne körperliche Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu keinem Zeitpunkt signifikant (Tabelle 4-78).

SF-36 v2 akut Domäne – Körperliche Rollenfunktion

In der mITT Population verbesserte sich die Domäne körperliche Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Rollenfunktion und somit der Lebensqualität.

In der mITT Population nahm die Verbesserung der körperlichen Rollenfunktion im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-79).

In den Subpopulationen A2 und A3 verbesserte sich die Domäne körperliche Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in der Subpopulation A2 von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Rollenfunktion und somit der Lebensqualität. In der Subpopulation A3 war dieser Unterschied zu keinem Zeitpunkt statistisch signifikant.

In den Subpopulationen A2 und A3 nahm die Verbesserung der körperlichen Rollenfunktion im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-79).

SF-36 v2 akut Domäne – Körperliche Schmerzen

In der mITT Population verbesserte sich die Domäne körperlichen Schmerzen von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war für beide Zeitpunkte statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Schmerzen und somit der Lebensqualität (Tabelle 4-80).

In den Subpopulationen A2 und A3 verbesserte sich die Domäne körperlichen Schmerzen von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in der Subpopulation A2 von Baseline bis Woche 52, jedoch nicht bis Woche 24, und in der Subpopulation A3 von Baseline bis Woche 24, jedoch nicht bis Woche 52, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Schmerzen und somit der Lebensqualität.

In der Subpopulation A2 nahm die Verbesserung der körperlichen Schmerzen im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-80).

SF-36 v2 akut Domäne – Allgemeiner Gesundheitsstatus

In der mITT Population verbesserte sich die Domäne allgemeiner Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung des allgemeinen Gesundheitszustandes und somit der Lebensqualität.

In der mITT Population nahm die Verbesserung des allgemeinen Gesundheitszustandes im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-81).

In den Subpopulationen A2 verbesserte sich die Domäne allgemeiner Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in der Subpopulation A2 zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung des allgemeinen Gesundheitszustandes und somit der Lebensqualität. In der Subpopulation A3 verbesserte sich der allgemeine Gesundheitszustand zu Woche 24 zugunsten von Adalimumab und zu Woche 52 zugunsten von Baricitinib. Dieser Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant.

In der Subpopulation A2 nahm die Verbesserung des allgemeinen Gesundheitszustandes im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-81).

SF-36 v2 akut Domäne – Vitalität

In der mITT Population verbesserte sich die Domäne Vitalität von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu Woche 52, jedoch nicht zu Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der Vitalität und somit der Lebensqualität.

In der mITT Population nahm die Verbesserung der Vitalität im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-82).

In den Subpopulationen A2 und A3 verbesserte sich die Domäne Vitalität von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in beiden Subpopulationen zu keinem Zeitpunkt statistisch signifikant.

In der Subpopulation A2 nahm die Verbesserung der Vitalität im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-82).

SF-36 v2 akut Domäne – Soziale Funktionsfähigkeit

In der mITT Population verbesserte sich die Domäne soziale Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Funktionsfähigkeit und somit der Lebensqualität (Tabelle 4-83).

In den Subpopulationen A2 und A3 verbesserte sich die Domäne soziale Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Für die Subpopulation A2 lag zu beiden Zeitpunkten ein statistisch signifikanter Unterschied vor. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Funktionsfähigkeit und somit der Lebensqualität. Der Unterschied war für die Subpopulation A3 zu beiden Zeitpunkten statistisch nicht signifikant.

In der Subpopulation A2 nahm die Verbesserung der sozialen Funktionsfähigkeit im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-83).

SF-36 v2 akut Domäne – Emotionale Rollenfunktion

In der mITT Population verbesserte sich die Domäne emotionale Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu Woche 52, jedoch nicht zu Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der emotionalen Rollenfunktion und somit der Lebensqualität.

In der mITT Population nahm die Verbesserung der emotionalen Rollenfunktion im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-84).

In der Subpopulationen A2 verbesserte sich die Domäne emotionalen Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in den Subpopulationen A2 zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der emotionalen Rollenfunktion und somit der Lebensqualität. In der Subpopulation A3 verbesserte sich die Domäne emotionale Rollenfunktion von Baseline bis Woche 24 zugunsten von Adalimumab und von Baseline bis Woche 52 zugunsten von Baricitinib. Dieser Unterschied war zu keinem Zeitpunkt statistisch signifikant (Tabelle 4-84).

SF-36 v akut Domäne – Mentaler Gesundheitszustand

In der mITT Population verbesserte sich die Domäne mentaler Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu keinem Zeitpunkt statistisch signifikant.

In der mITT Population nahm die Verbesserung des mentalen Gesundheitszustandes im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-85).

In den Subpopulationen A2 und A3 verbesserte sich die Domäne mentaler Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in beiden Subpopulationen zu keinem Zeitpunkt statistisch signifikant

In der Subpopulation A2 nahm die Verbesserung des mentalen Gesundheitszustandes im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-85).

Sicherheit

Die Bedeutung unerwünschter Arzneimittelwirkungen ist abhängig von ihrer Schwere, der Häufigkeit sowie der Reversibilität ihrer Symptome. Die Verringerung von unerwünschten Wirkungen ist ein wichtiger Aspekt der therapiebedingten Morbidität und ein vom IQWiG anerkannter patientenrelevanter Endpunkt [6].

Todesfälle

In der mITT Population traten an Woche 24 und Woche 52 vergleichbar wenige Todesfälle unter Baricitinib (Gesamtzahl Todesfälle bis Woche 52: n = 2) und unter Adalimumab (Gesamtzahl Todesfälle bis Woche 52: n = 1) auf. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-88).

In der Subpopulation A2 traten an Woche 24 und Woche 52 vergleichbar wenige Todesfälle unter Baricitinib (Gesamtzahl Todesfälle bis Woche 52: n = 2) und unter Adalimumab (Gesamtzahl Todesfälle bis Woche 52: n = 1) auf. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant.

In der Subpopulation A3 traten an Woche 24 und Woche 52 keine Todesfälle unter Baricitinib und unter Adalimumab auf (Tabelle 4-88).

Schwerwiegende Unerwünschte Ereignisse (SUE (Serious Adverse Events, SAE))

In der mITT Population traten mehr schwerwiegende unerwünschte Ereignisse an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab auf. Diese Unterschiede waren zu beiden Zeitpunkten statistisch signifikant (Tabelle 4-89).

Hierzu ist zu bemerken, dass an Woche 24 die Rate schwerwiegender unerwünschter Ereignisse unter Baricitinib auf dem gleichen Niveau wie unter Placebo war [9]. Das geringe Auftreten von schwerwiegenden unerwünschten Ereignissen unter Adalimumab im Vergleich zu Placebo ist überraschend angesichts vorangegangener Adalimumab Studien [58]. Die aufgetretenen SUE waren gleichmäßig über die *System Organ Classes* (SOCs) verteilt. In keiner SOC konnte ein statistisch signifikanter Unterschied zwischen der Baricitinib- und der Adalimumab-Behandlungsgruppe beobachtet werden.

In der Subpopulation A2 traten mehr schwerwiegende unerwünschte Ereignisse an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab auf. Diese Unterschiede waren zu beiden Zeitpunkten statistisch signifikant (Tabelle 4-89).

In der Subpopulation A3 unterschieden sich die schwerwiegenden unerwünschten Ereignisse in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander (Tabelle 4-89).

Unerwünschte Ereignisse (UE (Treatment-Emergent Adverse Events, TEAE))

In der mITT Population unterschied sich der Anteil der Patienten mit unerwünschten Ereignissen an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-90).

In den Subpopulationen A2 und A3 unterschied sich der Anteil der Patienten mit unerwünschten Ereignissen an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-90).

Schwere UE (TEAE rated as severe; ohne ausschließliche Laborereignisse)

Alle in der vorliegenden Nutzenbewertung aufgeführten schweren unerwünschten Ereignisse wurden ohne solche Laborereignisse ausgewertet, in deren Zusammenhang keine Symptome beschrieben wurden oder dies angenommen werden musste und die somit asymptomatisch waren (Abschnitt 4.2.5.2). Diese schweren UE waren ausschließlich Laborereignisse ohne assoziierte Symptomatik und daher für den Patienten nicht zu spüren, also nicht patientenrelevant. Sollte eine mit diesen Laborereignissen in Zusammenhang stehende Symptomatik aufgetreten sein, dann wurde diese Symptomatik innerhalb der unerwünschten Ereignisse erfasst.

In der mITT Population unterschied sich die Anzahl der Patienten mit mindestens einem schweren unerwünschten Ereignis an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-91).

In den Subpopulationen A2 und A3 unterschied sich die Anzahl der Patienten mit mindestens einem schweren unerwünschten Ereignis an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-91).

Studienabbruch aufgrund von UE

In der mITT Population unterschieden sich die Studienabbrüche aufgrund von unerwünschten Ereignissen an Woche 24 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander. An Woche 52 traten unter Baricitinib signifikant mehr Studienabbrüche aufgrund von unerwünschten Ereignissen auf als unter Adalimumab (Tabelle 4-92).

In Subpopulation A2 und Subpopulation A3 unterschieden sich die Studienabbrüche aufgrund von unerwünschten Ereignissen an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-92).

Therapieabbruch aufgrund von UE

In der mITT Population unterschieden sich die Therapieabbrüche aufgrund von unerwünschten Ereignissen an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-93).

In den Subpopulationen A2 und A3 unterschieden sich die Therapieabbrüche aufgrund von unerwünschten Ereignissen an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-93).

UE von besonderem Interesse

In der mITT Population konnte für keines der unerwünschten Ereignisse von besonderem Interesse an Woche 24 oder bis Woche 52 unter Baricitinib im Vergleich zu Adalimumab ein statistisch signifikanter Unterschied festgestellt werden (Tabelle 4-94).

In den Subpopulationen A2 und A3 konnte für keines der unerwünschten Ereignisse von besonderem Interesse an Woche 24 oder bis Woche 52 unter Baricitinib im Vergleich zu Adalimumab ein statistisch signifikanter Unterschied festgestellt werden (Tabelle 4-94).

Effektmodifikation in Subgruppen

Bei der Beurteilung der Ergebnisse von Subgruppenanalysen sollten die mit multiplem Testen in Zusammenhang stehende α -Fehler-Kumulierung, die die Zunahme der Wahrscheinlichkeit für Fehler 1. Art (falsch positive Testergebnisse) beschreibt, sowie bei der Interpretation der Resultate kleine Fallzahlen innerhalb der Subgruppen berücksichtigt werden.

Es wurden für 43 (Unter-) Endpunkte bis zu 18 Subgruppenanalysen durchgeführt. Unter der Annahme, dass keine Subgruppeneffekte vorliegen, wären daher zufallsbedingt ca. 38 – 39 Belege ($p < 0,05$) pro Population für eine Effektmodifikation zu erwarten.

In der mITT Population zeigten sich 37 Belege, in der Subpopulation A2 18 Belege sowie in der Subpopulation A3 9 Belege für eine Effektmodifikation.

Die Anzahl der beobachteten Effektmodifikationen liegt somit im oder deutlich unterhalb des Bereiches der zu erwarteten Ergebnisse. Auch sind keine auffälligen Häufungen über die Endpunkte oder Subgruppen hinweg zu beobachten. Auf Grund dessen wird insgesamt nicht von einer Modifikation des Behandlungseffekts ausgegangen (Abschnitt 4.3.1.3.5).

Zusammenfassung der Ergebnisse

Aus Sicht der *Lilly Deutschland GmbH* besteht aufgrund der Leitlinienempfehlungen und der Bestimmung der zVT durch den G-BA die Notwendigkeit, für den Nachweis des Zusatznutzens von Baricitinib in den Subpopulationen A2 und A3 die gesamte Studienpopulation (mITT Population) der Zulassungsstudie RA-BEAM (JADV) heranzuziehen.

Bei den Subpopulationen A2 und A3 handelt es sich um Patienten, die gemäß den Leitlinien für eine Zweitlinientherapie mit bDMARDs in Frage kommen. Die Leitlinien geben dabei eine Empfehlung, wann die Eskalation indiziert ist. So ist bei Patienten mit mehreren cDMARDs in der Vortherapie oder einem cDMARD in der Vortherapie und gleichzeitigem Vorliegen ungünstiger Prognosefaktoren eine Zweitlinientherapie mit bDMARDs indiziert. Konsequenterweise hat der G-BA für diese Patienten die gleiche zVT (bDMARDs in Kombination mit MTX (Adalimumab oder Etanercept oder Certolizumab-Pegol oder Golimumab oder Abatacept oder Tocilizumab); ggf. als Monotherapie unter Berücksichtigung des jeweiligen Zulassungsstatus bei MTX-Unverträglichkeit) festgelegt.

Dabei empfiehlt die Leitlinie in jedem Fall ein individuelles Vorgehen, d. h. neben der aktuellen Therapiesituation und dem Vorliegen von ungünstigen Prognosefaktoren ist immer die jeweilige individuelle Patientensituation zu berücksichtigen [4]. Um das Therapieziel Remission oder niedrige Krankheitsaktivität zu erreichen, gibt die Leitlinie somit eine Entscheidungshilfe bei der Fragestellung des Beginns der Zweitlinientherapie.

Die Studie RA-BEAM (JADV) vergleicht Baricitinib mit der vom G-BA bestimmten zVT (in diesem Fall Adalimumab) bei jenen Patienten, für die gemäß der S1-Leitlinie eine erstmalige Behandlung mit Biologika vorgesehen ist [5]. Die Studie umfasst Eskalations-würdige Patienten und ist somit in ihrer Gesamtheit geeignet, die Fragestellung nach einem Zusatznutzen von Baricitinib gegenüber dem „Standard of Care“ in der Zweitlinie zu beantworten. Ein Auftrennen der Studie generiert für die Nutzenbewertung keine zusätzliche Information für die Bewertung der Zweitlinientherapie.

Vor diesem Hintergrund wurde für jeden Endpunkt der RA-BEAM (JADV) untersucht, ob der Behandlungseffekt zwischen den verschiedenen Subpopulationen A2 und A3 heterogen, d. h. unterschiedlich ist (Abschnitt 4.3.1.3). Da diese Heterogenitätstests an Woche 52 unauffällig waren und demnach keine Heterogenität zeigten, gehen wir davon aus, dass die über die gesamte Studienpopulation zu beobachtenden Effekte auch auf die verschiedenen Subpopulationen übertragbar sind. Demnach können die Ergebnisse der mITT Population herangezogen werden, um den Zusatznutzen von Baricitinib in der Zweitlinientherapie abzuleiten¹.

Im Folgenden werden deshalb lediglich die Ergebnisse der mITT Population dargestellt. Die Ergebnisse der Subpopulationen A2 und A3 werden der Vollständigkeit halber im Ergebnisteil (Abschnitt 4.3.1.3) dargestellt.

¹ Der Anteil Patienten, die in der mITT Population neben einer Begleitbehandlung mit MTX ein weiteres cDMARD erhalten haben, liegt bei 15,5 % (127/817 Patienten). Da die Studie somit zu mindestens 80 % das Einschlusskriterium Prüfintervention erfüllt[6], ist sie grundsätzlich zur Ableitung eines Zusatznutzens in der hier adressierten Fragestellung geeignet.

Für die mITT Population konnte für folgende Endpunkte eine signifikante Überlegenheit entweder konsistent zu beiden Zeitpunkten oder zum ausschlaggebenden Zeitpunkt von 52 Wochen gegenüber der zVT Adalimumab festgestellt werden (Tabelle 4-1):

- Veränderung der Anzahl der druckschmerzhaften Gelenke
- Veränderung der Anzahl der geschwollenen Gelenke
- Veränderung des DAS28-hsCRP (mBOCF, mLOCF, MMRM)
- Veränderung des SDAI
- Veränderung des CDAI
- Veränderung des PAP (VAS)
- Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm
- Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm
- Veränderung des PtGADA (VAS)
- Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm
- Veränderung des HAQ-DI (mBOCF, mLOCF, MMRM)
- Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$
- Veränderung der Dauer der Morgensteifigkeit (min)
- Veränderung der EQ-5D VAS
- Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm
- Veränderung der schwersten Müdigkeit
- Veränderung des SF-36 v2 akut PCS
- Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5
- Veränderung der SF-36 v2 akut Domänen
 - Körperliche Funktionsfähigkeit
 - Körperliche Rollenfunktion
 - Körperliche Schmerzen

- Allgemeiner Gesundheitszustand
- Vitalität
- Soziale Funktionsfähigkeit
- Emotionale Rollenfunktion

Statistisch signifikante Ergebnisse zuungunsten von Baricitinib konnten für folgende Endpunkte entweder konsistent zu beiden Zeitpunkten oder zum ausschlaggebenden Zeitpunkt von 52 Wochen nachgewiesen werden:

- SUE
- Studienabbruch aufgrund von UE

Bezüglich der anderen patientenrelevanten Endpunkte der Nutzendimension *Sicherheit*, (Todesfälle, Unerwünschte Ereignisse, Therapieabbruch aufgrund von UE und UE von besonderem Interesse) erwiesen sich Baricitinib und Adalimumab als vergleichbar.

Gemäß der in Abschnitt 4.4.1 beschriebenen Methodik wurde der Zusatznutzen auf Endpunktebene für statistisch signifikante Ergebnisse für die mITT Population wie folgt abgeleitet (Tabelle 4-1).

Tabelle 4-1: Zusammenfassende Darstellung zum Ausmaß des Zusatznutzens auf Endpunktebene für die mITT Population

Endpunkt	Baricitinib vs. Adalimumab Effektschätzer ^a (95% KI); p-Wert		Ausmaß des Zusatznutzens ^f (Aussagesicherheit)
	Woche 24	Woche 52	
Mortalität^b	-	-	
Morbidität			
Druckschmerzhaftige Gelenke und geschwollene Gelenke			
Veränderung der Anzahl der druckschmerzhaften Gelenke (basierend auf 28 Gelenken)	-0,6 (-1,5; 0,2); 0,135	-0,9 (-1,6; -0,1); 0,032 Hedges' g: -0,1 (-0,3; 0,0)	geringer Zusatznutzen (hoch)
Veränderung der Anzahl der geschwollenen Gelenke (basierend auf 28 Gelenken)	-0,5 (-1,2; 0,2); 0,159	-0,9 (-1,5; -0,2); 0,007 Hedges' g: -0,2 (-0,3; -0,0)	geringer Zusatznutzen (hoch)
DAS28-hsCRP			
Veränderung des DAS28-hsCRP (mBOCF)	-0,21 (-0,39; -0,03); 0,022	-0,29 (-0,49; -0,08); 0,007	kein Zusatznutzen (hoch)
Veränderung des DAS28-hsCRP (mLOCF)	-	-0,35 (-0,54; -0,16); < 0,001	kein Zusatznutzen (hoch)
Veränderung des DAS28-hsCRP (MMRM)	-0,19 (-0,36; -0,02); 0,029	-0,23 (-0,41; -0,05); 0,014	kein Zusatznutzen (hoch)
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	1,08 (0,95; 1,24); 0,251	1,14 (1,00; 1,30); 0,059	kein Zusatznutzen (hoch)
Anteil Patienten mit DAS28-hsCRP < 2,6	1,09 (0,90; 1,33); 0,385	1,01 (0,85; 1,20); 0,904	kein Zusatznutzen (hoch)
SDAI			
Veränderung des SDAI	-2,43 (-4,28; -0,59); 0,010	-3,18 (-5,04; -1,32); < 0,001	kein Zusatznutzen (hoch)
Anteil Patienten mit SDAI ≤ 3,3	1,22 (0,86; 1,71); 0,261	1,26 (0,96; 1,67); 0,101	kein Zusatznutzen (hoch)
CDAI			
Veränderung des CDAI	-2,06 (-3,83; -0,29); 0,023	-2,85 (-4,62; -1,09); 0,002	kein Zusatznutzen (hoch)
Anteil Patienten mit CDAI ≤ 10	1,04 (0,91; 1,20); 0,567	1,14 (1,00; 1,29); 0,055	kein Zusatznutzen (hoch)
EULAR Responsekriterien („good Response“, DAS28-basiert)			
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	1,07 (0,93; 1,23); 0,323	1,15 (1,00; 1,31); 0,051	kein Zusatznutzen (hoch)

Endpunkt	Baricitinib vs. Adalimumab Effektschätzer ^a (95% KI); p-Wert		Ausmaß des Zusatznutzens ^f (Aussagesicherheit)
	Woche 24	Woche 52	
	PAP (VAS)		
Veränderung des PAP (VAS)	-4,8 (-8,0; -1,6); 0,003	-5,9 (-9,1; -2,6); < 0,001	<i>nicht quantifizierbarer Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	1,06 (0,99; 1,14); 0,111	1,13 (1,05; 1,22); 0,001	<i>geringer Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	1,09 (0,99; 1,20); 0,090	1,18 (1,07; 1,30); 0,001	<i>geringer Zusatznutzen (hoch)</i>
PtGADA (VAS)			
Veränderung des PtGADA (VAS)	-4,0 (-7,1; -0,8); 0,013	-6,0 (-9,2; -2,8); < 0,001	<i>kein Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 10 mm	1,04 (0,97; 1,11); 0,298	1,05 (0,98; 1,13); 0,167	<i>kein Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm	1,08 (0,98; 1,18); 0,124	1,10 (1,00; 1,21); 0,043	<i>kein Zusatznutzen (hoch)</i>
HAQ-DI			
Veränderung des HAQ-DI (mBOCF)	-0,11 (-0,18; -0,03); 0,008	-0,09 (-0,17; -0,00); 0,039	<i>nicht quantifizierbarer Zusatznutzen (hoch)</i>
Veränderung des HAQ-DI (mLOCF)	-	-0,11 (-0,19; -0,03); 0,007	<i>nicht quantifizierbarer Zusatznutzen (hoch)</i>
Veränderung des HAQ-DI (MMRM)	-0,12 (-0,2; -0,0); 0,002	-0,08 (-0,2; -0,0); 0,044	<i>nicht quantifizierbarer Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,22	1,12 (1,02; 1,23); 0,021	1,14 (1,03; 1,27); 0,016	<i>geringer Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,3	1,11 (1,00; 1,23); 0,060	1,10 (0,98; 1,24); 0,103	<i>kein Zusatznutzen (hoch)</i>

Endpunkt	Baricitinib vs. Adalimumab		Ausmaß des Zusatznutzens ^f (Aussagesicherheit)
	Effektschätzer ^a (95% KI); p-Wert		
	Woche 24	Woche 52	
Dauer der Morgensteifigkeit (min)			
Veränderung der Dauer der Morgensteifigkeit (min)	-20,65 (-45,26; 3,95); 0,100	-27,34 (-51,35; -3,32); 0,026 Hedges'g: -0,18 (-0,36; 0,01)	<i>geringer Zusatznutzen (hoch)</i>
EQ-5D VAS			
Veränderung der EQ-5D VAS	5,6 (2,4; 8,8); < 0,001	7,4 (4,2; 10,7); < 0,001	<i>nicht quantifizierbarer Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung der EQ-5D VAS \geq 10 mm	1,15 (1,02; 1,29); 0,020	1,16 (1,04; 1,30); 0,009	<i>geringer Zusatznutzen (hoch)</i>
Schwerste Müdigkeit			
Veränderung der schwersten Müdigkeit	-0,27 (-0,70; 0,16); 0,222	-0,59 (-1,02; -0,16); 0,007 Hedges'g: -0,22 (-0,40; -0,03)	<i>kein Zusatznutzen (hoch)</i>
Lebensqualität			
SF-36 v2 akut			
Veränderung des SF-36 v2 akut PCS	1,7 (0,7; 2,8); 0,001	1,7 (0,6; 2,9); 0,004	<i>nicht quantifizierbarer Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS \geq 5	1,15 (1,02; 1,28); 0,018	1,14 (1,00; 1,29); 0,047	<i>geringer Zusatznutzen (hoch)</i>
Veränderung des SF-36 v2 akut MCS	0,5 (0,9; 1,8); 0,503	0,9 (0,4; 0,2); 0,169	<i>kein Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS \geq 5	1,13 (0,94; 1,36); 0,202	1,14 (0,93; 1,39); 0,219	<i>kein Zusatznutzen (hoch)</i>
Veränderung der SF-36 v2 akut Domänen			
Körperliche Funktionsfähigkeit	1,7 (0,5; 3,0); 0,006	1,4 (0,1; 2,7); 0,029 Hedges'g: 0,14 (-0,00; 0,28)	<i>kein Zusatznutzen (hoch)</i>
Körperliche Rollenfunktion	1,3 (0,2; 2,5); 0,025	1,9 (0,7; 3,1); 0,002 Hedges'g: 0,20 (0,06; 0,34)	<i>kein Zusatznutzen (hoch)</i>

Endpunkt	Baricitinib vs. Adalimumab		Ausmaß des Zusatznutzens ^f
	Effektschätzer ^a (95% KI); p-Wert		(Aussagesicherheit)
	Woche 24	Woche 52	
Körperliche Schmerzen	1,8 (0,6; 3,0); 0,002	1,5 (0,2; 2,8); 0,028 Hedges'g: 0,14 (-0,00; 0,28)	kein Zusatznutzen (hoch)
Allgemeiner Gesundheitszustand	0,7 (0,4; 1,8); 0,209	1,3 (0,2; 2,5); 0,023 Hedges'g: 0,14 (0,00; 0,28)	kein Zusatznutzen (hoch)
Vitalität	0,8 (0,4; 2,1); 0,195	1,3 (0,0; 2,6); 0,045 Hedges'g: 0,13 (-0,02; 0,27)	kein Zusatznutzen (hoch)
Soziale Funktionsfähigkeit	2,0 (0,7; 3,3); 0,002	1,9 (0,7; 3,2); 0,003 Hedges'g: 0,19 (0,05; 0,33)	kein Zusatznutzen (hoch)
Emotionale Rollenfunktion	0,7 (0,7; 2,0); 0,355	1,4 (0,0; 2,7); 0,048 Hedges'g: 0,12 (-0,02; 0,27)	kein Zusatznutzen (hoch)
Mentaler Gesundheitszustand	0,5 (0,8; 1,9); 0,441	0,7 (0,7; 2,0); 0,327 Hedges'g: 0,06 (-0,08; 0,20)	kein Zusatznutzen (hoch)
Sicherheit			
Todesfälle	5,36 (0,318; 90,546); 0,518	1,34 (0,133; 13,502); 1,000	kein höheres/geringeres Schadenspotenzial (hoch)
SUE	2,60 (1,069; 6,310); 0,033	1,98 (1,072; 3,660); 0,027	höheres Schadenspotenzial: Ausmaß: gering (hoch)
UE	1,05 (0,956; 1,152); 0,302	1,03 (0,954; 1,109); 0,460	kein höheres/geringeres Schadenspotenzial (hoch)
Schwere UE (ohne ausschließliche Laborereignisse) ^c	2,44 (0,915; 6,506); 0,084	1,72 (0,871; 3,416); 0,133	kein höheres/geringeres Schadenspotenzial (hoch)

Endpunkt	Baricitinib vs. Adalimumab		Ausmaß des Zusatznutzens ^f (Aussagesicherheit)
	Effektschätzer ^a (95% KI); p-Wert		
	Woche 24	Woche 52	
Studienabbruch aufgrund von UE	2,13 (0,920; 4,928); 0,083	1,92 (1,009; 3,652); 0,045	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
Therapieabbruch aufgrund von UE	2,23 (0,967; 5,129); 0,059	1,65 (0,897; 3,018); 0,129	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
UE von besonderem Interesse			
Infektionen	1,08 (0,894; 1,315); 0,409	1,09 (0,934; 1,269); 0,272	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
Schwere Infektionen	1,02 (0,170; 6,093); 1,000	1,36 (0,411; 4,464); 0,771	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
Kardiovaskuläre Ereignisse	5,35 (0,099; 290,613); 1,000	1,34 (0,133; 13,502); 1,000	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
Malignität	5,36 (0,318; 90,546); 0,518	5,37 (0,534; 54,094); 0,277	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
Depressionen: QIDS-SR16	-0,0 (-0,5; 0,5); 0,937	0,2 (-0,4; 0,8); 0,449	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
Allergische Reaktionen/Hypersensitivität ^d	-	-	<i>kein höheres/geringeres Schadenspotenzial</i>
Gastrointestinale Perforation ^e	-	-	<i>kein höheres/geringeres Schadenspotenzial</i>

Endpunkt	Baricitinib vs. Adalimumab Effektschätzer ^a (95% KI); p-Wert		Ausmaß des Zusatznutzens ^f (Aussagesicherheit)
	Woche 24	Woche 52	
<p>a: Für kontinuierliche Endpunkte wird der LSMD dargestellt. Für binäre Endpunkte wird das RR dargestellt. Für Todesfälle und UE von besonderem Interesse wird das Peto OR dargestellt.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>c: Alle in der vorliegenden Nutzenbewertung aufgeführten schweren unerwünschten Ereignisse wurden ohne Laborereignisse ausgewertet, in deren Zusammenhang keine Symptome beschrieben wurden oder dies angenommen werden musste und die somit asymptomatisch waren (Abschnitt 4.2.5.2). Diese schweren UE waren ausschließlich Laborereignisse ohne assoziierte Symptomatik und daher für den Patienten nicht zu spüren, also nicht patientenrelevant. Sollte eine mit diesen Laborereignissen in Zusammenhang stehende Symptomatik aufgetreten sein, dann wurde diese Symptomatik innerhalb der unerwünschten Ereignisse erfasst.</p> <p>d: Für das UE von besonderem Interesse Allergische Reaktionen/Hypersensitivität trat ein Ereignis auf.</p> <p>e: Für das UE von besonderem Interesse Gastrointestinale Perforation trat kein Ereignis auf.</p> <p>f: Sofern eine Responderanalyse für einen Endpunkt einen Zusatznutzen ergeben hat, wurde das Ausmaß des Zusatznutzens für die kontinuierliche Auswertung dieses Endpunkts auf <i>nicht quantifizierbar</i> gesetzt.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59-60].</p>			

Schlussfolgerungen zum Zusatznutzen und zum therapeutisch bedeutsamen Zusatznutzen

Bei der der Bewertung zugrunde liegenden Studie handelt es sich um eine randomisierte, aktiv-kontrollierte, klinische Studie der Evidenzstufe Ib. Die Methodik der Studie entspricht den internationalen Standards der *Good Clinical Practice* (GCP) [61]. Es handelt sich somit um eine hochwertige Studie, die zudem von ausreichender Größe und daher grundsätzlich geeignet ist, *Hinweise* für einen Zusatznutzen zu liefern.

Subpopulation A1

Für Patienten der Subpopulation A1, für Patienten, bei denen **keine ungünstigen Prognosefaktoren**¹ vorliegen und die unzureichend auf eine vorangegangene Behandlung mit **einem** krankheitsmodifizierenden Antirheumatikum (klassische DMARDs, inklusive MTX) ansprachen, hat der G-BA die zVT alternative klassische DMARDs, sofern geeignet (z. B. MTX, Leflunomid) als Mono- oder Kombinationstherapie, bestimmt. Derzeit liegt keine direkte Evidenz zum Vergleich von Baricitinib mit der vom G-BA genannten zVT für diese Subpopulation vor und somit kann **kein Zusatznutzen** gegenüber einer der zweckmäßigen Vergleichstherapien festgestellt werden.

Subpopulation A2 und Subpopulation A3

Wie oben ausgeführt, werden für die Ableitung des Zusatznutzens für die Subpopulationen A2 und A3 die Ergebnisse der mITT Population herangezogen.

Für die mITT Population wird basierend auf der direkt vergleichenden Studie RA-BEAM (JADV) (Baricitinib vs. Adalimumab vs. Placebo, jeweils in Kombination mit MTX) ein *Hinweis* auf einen **geringen Zusatznutzen** gegenüber der zVT Adalimumab festgestellt (Tabelle 4-2).

Ein Zusatznutzen zeigt sich durch die im Vergleich zur zVT statistisch signifikant besseren Ergebnisse der **Morbiditätsendpunkte Anzahl druckschmerzhafter und geschwollener Gelenke, Schmerz** (Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm und mit einer Verbesserung des PAP (VAS) ≥ 20 mm), **körperliche Funktion** (Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$), **Dauer der Morgensteifigkeit** und **allgemeiner Gesundheitszustand** (Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm) (Tabelle 4-1). Der Zusatznutzen zeigt sich zudem durch die im Vergleich zur zVT statistisch signifikant besseren Ergebnisse des Endpunktes der **Lebensqualität** (Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5) (Tabelle 4-1).

¹ Beispiele für ungünstige Prognosefaktoren:

- Nachweis von Autoantikörpern (z. B. Rheumafaktoren, hohe Spiegel von Antikörpern gegen citrullinierte Peptid Antigene)
- Hohe Krankheitsaktivität (nachgewiesen durch DAS- bzw. DAS28-Bewertungssystem, geschwollene Gelenke, Parameter der Akute-Phase-Reaktion wie z. B. C-Reaktives Protein, Erythrozytensedimentationsrate)
- Frühes Auftreten von Gelenkerosionen.

Quelle: [2].

Zusätzlich zeigen sich statistisch signifikante Vorteile für Baricitinib im Vergleich zu Adalimumab für weitere Endpunkte (Tabelle 4-1). Auch wenn diese Endpunkte nach den formalen Kriterien des IQWiG nicht zu der Ableitung eines Zusatznutzens auf Endpunktebene führen, so legen sie doch in ihrer Gesamtsumme einen geringen Zusatznutzen für Baricitinib nahe.

Bezogen auf die **Sicherheitsendpunkte** insgesamt ergab sich für Baricitinib ein höheres Schadenspotenzial durch den Endpunkt **schwerwiegende unerwünschte Ereignisse** (Tabelle 4-1). Hierzu ist zu bemerken, dass die SUE unter Baricitinib an Woche 24 auf Placeboniveau waren. Weiterhin waren die SUE der Adalimumab-Behandlungsgruppe, vor dem Hintergrund der bekannten Datenlage [58], auf unerwartet niedrigem Niveau. Die aufgetretenen SUE waren gleichmäßig über die SOCs verteilt. In keiner SOC konnte ein statistisch signifikanter Unterschied zwischen der Baricitinib- und der Adalimumab-Behandlungsgruppe beobachtet werden.

In der Gesamtschau überwiegen die positiven Effekte der patientenrelevanten Endpunkte der Nutzendimensionen *Morbidität* und *Lebensqualität* den Nachteil der Nutzendimensionen *Sicherheit*. Daher wird der Zusatznutzen von Baricitinib als eine bisher nicht erreichte moderate und nicht nur geringfügige Verbesserung des therapielevanten Nutzens bewertet.

Tabelle 4-2: Gesamtaussage zum medizinischen Zusatznutzen für Subpopulation A2 und A3 (basierend auf den Ergebnissen der mITT Population)

Studie RA-BEAM (JADV)	
Nutzendimension Aussage zum Zusatznutzen	Gesamtaussage zum Zusatznutzen
Mortalität^a	Hinweis auf geringen Zusatznutzen
-	
Morbidität	
Hinweis auf einen geringen Zusatznutzen	
Lebensqualität	
Hinweis auf einen geringen Zusatznutzen	
Sicherheit	
Hinweis auf ein höheres Schadenspotenzial: Ausmaß: gering	
<p>a: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p>	

Fazit

Die Rheumatoide Arthritis ist eine systemische entzündliche Autoimmunerkrankung. Die Krankheit hat unterschiedliche Verlaufsformen und Schweregrade, die sich von einer milden Erkrankung bis hin zu einer schwerwiegenden Erkrankung, assoziiert mit zunehmender Gelenkzerstörung, bedeutend eingeschränkter Lebensqualität und reduziertem Überleben, erstrecken können [62-63]. Die stärkste Belastung für Patienten stellen Schmerzen, Fatigue, Morgensteifigkeit sowie Funktionseinschränkungen und Depressionen dar. Die negativen psychologischen Effekte der RA scheinen der Kombination aus Schmerzen und Funktionseinschränkungen zu folgen. Chronische Schmerzen sind eine der Hauptkomponenten der Krankheitslast für RA-Patienten. Die Patienten leiden an stärkeren Schmerzen als die Allgemeinbevölkerung und erfahren ein Schmerzniveau, das mit dem von Schmerzpatienten vergleichbar ist. Höhere Schmerzniveaus korrelieren mit der Funktionseinschränkung sowie mit Depressionen, was sich negativ auf die Lebensqualität der Patienten auswirkt.

In den vergangenen Jahren sind weitere Optionen zur Therapie der RA verfügbar geworden. Trotz einer Vielzahl zugelassener Medikamente für die Behandlung der RA ist die komplette oder anhaltende Krankheitsremission noch immer die Ausnahme. Ein Großteil der Patienten in klinischen Studien erreicht keine klinische Remission [64]. Selbst in Patienten, die eine scheinbar ausreichende Kontrolle der Anzeichen und Symptome der RA durch eine Therapie mit cDMARDs und/oder Biologika erreichen, kann die Erkrankung voranschreiten [65-66]. Bei derzeit auf dem Markt befindlichen Therapeutika zur Behandlung der RA muss zudem mit einem verzögerten Wirkeintritt gerechnet werden.

Auch wenn eine Kontrolle der Schmerzen ein wichtiges Kriterium einer erfolgreichen Therapie darstellt, erfährt ein Großteil der Patienten trotz einer Therapie und der aktuellen Therapieoptionen noch immer bedeutsame Schmerzen [36-40] und eine damit einhergehende Einschränkungen der Lebensqualität [40,67]. Dementsprechend besteht weiterhin ein erheblicher, ungedeckter Bedarf für alternative Therapieoptionen zur Behandlung der RA.

Baricitinib ist ein oral verfügbarer, selektiver Januskinase 1 (JAK1)-/Januskinase 2 (JAK2)-Inhibitor zur Behandlung der mittelschweren bis schweren aktiven rheumatoiden Arthritis, der gezielt in Signalkaskaden eingreift und der nicht auf Signalwege, die für ein funktionierendes Immunsystem wichtig sind, wirkt.

Durch die einmal tägliche orale Gabe, ohne die Notwendigkeit subkutaner oder intravenöser Injektionen/Infusionen, ist Baricitinib zudem patientenfreundlich und kann zu einer verbesserten *Compliance* führen. Baricitinib zeichnet sich durch einen vergleichsweise schnelleren Wirkeintritt aus und führt zu einer schnelleren Verbesserung patientenrelevanter Endpunkte (wie z. B. Schmerz, Anzahl schmerzhafter Gelenke,...). Das Risiko, dass die Erkrankung im kritischen Zeitfenster bedeutend voranschreitet, verringert sich. Die Gefahr des sekundären Wirkverlustes besteht nicht, da keine Antikörper gegenüber Baricitinib gebildet werden.

Damit kann durch Baricitinib als Alternative zu Adalimumab eine „*Verbesserung des Gesundheitszustands und der Lebensqualität*“ für die Patienten erreicht werden [3].

Aufgrund der Strukturgleichheit, v. a. bezüglich der klinischen Parameter, zwischen der Studienpopulation und der Zielpopulation im Anwendungsgebiet A wird davon ausgegangen, dass die in der Studie RA-BEAM (JADV) beobachteten klinischen Effekte auch in der medizinischen Versorgung unter Alltagsbedingungen auftreten und die Studienergebnisse somit auf den deutschen Versorgungskontext übertragbar sind (Abschnitt 4.3.1.2.1).

4.2 Methodik

Abschnitt 4.1 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

Unter Endpunkte sind dabei alle für die frühe Nutzenbewertung relevanten Endpunkte anzugeben (d. h. nicht nur solche, die ggf. in den relevanten Studien untersucht wurden).

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.

Fragestellung

Gemäß Fachinformation ist:

„Olumiant angezeigt zur Behandlung von mittelschwerer bis schwerer aktiver rheumatoider Arthritis bei erwachsenen Patienten, die auf eine vorangegangene Behandlung mit einem oder mehreren krankheitsmodifizierenden Antirheumatika (DMARDs) unzureichend angesprochen oder diese nicht vertragen haben. Olumiant kann als Monotherapie oder in Kombination mit Methotrexat eingesetzt werden [1].“

Es sollen der medizinische Nutzen und der medizinische Zusatznutzen gemäß § 35a SGB V von Baricitinib zur Behandlung erwachsenen Patienten mit mittelschwerer bis schwerer aktiver RA, die auf eine vorangegangene Behandlung mit einem oder mehreren cDMARDs unzureichend angesprochen oder diese nicht vertragen haben, vergleichend zu einer vom G-BA genannten zVT bewertet werden [2]. Baricitinib kann als Monotherapie oder in Kombination mit MTX eingesetzt werden.

Ursprünglich wurde eine Anfrage an den G-BA gestellt, die zVT für Baricitinib für den Fall zu bestimmen, dass eine Zulassung von Baricitinib neben der Zweit- und Drittlinien-Behandlung auch für die Erstlinien-Behandlung erfolgt. Eine Behandlung Therapie-naiver Patienten ohne vorherige Behandlung ist jedoch außerhalb der Zulassung von Baricitinib und somit für dieses Nutzendossier nicht von Relevanz.

Gemäß G-BA-Beratungsgespräch vom 10.12.2015 [2] sind im vorliegenden Anwendungsgebiet verschiedene Therapiesituationen zu unterscheiden, die in:

- Therapie-naive Patienten ohne vorherige Behandlung oder
- Patienten mit vorangegangenen Behandlungen mit einem oder mehreren **cDMARDs**, einschließlich MTX, oder (Anwendungsgebiet A)
- Patienten mit vorangegangenen Behandlungen mit einem oder mehreren **bDMARDs** (Anwendungsgebiet B)

zu differenzieren sind.

Die Zielpopulation wurde, aufgrund der verschiedenen durch den G-BA festgelegten Therapiesituationen, in zwei Anwendungsgebiete aufgeteilt, die nachfolgend als Anwendungsgebiet A (Modul 4 A) und Anwendungsgebiet B (Modul 4 B) bezeichnet werden (Modul 2 und Modul 3 (Abschnitt 3.1)).

Die Bewertung erfolgt in den Kategorien *Mortalität, Morbidität und Lebensqualität* sowie *Sicherheit* (Abschnitt 4.3.1.3.1 bis 4.3.1.3.4) und auf Grundlage derjenigen klinischen Endpunkte, mit denen diese Kategorien in der zur Bewertung herangezogenen Studie operationalisiert wurden.

Patientenpopulation

Im vorliegenden Dossier werden der medizinische Nutzen und der medizinischen Zusatznutzen von Baricitinib als Mono- oder Kombinationstherapie zur Behandlung erwachsenen Patienten mit mittelschwerer bis schwerer aktiver RA, die auf eine vorangegangene Behandlung mit einem oder mehreren cDMARDs unzureichend angesprochen oder diese nicht vertragen haben, bewertet.

Die Zielpopulation im Anwendungsgebiet A wurde vom G-BA [2] in folgende Subpopulationen unterteilt:

Subpopulation A1

Patienten, bei denen **keine ungünstigen Prognosefaktoren**¹ vorliegen und die unzureichend auf eine vorangegangene Behandlung mit **einem** cDMARDs (inklusive MTX) ansprachen.

Subpopulation A2

Patienten **mit ungünstigen Prognosefaktoren**¹, die unzureichend auf eine vorangegangene Behandlung mit **einem** cDMARDs (inklusive MTX) ansprachen.

Subpopulation A3

Patienten, die unzureichend auf eine vorangegangene Behandlung mit **mehreren** cDMARDs (inklusive MTX) ansprachen.

Intervention

Die zu untersuchende Intervention ist Baricitinib. Baricitinib ist gemäß Fachinformation indiziert zur Behandlung von mittelschwerer bis schwerer aktiver RA bei erwachsenen Patienten, die auf eine vorangegangene Behandlung mit **einem** oder **mehreren** DMARDs unzureichend angesprochen oder diese nicht vertragen haben. Baricitinib kann als Monotherapie oder in Kombination mit MTX eingesetzt werden [1].

Vergleichstherapie

Gemäß G-BA-Beratungsgespräch am 10.12.2015 [2] deckt das Anwendungsgebiet von Baricitinib (Olumiant®) sowohl die Zweit- als auch Drittlinien-Behandlung erwachsener Patienten mit mittelschwerer bis schwerer aktiver RA ab. Die zVT wurde daher für Patientenpopulationen mit Vorbehandlung unter Berücksichtigung der Eignung für eine kombinierte MTX-Gabe als auch für eine Monotherapie bestimmt.

Für Patienten der Subpopulation A1 hat der G-BA die zVT alternative cDMARDs, sofern geeignet (z. B. MTX, Leflunomid) als Mono- oder Kombinationstherapie, bestimmt. Derzeit liegt keine direkte Evidenz zum Vergleich von Baricitinib mit der vom G-BA genannten zVT für diese Subpopulation vor.

Für die Subpopulationen A2 und A3 hat der G-BA die zVT bDMARDs in Kombination mit MTX (Adalimumab oder Etanercept oder Certolizumab-Pegol oder Golimumab oder Abatacept oder Tocilizumab); ggf. als Monotherapie unter Berücksichtigung des jeweiligen Zulassungsstatus bei MTX-Unverträglichkeit benannt.

¹ Beispiele für ungünstige Prognosefaktoren:

- Nachweis von Autoantikörpern (z. B. Rheumafaktoren, hohe Spiegel von Antikörpern gegen citrullinierte Peptid Antigene)
- Hohe Krankheitsaktivität (nachgewiesen durch DAS- bzw. DAS28-Bewertungssystem, geschwollene Gelenke, Parameter der Akute-Phase-Reaktion wie z. B. C-Reaktives Protein, Erythrozytensedimentationsrate)
- Frühes Auftreten von Gelenkerosionen.

Quelle: [2].

Für die Subpopulationen A2 und A3 werden Ergebnisse von Baricitinib im Vergleich zu Adalimumab aus der randomisierten, aktiv-kontrollierten Phase-III-Studie RA-BEAM (JADV) von Baseline bis Woche 24 und Woche 52 dargestellt. Die *Lilly Deutschland GmbH* folgt somit der vom G-BA genannten zVT.

Endpunkte

Die Bewertung erfolgte in den Kategorien *Mortalität*, *Morbidität* und *Lebensqualität* sowie *Sicherheit* und auf Grundlage derjenigen klinischen Endpunkte, mit denen diese Kategorien in der zur Bewertung herangezogenen Studie operationalisiert wurden.

Studientypen

Die Evidenzgrundlage für die Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens von Baricitinib im Anwendungsbiet ist eine RCT. Sofern zur Darstellung des Zusatznutzens keine direkt vergleichende Studie gegen die zVT zur Verfügung steht, wird ein indirekter Vergleich herangezogen.

4.2.2 Kriterien für den Einschluss von Studien in die Nutzenbewertung

Die Untersuchung der in Abschnitt 4.2.1 benannten Fragestellung soll auf Basis von klinischen Studien vorgenommen werden. Für die systematische Auswahl von Studien für diese Untersuchung sollen Ein- und Ausschlusskriterien für die Studien definiert werden. Dabei ist zu beachten, dass eine Studie nicht allein deshalb ausgeschlossen werden soll, weil keine in einer Fachzeitschrift veröffentlichte Vollpublikation vorliegt. Eine Bewertung der Studie kann beispielsweise auch auf Basis eines ausführlichen Ergebnisberichts aus einem Studienregister erfolgen, während ein Kongressabstract allein in der Regel nicht für eine Studienbewertung ausreicht.

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

Ein-/Ausschlusskriterien für Studien zur Bewertung von direkter Evidenz

Für die Auswahl in die Bewertung einzubeziehender Studien wurden Ein- und Ausschlusskriterien bezüglich der Studienpopulation, der Intervention, der Vergleichstherapie, der Endpunkte, des Studientyps, der Studiendauer und des Publikationstyps formuliert (Tabelle 4-3). Studien wurden nach folgenden Kriterien für die Bewertung des medizinischen Nutzens und Zusatznutzens ein- bzw. ausgeschlossen.

Tabelle 4-3: Einschlusskriterien für die Studienbewertung

Kriterium	Einschlusskriterien	Ausschlusskriterien	Begründung
Population	<p>Erwachsene Patienten (≥ 18 Jahre) mit mittelschwerer bis schwerer aktiver rheumatoider Arthritis,</p> <p><u>Subpopulation A1</u></p> <p>bei denen keine ungünstigen Prognosefaktoren^b vorliegen und die unzureichend auf eine vorangegangene Behandlung mit einem cDMARD (inklusive MTX) ansprachen.</p> <p><u>Subpopulation A2</u></p> <p>mit ungünstigen Prognosefaktoren^b, die unzureichend auf eine vorangegangene Behandlung mit einem cDMARD (inklusive MTX) ansprachen.</p>	<p>Patienten, die nicht der Definition gemäß Einschlusskriterien entsprechen.</p>	<p>gem. Fachinformation [1]</p>

Kriterium	Einschlusskriterien	Ausschlusskriterien	Begründung
	<u>Subpopulation A3</u> die unzureichend auf eine vorangegangene Behandlung mit mehreren cDMARDs (inklusive MTX) ansprachen.		
Intervention	Baricitinib, das als Monotherapie oder in Kombination mit MTX verabreicht werden kann.	Andere Interventionen	gem. Fachinformation [1]
Vergleichstherapie	<u>Subpopulation A1</u> alternative cDMARDs, sofern geeignet (z. B. MTX, Leflunomid) als Mono- oder Kombinationstherapie. <u>Subpopulationen A2/ Subpopulationen A3</u> bDMARDs in Kombination mit MTX (Adalimumab oder Etanercept oder Certolizumab-Pegol oder Golimumab oder Abatacept oder Tocilizumab); ggf. als Monotherapie unter Berücksichtigung des jeweiligen Zulassungsstatus bei MTX-Unverträglichkeit.	Andere Vergleichstherapie	Erfüllung der Anforderungen für eine zweckmäßige Vergleichstherapie nach § 6 VerfO [68], gem. Festlegung der zVT durch den G-BA [2], (Modul 3)
Endpunkte^a	Endpunkte, die zur Bewertung der <i>Mortalität</i> , <i>Morbidität</i> , <i>Lebensqualität</i> und <i>Sicherheit</i> herangezogen werden können.	Keine Endpunkte, die zur Bewertung der <i>Mortalität</i> , <i>Morbidität</i> , <i>Lebensqualität</i> oder der <i>Sicherheit</i> herangezogen werden können.	Auswahl der Nutzendimensionen nach § 5 Abs. 2 VerfO [68], <i>Sicherheit</i> zur Schadensabwägung [6]
Studientyp	Randomisierte kontrollierte klinische Studien	Studien, die nicht randomisiert, nicht kontrolliert oder nicht klinisch sind.	Erfüllung der Anforderungen an den Nachweis des Zusatznutzens gem. § 5 Abs. 3 VerfO [68]
Studiendauer	mindestens 24 Wochen	weniger als 24 Wochen	Erfüllung der Anforderungen an die Studiendauer zur Untersuchung chronischer Erkrankungen [69]

Kriterium	Einschlusskriterien	Ausschlusskriterien	Begründung
Publikationstyp	Berichterstattung liefert ausreichende Informationen zur Beurteilung von Methodik/Ergebnissen (z. B. Vollpublikation, Ergebnisbericht aus einem Studienregister oder Studienbericht verfügbar)	Berichterstattung liefert keine ausreichenden Informationen zur Beurteilung von Methodik/Ergebnissen (z. B. narrativer Review, Letter, Editorial, Errata, Note, Konferenz-Abstract oder Paper, Short Survey, unsystematische Übersichtsarbeit)	Ausreichende Datenverfügbarkeit zur Bewertung von Ergebnissen
<p>a: Auf eine Aufzählung aller möglichen oder wahrscheinlich verwendeten Messinstrumente und Operationalisierungen von Endpunkten wird an dieser Stelle verzichtet. Im Gegensatz zu einer prospektiv geplanten Datenerhebung ist zum Zeitpunkt der Planung eines Nutzendossiers nicht bekannt, welche Messinstrumente in den zu einem späteren Zeitpunkt aufzufindenden Studien verwendet werden. Von Relevanz ist vielmehr die Frage, wie die Nutzendimensionen <i>Mortalität</i>, <i>Morbidität</i>, <i>Lebensqualität</i> und <i>Sicherheit</i> in den der Nutzenbewertung zugrunde gelegten RCT operationalisiert werden (Abschnitt 4.3.1.3.1 bis 4.3.1.3.4).</p> <p>b: Beispiele für ungünstige Prognosefaktoren:</p> <ul style="list-style-type: none"> • Nachweis von Autoantikörpern (z. B. Rheumafaktoren, hohe Spiegel von Antikörpern gegen citrullinierte Peptid Antigene) • Hohe Krankheitsaktivität (nachgewiesen durch DAS- bzw. DAS28-Bewertungssystem, geschwollene Gelenke, Parameter der Akute-Phase-Reaktion wie z. B. C-Reaktives Protein, Erythrozytensedimentationsrate) • Frühes Auftreten von Gelenkerosionen. <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [1-2,68-69].</p>			

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 erforderlich, um sicherzustellen, dass ein vollständiger Studienpool in die Bewertung einfließt.

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

Das Datum der Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Die bibliografische Literaturrecherche soll mindestens in den Datenbanken MEDLINE und EMBASE sowie in der Cochrane-Datenbank „Cochrane Central Register of Controlled Trials (Clinical Trials)“ 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. Alle Suchstrategien sind Anhang 4-A zu dokumentieren.

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

Zur Identifizierung von relevanten Primärpublikationen für Studien mit dem zu bewertenden Arzneimittel wurde eine systematische bibliographische Literaturrecherche durchgeführt (Suchzeitpunkt: 02.02.2017). Diese erfolgte für die in Abschnitt 4.2.1 formulierte Fragestellung. Für die Recherche wurde auf der Suchoberfläche Ovid jeweils einzeln in den Datenbanken Cochrane, MEDLINE und EMBASE nach relevanten Treffern gesucht. Die Suche wurde auf englisch- und deutschsprachige Literatur beschränkt. Für die Einschränkung der Suche hinsichtlich des Studientyps auf randomisierte kontrollierte Studien wurde der validierte und in der Verfo empfohlene Filter nach Wong verwendet [70].

Die Suchstrategien sind in Anhang 4-A dokumentiert.

4.2.3.3 Suche in Studienregistern

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

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

Das Datum der Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Die Suche soll mindestens in den Studienregistern [clinicaltrials.gov](http://www.clinicaltrials.gov) (www.clinicaltrials.gov), EU Clinical Trials Register (EU-CTR, www.clinicaltrialsregister.eu), Klinische Prüfungen PharmNet.Bund (<http://www.pharmnet-bund.de/dynamic/de/klinische-pruefungen/index.htm>) sowie über das International Clinical Trials Registry Platform Search Portal (ICTRP Search Portal, Suchportal der WHO: <http://apps.who.int/trialsearch/>) durchgeführt werden. Optional kann zusätzlich eine Suche in weiteren themenspezifischen Studienregistern (z. B. krankheitsspezifische Studienregister oder Studienregister einzelner pharmazeutischer Unternehmen) durchgeführt werden. Die Suche in Studienregistern anderer pharmazeutischer Unternehmen ist insbesondere bei indirekten Vergleichen sinnvoll, wenn Studien zu anderen Arzneimitteln identifiziert werden müssen.

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. Alle Suchstrategien sind in Anhang 4-B zu dokumentieren.

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

Für die Identifizierung relevanter Studien mit dem zu bewertenden Arzneimittel wurden die Studienregister *ClinicalTrials.gov* (www.clinicaltrials.gov), *EU Clinical Trials Register* (www.clinicaltrialsregister.eu), die Datenbank zu in Deutschland genehmigten klinischen Prüfungen mit Arzneimitteln PharmNet.Bund (www.pharmnet-bund.de) und das *International Clinical Trials Registry Platform* Suchportal der WHO (*ICTRP Search Portal*; www.who.int/ictrp) nach abgeschlossenen, abgebrochenen und noch laufenden relevanten Studien durchsucht.

Die Suche wurde in jedem Studienregister einzeln und mit einer an das jeweilige Register adaptierten Suchstrategie durchgeführt.

Die detaillierte Suchstrategie und deren Ergebnisse sind in Anhang 4-B dokumentiert.

4.2.3.4 Selektion relevanter Studien

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

Bibliographische Literaturrecherche

Die Liste der im Rahmen der bibliographischen Literaturrecherche identifizierten Publikationen wurde unabhängig von zwei Personen unter Berücksichtigung der Einschlusskriterien aus Tabelle 4-3 auf ihre Relevanz hin überprüft. In einem ersten Schritt wurden Publikationen ausgeschlossen, deren Titel eindeutig eine Einstufung als nicht relevant zuließ. In nachfolgenden Schritten wurde zunächst das Abstract und bei Unklarheiten der Volltext gesichtet, um festzustellen, ob alle Einschlusskriterien erfüllt waren. Voneinander abweichende Einstufungen der beiden unabhängigen Bewerter wurden diskutiert und ggf. unter Einbeziehung eines dritten Bewerter behoben. Bei Erfüllung aller Einschlusskriterien wurde die betreffende Studie in den Studienpool der zu bewertenden Studien (Abschnitt 4.3.1.1.4) eingeschlossen.

Suche in Studienregistern

Die aus der Registersuche zusammengestellten Feldeinträge wurden unter Berücksichtigung der Einschlusskriterien aus Abschnitt 4.2.2 von zwei Personen unabhängig voneinander bewertet und auf ihre Relevanz hin überprüft. Voneinander abweichende Einstufungen der beiden unabhängigen Bewerter wurden diskutiert und ggf. unter Einbeziehung eines dritten Bewerter behoben. Bei Erfüllung aller Einschlusskriterien wurde die betreffende Studie in den Studienpool der zu bewertenden Studien (Abschnitt 4.3.1.1.4) eingeschlossen.

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-A):

A: Verzerrungsaspekte der Ergebnisse auf Studienebene

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

B: Verzerrungsaspekte der Ergebnisse auf Endpunktebene

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

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

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

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

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

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

Die Bewertung der eingeschlossenen Studien erfolgte auf Studien- und Endpunktebene.

Verzerrungspotenzial auf Studienebene

In diese Beurteilung flossen Aspekte des Studiendesigns ein, darunter die Erzeugung der Randomisierungssequenz, die Verdeckung der Gruppenzuteilung und die Verblindung von Patienten und Behandlern. Zudem wurde untersucht, ob Verzerrungen bei der Berichterstattung der Ergebnisse vorlagen. Hierzu wurden die Angaben im Studienprotokoll und im Studienbericht verglichen. Um weitere Verzerrungen zu erkennen, wurden die Transparenz und Plausibilität des Patientenflusses sowie die Vergleichbarkeit der Gruppen (Alter, Geschlecht, Konstitution, Begleiterkrankungen und -behandlungen) bewertet. Die Transparenz der Beschreibung und Durchführung der Methodik (auch von Interimsanalysen) und ggf. nachträgliche Änderungen hierzu wurden beurteilt. Ausgehend von dieser Bewertung wurde das Verzerrungspotenzial auf Studienebene als *niedrig* oder *hoch* eingestuft. Eine Einstufung als *niedrig* wurde vorgenommen, wenn eine relevante Verzerrung der Ergebnisse unwahrscheinlich und keine Anhaltspunkte für verzerrende Aspekte vorhanden waren, die bei Behebung die Grundaussage der Ergebnisse verändert hätten.

Verzerrungspotenzial auf Endpunktebene

Auf Endpunktebene wurden Aspekte u. a. zur Verblindung der Endpunkterheber, Analyse der Daten, zur Berichterstattung und weitere potenziell verzerrende Punkte berücksichtigt. Das Verzerrungspotenzial wurde als *niedrig* eingestuft, wenn sich keine Anhaltspunkte für Verzerrungen finden, die die Aussage der Ergebnisse bei Behebung grundlegend verändert hätten.

4.2.5 Informationssynthese und -analyse

4.2.5.1 Beschreibung des Designs und der Methodik der eingeschlossenen Studien

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

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

Die Informationen zu den eingeschlossenen Studien wurden anhand der Items 2b bis 14 des CONSORT-Statements 2010 beschrieben [71-72]. Der Patientenfluss wurde gemäß CONSORT-Flow-Chart in Anhang 4-E dargestellt. Dazu wurden Angaben zu Studienziel, Methodik (Design, Probanden, Interventionen, Zielkriterien, Fallzahl, Erzeugung und Geheimhaltung der Randomisierungssequenz, Durchführung der Randomisierung, Verblindung, Statistische Methoden) und Resultaten (Patientenfluss, Anzahl Studienteilnehmer, Aufnahme/Rekrutierung) gemacht. Informationen zu den Items wurden den Studienunterlagen (Studienprotokoll, statistischer Analyseplan und Studienbericht) entnommen.

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.

4.2.5.2 Gegenüberstellung der Ergebnisse der Einzelstudien

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

Benennen Sie die Patientencharakteristika und patientenrelevanten Endpunkte, die in den relevanten Studien erhoben wurden. Begründen Sie, wenn Sie von den oben benannten Vorgaben abgewichen sind. Beschreiben Sie für jeden Endpunkt, warum Sie ihn als patientenrelevant einstufen, und machen Sie Angaben zur Validität des Endpunkts (z. B. zur Validierung der eingesetzten Fragebögen). Geben Sie für den jeweiligen Endpunkt an, ob unterschiedliche Operationalisierungen innerhalb der Studien und zwischen den Studien verwendet wurden. Benennen Sie die für die Bewertung herangezogene(n) Operationalisierung(en) und begründen Sie die Auswahl. Beachten Sie bei der Berücksichtigung von Surrogatendpunkten Abschnitt 4.5.4.

Patientencharakteristika

Die Charakteristika der Patienten sowie eine graphische Darstellung zu den Studienpopulationen finden sich im Anhang 4-E. Die Baseline-Patientencharakteristika wurden erfasst und sind in Abschnitt 4.3.1.2 (Tabelle 4-14 bis Tabelle 4-16) dargestellt.

Patientenrelevante Endpunkte

Nach § 35a SGB V wird der patientenrelevante Effekt in § 3 Abs. 1 als Verbesserung des Gesundheitszustands, Verkürzung der Krankheitsdauer, Verlängerung des Überlebens, Verringerung von Nebenwirkungen oder Verbesserung der Lebensqualität definiert [73]. Da die „Verlängerung des Überlebens“ bzw. die *Mortalität* nicht als primärer oder sekundärer Endpunkt vorgesehen war und nicht im Rahmen einer Überlebenszeitanalyse untersucht wurde, wurde die Anzahl der Todesfälle im Rahmen der schwerwiegenden unerwünschten Ereignisse als Annäherung zur Beurteilung dieses patientenrelevanten Endpunktes ausgewertet. Die „Verbesserung des Gesundheitszustands“, die „Verkürzung der Krankheitsdauer“, die „Verringerung von Nebenwirkungen“ und „Verbesserung der Lebensqualität“ wurden durch die im Folgenden beschriebenen patientenrelevanten Endpunkte berücksichtigt (Abschnitt 4.2).

Validität/Operationalisierung/Patientenrelevanz der Endpunkte

Zur Bewertung des Ausmaßes des medizinischen Zusatznutzens wurden gemäß § 3 Absatz 1 und 2, 5. Kapitel der Verfo des G-BA Ergebnisse bezüglich der patientenrelevanten Nutzendimensionen *Mortalität*, *Morbidität*, *Lebensqualität* und *Sicherheit* herangezogen [68]. Insgesamt wurden dazu die im Folgenden beschriebenen patientenrelevanten Endpunkte im Dossier dargestellt. Eine Übersicht der patientenrelevanten Endpunkte befindet sich in Tabelle 4-19.

Mortalität

Da *Mortalität* nicht als primärer oder sekundärer Endpunkt vorgesehen war und nicht im Rahmen einer Überlebenszeitanalyse untersucht wurde, wurde die Anzahl der Todesfälle im Rahmen der schwerwiegenden unerwünschten Ereignisse als Annäherung zur Beurteilung dieses patientenrelevanten Endpunktes ausgewertet (Abschnitt 4.3.1.3.4).

Morbidität

- Druckschmerzhaftes Gelenke und geschwollene Gelenke (basierend auf 28 Gelenken)
 - Veränderung der Anzahl der druckschmerzhaften Gelenke
 - Veränderung der Anzahl der geschwollenen Gelenke
- *Disease Activity Score 28 High Sensitivity C-Reactive Protein* (DAS28-hsCRP)
 - Veränderung des DAS28-hsCRP
 - Anteil Patienten mit $\text{DAS28-hsCRP} \leq 3,2$
 - Anteil Patienten mit $\text{DAS28-hsCRP} < 2,6$
- *Simplified Disease Activity Index* (SDAI)
 - Veränderung des SDAI
 - Anteil Patienten mit $\text{SDAI} \leq 3,3$
- *Clinical Disease Activity Index* (CDAI)
 - Veränderung des CDAI
 - Anteil Patienten mit $\text{CDAI} \leq 10$
- EULAR Responsekriterien („*good Response*“, DAS28-basiert)
 - Anteil Patienten mit „*good Response*“ gemäß EULAR Responsekriterien (DAS28-basiert)
- *Patient Assessment of Pain* (PAP) (VAS)
 - Veränderung des PAP (VAS)
 - Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm
 - Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm

- *Patient's Global Assessment of Disease Activity (PtGADA) (VAS)*
 - Veränderung des PtGADA (VAS)
 - Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm
 - Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm
- *Health Assessment Questionnaire-Disability Index (HAQ-DI)*
 - Veränderung des HAQ-DI
 - Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$
 - Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$
- Dauer der Morgensteifigkeit (min)
 - Veränderung der Dauer der Morgensteifigkeit (min)
- *European Quality of Life-5 Dimensions (EQ-5D) VAS*
 - Veränderung des EQ-5D VAS
 - Anteil Patienten mit einer Verbesserung des EQ-5D VAS ≥ 10 mm
- Schwerste Müdigkeit (*Worst Tiredness*)
 - Veränderung der schwersten Müdigkeit

Lebensqualität

- *Short Form (36 item) Health Survey Version 2 akut (SF-36 v2 akut)*
 - Veränderung des SF-36 v2 akut PCS
 - Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5
 - Veränderung des SF-36 v2 akut MCS
 - Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5
 - Veränderung der SF-36 v2 akut Domänen

Sicherheit

- Todesfälle
- Schwerwiegende Unerwünschte Ereignisse (SUE)
- Unerwünschte Ereignisse (UE)
- Schwere UE (ohne ausschließliche Laborereignisse)
- Studienabbruch aufgrund von UE
- Therapieabbruch aufgrund von UE
- UE von besonderem Interesse
 - Infektionen
 - Schwere Infektionen
 - Kardiovaskuläre UE
 - Malignität
 - Depressionen: *16-Item Quick Inventory of Depressive Symptomatology-Self-Report* (QIDS-SR16)
 - Allergische Reaktionen/Hypersensitivität
 - Gastrointestinale Perforation

Die Validierung, Operationalisierung sowie Patientenrelevanz der Endpunkte wird im Folgenden erläutert.

Um das primäre Ziel der Behandlung der RA, die Maximierung der langfristigen gesundheitsbezogenen Lebensqualität durch eine Kontrolle der Symptome, der Prävention von strukturellen Schäden, eine Normalisierung der körperlichen Funktion sowie der Teilhabe am sozialen Leben und an der Arbeit, zu erreichen, wird bei der Behandlung der RA die Remission (insbesondere in der frühen RA) oder das Erreichen einer niedrigen Krankheitsaktivität (hauptsächlich bei längerfristiger Erkrankung) angestrebt [12-13]. Die klinische Remission ist definiert als die Abwesenheit von Anzeichen und Symptomen signifikanter entzündlicher Krankheitsaktivität [14].

ACR und EULAR haben kürzlich neue Remissionskriterien entwickelt, die auf einem Booleschen Ansatz oder einem Indexansatz beruhen und die Kriterien des SDAI oder des CDAI verwenden (Tabelle 4-4).

Tabelle 4-4: Einteilung der Krankheitsaktivität der Rheumatoiden Arthritis

Komponenten		Trennpunkte			
		Krankheitsaktivität			
		Remission	gering	mittelschwer	hoch
DAS28-ESR	<ul style="list-style-type: none"> • Anzahl geschwollener und druckschmerzhafter Gelenke (von 28) • ESR (mm) • Gesundheitszustand 	< 2,6	≥ 2,6 bis < 3,2	≥ 3,2 bis ≤ 5,1	> 5,1
DAS28-CRP	<ul style="list-style-type: none"> • Anzahl geschwollener und druckschmerzhafter Gelenke (von 28) • CRP (mg/dl) • Gesundheitszustand 	< 2,6	≥ 2,6 bis < 3,2	≥ 3,2 bis ≤ 5,1	> 5,1
SDAI	<ul style="list-style-type: none"> • Anzahl geschwollener und druckschmerzhafter Gelenke (von 28) • PtGADA (cm) • PhGADA (cm) • CRP (mg/dl) 	≤ 3,3	> 3,3 bis ≤ 11	> 11 bis ≤ 22	> 26
CDAI	<ul style="list-style-type: none"> • Anzahl geschwollener und druckschmerzhafter Gelenke (von 28) • PtGADA (cm) • PhGADA (cm) 	≤ 2,8	> 2,8 bis ≤ 10	> 10 bis ≤ 22	> 22
ACR-EULAR Remission	<ul style="list-style-type: none"> • Index: SDAI, CDAI • Boolean: <ul style="list-style-type: none"> ○ Anzahl geschwollener und druckschmerzhafter Gelenke (von 28) ○ PtGADA (cm) ○ PhGADA (cm) ○ CRP (mg/dl) 	SDAI ≤ 3,3 CDAI ≤ 2,8 Boolean gesamt ≤ 1	-	-	-
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: modifiziert nach [15].					

Weiterhin ist es von hoher Wichtigkeit, das strukturelle Voranschreiten der Erkrankung zu evaluieren. Die Behandlung der RA sollte strukturelle Veränderung verhindern oder aufhalten und dadurch körperliche Einschränkungen minimieren oder umkehren [15].

Zur Messung der Krankheitsaktivität werden vornehmlich Instrumente verwendet, die die Anzahl (druck-)schmerzhafter bzw. geschwollener Gelenke einschließen. Sofern nach drei Monaten eine niedrige Krankheitsaktivität erreicht wird oder eine ca. 80%ige Verbesserung des SDAI oder CDAI eintritt, ist die Wahrscheinlichkeit sehr hoch, dass das Behandlungsziel einer Remission oder zumindest niedrigen Krankheitsaktivität sechs Monate nach Therapiebeginn erreicht wird [74]. Wenn die Verbesserung nach drei Monaten nur gering ist, sollte die Behandlung angepasst werden. Gleichermaßen sollte die Therapie überprüft werden, wenn das Ziel einer niedrigen Krankheitsaktivität oder einer Remission nach sechs Monaten nicht erreicht ist. Eine mögliche Eskalation der Therapie muss jedoch immer vor dem Hintergrund der individuellen Patientensituation und potentieller Nebenwirkungen betrachtet werden [14].

Zur Beurteilung der Morbidität bzw. Krankheitsaktivität, Symptomatik der RA und des körperlichen Funktionsstatus werden daher folgende patientenrelevanten Endpunkte herangezogen.

Druckschmerzhafte Gelenke (Tender Joint Count, TJC) und geschwollene Gelenke (Swollen Joint Count, SJC)

Gelenkentzündungen und die damit einhergehende Schwellung (Synovitis) und Druckschmerzhaftigkeit der betroffenen Gelenke sind wichtige und für den Patienten spürbare Symptome und eines der Hauptmerkmale der RA. Zur Beurteilung dieser Symptome werden *Joint Counts* durchgeführt, d. h. die Anzahl der betroffenen Gelenke wird erfasst. Dabei variieren die angewandten Erfassungsmethoden bezüglich der Anzahl der untersuchten Gelenke, bezüglich einer Gewichtung der Gelenke nach ihrer Größe und einer Skalierung der erfassten Schwellung oder Druckschmerzhaftigkeit [75]. Das bloße Zählen der betroffenen Gelenke gilt derzeit als zuverlässigste Messmethode. Zu den üblichen Methoden gehören der 66/68 *Joint Count*, der Ritchie Index [76] und der reduzierte 28 *Joint Count* [77-79].

Beim 66/68 *Joint Count* werden zur Beurteilung der Krankheitsaktivität bei Arthritis-Patienten 66 Gelenke auf Schwellung und 68 Gelenke auf ihre Druckschmerzhaftigkeit hin untersucht werden. Zu den zu untersuchenden Gelenken, die über den gesamten Körper verteilt sind, gehören u. a. das Kiefer-, Sternoklavikulär-, Schulterreck-, Schulter-, Ellenbogen-, Hand-, Daumengrund-, Fingermittel-, Fingerend-, Hüft-, Knie-, Knöchel- sowie Großzehengrundgelenk und die distalen Interphalangealgelenke der Füße sowie der Tarsus. Dabei werden die gleichen 66 Gelenke auf Schwellung und Druckschmerzhaftigkeit hin untersucht, die beiden Hüftgelenke werden lediglich auf Druckschmerzhaftigkeit untersucht, da die Hüftgelenke konstitutionsbedingt nicht bei jedem Menschen im Hinblick auf Schwellung beurteilt werden können. Endoprothetisch versorgte und ankylosierende Gelenke werden von der Bewertung ausgeschlossen. Mittels einer 2-Punkte-Skala wird der Grad der Schwellung und des Druckschmerzes der Gelenke ermittelt:

0 Punkte = Keine Schwellung/Druckschmerz

1 Punkt = Gelenkflüssigkeitsbedingte Verdickung/Druckschmerz vorhanden

Im Falle der RA werden häufig vereinfachte *Joint Counts* herangezogen und beispielsweise lediglich 28 Gelenke untersucht. In einer Studie von Smolen et al. konnte gezeigt werden, dass die Analyse von 28 Gelenken äquivalente Ergebnisse im Vergleich zur Untersuchung von 66/68 Gelenken brachte [80]. Die Reduktion der zu zählenden Gelenke auf 28 zur Ermittlung des Ausmaßes der RA erwies sich als valide und änderungssensitiv [78-79,81-85].

Beim 28 *Joint Count* werden 28 Gelenke als druckschmerzhaft oder nicht druckschmerzhaft bzw. als geschwollen oder nicht geschwollen für je 14 Gelenke einer Körperseite evaluiert (insgesamt zwei Schultergelenke, zwei Ellbogengelenke, zwei Handgelenke, zehn Großzehengrundgelenke, zehn Fingermittelgelenke und zwei Kniegelenke) [86].

In der Studie RA-BEAM (JADV) wurden 66 (Schwellung) bzw. 68 (Druckschmerz) Gelenke untersucht. Im Dossier wird die Änderung der Anzahl der druckschmerzhaften und geschwollenen Gelenke von Baseline von Baseline bis Woche 24 und Woche 52 als separater Endpunkt geführt, basierend auf dem 28 *Joint Count*. Darüber hinaus ist die Anzahl der SJC und TJC Grundlage verschiedener Konstrukte wie dem DAS28-hsCRP, SDAI und CDAI, mit denen das komplexe Krankheitsgeschehen der RA abgebildet wird. Für die Berechnung dieser Konstrukte wurde der 28 *Joint Count* herangezogen.

Die Druckschmerzhaftigkeit und Schwellung betroffener Gelenke sind für den Patienten unmittelbar spürbar, verbunden mit körperlichen Beeinträchtigungen und gehen folglich mit Einschränkungen des alltäglichen Lebens einher. Bei der Untersuchung der o. g. Gelenke handelt sich um eine validierte Methode, die auch in den Leitlinien zur Bewertung der Wirksamkeit von Therapien zur Behandlung der RA in klinischen Studien und auch vom IQWiG anerkannt ist und empfohlen wird [5,41,87-92]. Dieser Endpunkt ist somit patientenrelevant.

Disease Activity Score 28 High Sensitivity C-Reactive Protein (DAS28-hsCRP)

Der DAS ist ein klinisches Konstrukt zur Bewertung der Krankheitsaktivität der RA und kombiniert Informationen zu druckschmerzhaften und geschwollenen Gelenken, zur Akute-Phase-Reaktion und dem allgemeinen Gesundheitszustand. Der DAS28 basiert auf der Weiterentwicklung des 1983 entwickelten DAS und bezieht die Anzahl der betroffenen Gelenke, die Einschätzung der Krankheitsaktivität durch den Patienten (*Patient's Global Assessment of Disease Activity* (PtGADA)) sowie unspezifische serologische Entzündungsparameter wie die Blutkörperchengeschwindigkeit (BSG), auch Erythrozytensedimentationsrate (ESR) genannt, oder das Entzündungsprotein C-reaktives Protein (CRP) mit ein.

Eine Version des DAS28 ist der DAS-hsCRP, der statt der ESR das *High Sensitivity C-Reactive Protein* (hsCRP) beinhaltet, das eine Entzündung im Körper des Patienten quantitativ genauer und schneller im Verlauf anzeigt als die ESR und somit besser die Akute-Phase-Reaktion reflektiert [17] (Kapitel 2 - Diagnose und Prognose der frühen RA). Beim DAS28-hsCRP handelt es sich um ein validiertes und auf den EULAR-Kriterien basiertes Messinstrument, das die Schwere und den Verlauf der RA bewertet [16]. Höhere Werte spiegeln eine höhere Krankheitsaktivität wider. Der DAS28-hsCRP setzt sich aus folgenden patientenrelevanten Endpunkten sowie aus dem hsCRP (mg/dl) zusammen:

- Anzahl TJC (28 von 68 druckschmerzhafte Gelenken)
- Anzahl SJC (28 von 66 geschwollenen Gelenken)
- PtGADA (VAS; 0-100 mm) [93]

Der DAS28-hsCRP wird anhand der folgenden mathematischen Formel berechnet:

$$\text{DAS28-hsCRP} = 0,56 \sqrt{\text{TJC}28} + 0,28 \sqrt{\text{SJC}28} + 0,36 (\ln(\text{hsCRP}+1)) + 0,014 (\text{VAS}) + 0,96$$

Die Krankheitsaktivität gemessen mittels DAS28 wird gemäß ACR und EULAR in verschiedene Stufen eingeteilt. Bei einem DAS28-Score von < 2,6 liegt eine Remission vor [16]. Des Weiteren erfolgt die Einteilung der Krankheitsaktivität mittels DAS28 in die Stufen gering ($\geq 2,6$ bis < 3,2), mittelschwer ($\geq 3,2$ bis $\leq 5,1$) und hoch ($> 5,1$) (Tabelle 4-4) [15]. Besonders eine Remission, aber auch das Erreichen einer geringen Krankheitsaktivität, die die Lebensqualität, soziale Teilhabe und die radiologisch nachweisbare Gelenkschädigung (radiologische Progression) jeweils positiv beeinflussen [17], stellen wichtige und anerkannte Therapieziele dar [16,18-31].

Im Dossier wird die nachfolgende Auswertung des Endpunktes herangezogen:

- Veränderung des DAS28-hsCRP von Baseline bis Woche 24 und Woche 52
- Anteil Patienten mit DAS28-hsCRP < 2,6 an Woche 24 und Woche 52
- Anteil Patienten mit DAS28-hsCRP $\leq 3,2$ an Woche 24 und Woche 52

Somit kann das Erreichen wichtiger und anerkannter Therapieziele über das Erfassen des Anteils Patienten mit einer Remission (DAS28-hsCRP < 2,6) und des Anteils Patienten mit einer geringen Krankheitsaktivität (DAS28-hsCRP $\leq 3,2$) beurteilt werden [16,18-31].

Zusammenfassend lässt sich festhalten, dass der DAS bzw. DAS28 innerhalb der Fachgesellschaften und Rheumatologen ein akzeptiertes und validiertes Messinstrument zur Beurteilung der Krankheitsaktivität darstellt [5]. Dieser Score besteht aus patientenrelevanten Komponenten und besitzt somit als Konstrukt, mit dem das komplexe Krankheitsgeschehen der RA abgebildet wird, ebenfalls eine hohe Patientenrelevanz. Auch im IQWiG Bericht „*Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis*“ (Kapitel 4.1.3) ist die DAS28-Remission als patientenrelevanter Endpunkt anerkannt [5].

Simplified Disease Activity Index (SDAI)

Der SDAI ist ein zusammengesetztes und validiertes Konstrukt, das die Krankheitsaktivität der RA erfasst. Der SDAI besteht analog zu dem DAS28-hsCRP aus patientenrelevanten Einzelkomponenten (Anzahl TJC/SJC sowie PtGADA) und dem Akut-Phase-Protein hsCRP und bezieht zusätzlich Informationen zur Einschätzung der Krankheitsaktivität durch den Arzt (*Physician's Global Assessment of Disease Activity*, PhGADA) mit ein. Eine Remission liegt gemäß ACR und EULAR bei einem SDAI-Score von $\leq 3,3$ vor [15,20,27,32]. Des Weiteren wird die Krankheitsaktivität in die Stufen gering ($> 3,3$ bis 11), mittelschwer (> 11 bis ≤ 22) und hoch (> 26) eingeteilt [15] (Tabelle 4-4).

Der SDAI setzt sich somit aus folgenden Einzelkomponenten zusammen:

- Anzahl TJC (28 von 68 druckschmerzhaft Gelenken)
- Anzahl SJC (28 von 66 geschwollenen Gelenken)
- PtGADA (VAS 0-100 mm)
- PhGADA (VAS 0-100 mm)
- hsCRP (mg/dl)

Die einzelnen Messwerte werden ohne Umrechnung oder Gewichtung wie folgt addiert:

$$\text{SDAI} = \text{SJC (28)} + \text{TJC (28)} + \text{hsCRP (mg/dl) (0,1-10,0)} + \text{PtGADA} + \text{PhGADA}$$

Im Dossier werden die nachfolgenden Auswertungen des Endpunktes herangezogen:

- Veränderung des SDAI von Baseline bis Woche 24 und Woche 52
- Anteil Patienten mit $\text{SDAI} \leq 3,3$ an Woche 24 und Woche 52

Clinical Disease Activity Index (CDAI)

Der CDAI ist ein validiertes Konstrukt (ohne Akute-Phase-Protein), mit dem das komplexe Krankheitsgeschehen der RA abgebildet wird. Dieser Score setzt sich aus den folgenden Einzelkomponenten zusammen [94]:

- Anzahl TJC (28 von 68 druckschmerzhaft Gelenken)
- Anzahl SJC (28 von 66 geschwollenen Gelenken)
- PtGADA (VAS 0-100 mm)
- PhGADA (VAS 0-100 mm)

Die Berechnung dieses Scores erfolgt anhand der folgenden mathematischen Formel ohne Umrechnung und Gewichtung:

$$\text{CDAI} = \text{SJC (28)} + \text{TJC (28)} + \text{PtGADA} + \text{PhGADA}$$

Ebenso wie der SDAI ist der CDAI ein validiertes Konstrukt, mit dem das komplexe Krankheitsgeschehen der RA abgebildet wird. Ein CDAI-Score von $\leq 2,8$ zeigt gemäß ACR und EULAR eine Remission an [15,20,27,32]. Des Weiteren wird die Krankheitsaktivität in die Stufen gering ($> 2,8$ bis ≤ 10), mittelschwer (> 10 bis ≤ 22) und hoch (> 22) eingeteilt [15] (Tabelle 4-4). Der CDAI besteht analog zu dem SDAI aus patientenrelevanten Einzelkomponenten wie der Anzahl TJC/SJC sowie PtGADA und PhGADA. Das Akut-Phase-Protein hsCRP wird im Gegensatz zum SDAI für die Berechnung des CDAI nicht berücksichtigt. Der CDAI besitzt wie der SDAI als Konstrukt ebenfalls eine hohe Patientenrelevanz [15,33-34].

Im Dossier wird die nachfolgende Auswertung des Endpunktes herangezogen:

- Veränderung des CDAI von Baseline bis Woche 24 und Woche 52
- Anteil Patienten mit CDAI ≤ 10 an Woche 24 und Woche 52

EULAR Responsekriterien („good Response“, DAS28-basiert)

Mit Hilfe der validierten EULAR Responsekriterien wird das Therapieansprechen eines Patienten auf Grundlage der Veränderung und des Post-Baseline Levels des DAS28 beurteilt [35]. Patienten werden als *Responder* oder *Non-Responder* klassifiziert. Das Ansprechen wird weiterhin unterteilt in *gutes Ansprechen* oder *moderates Ansprechen*. Ein *gutes Ansprechen* liegt bei einem Post-Baseline DAS28 $\leq 3,2$ und einer Verbesserung im Vergleich zu Baseline von DAS28 $> 1,2$ vor. Ein *moderates Ansprechen* liegt bei einem Post-Baseline DAS28 $> 3,2$ bis $\leq 5,1$ und einer Verbesserung im Vergleich zu Baseline von DAS28 $\leq 1,2$ und $> 0,6$ vor, wohingegen bei einem Post-Baseline DAS28 $> 5,1$ und einer Verbesserung im Vergleich zu Baseline von DAS28 $\leq 0,6$ kein Ansprechen vorliegt [19].

Die EULAR *Response* (DAS28-basiert) wird wie folgt kategorisiert (Tabelle 4-5):

Tabelle 4-5: Kategorisierung der Patienten ohne Ansprechen (*Non-Responder*), mit *moderatem Ansprechen* oder mit *gutem Ansprechen* mittels EULAR Responsekriterien (DAS28-basiert)

DAS28 Post-Baseline Level	Verbesserung im Vergleich zu Baseline (DAS28)		
	$> 1,2$	$\leq 1,2$ und $> 0,6$	$\leq 0,6$
$\leq 3,2$	gutes Ansprechen	moderates Ansprechen	kein Ansprechen
$> 3,2$ bis $\leq 5,1$			
$> 5,1$			

Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.
Quelle : [19].

Im Dossier wird die Anzahl Patienten mit EULAR „good Response“ (DAS28-basiert) an Woche 24 und Woche 52 dargestellt.

Sowohl der DAS28 als Konstrukt, mit dem das Krankheitsgeschehen der RA abgebildet werden soll, als auch seine Einzelkomponenten sind patientenrelevant. Somit sind die EULAR Responsekriterien, deren Grundlage der DAS28 ist, ebenfalls als patientenrelevant anzusehen.

Patient Assessment of Pain (PAP) (VAS)

Schmerz ist ein unmittelbar vom Patienten wahrgenommener und somit patientenrelevanter Endpunkt. Schmerz ist eines der Hauptsymptome der RA und ist aus Patientensicht eines der schwerwiegendsten Symptome [95]. In einer Untersuchung von Gossec et al. konnte gezeigt werden, dass die sieben am wichtigsten durch die Patienten eingestuften Domänen, inklusive ihrer Gewichtung, die folgenden sind: Schmerz (21%), funktionelle Beeinträchtigung (16%), Fatigue (15%), emotionales Wohlbefinden (12%), Schlaf (12%), Krankheitsverarbeitung (12%) und physischem Wohlbefinden (12%) [95]. Die meisten RA-Patienten bewerten den Schmerz als größtes Problem mit der höchsten Priorität [40,96-101]. Ca. 70% der Patienten würden eine Verbesserung der Behandlung von Schmerzen begrüßen [98]. Gemäß ACR *Pain Management Task Force* ist Schmerz wahrscheinlich der wichtigste durch Patienten berichtete Endpunkt in der Rheumatologie, der eventuell mehr zur Beeinträchtigung der RA-Patienten beiträgt als strukturelle Gelenkschäden [102-103]. Die *Quantitative Standard Monitoring of Patients with Rheumatoid Arthritis* (QUEST-RA) Studie [104], die 7.028 Patienten untersuchte, zeigte, dass Schmerz auch der entscheidendste Faktor bezüglich der globalen Einschätzung der Krankheitsaktivität durch den Patienten ist und ferner der wichtigste prädiktive Faktor der psychosozialen Gesundheit der RA-Patienten [105]. Das individuelle Schmerzempfinden beeinträchtigt das tägliche Wohlempfinden und hat somit einen erheblichen Einfluss auf die Lebensqualität der Betroffenen [41-44].

Visuelle Analogskalen (VAS) sind vor allem bei der Messung von Schmerz und Stimmungslagen etablierte Messinstrumente [106]. Mittels VAS (0 - 100 mm) wird der Patient gebeten, seine empfundene Schmerzintensität zu beurteilen, wobei ein Wert von 0 mm bedeutet, dass der Patient keine Schmerzen bzw. ein Wert von 100 mm bedeutet, dass der Patient die stärksten vorstellbaren Schmerzen verspürt. Eine Abnahme des Wertes spiegelt somit eine Verbesserung wider. In der Literatur werden als *Minimal Clinically Important Difference (MCID)* des PAP (VAS), also des Unterschieds auf einer visuellen Analogskala der mindestens notwendig ist, um als klinisch relevant eingestuft zu werden, unterschiedliche Werte angegeben. Dabei reichen die MCIDs von 4 mm bis 20,4 mm [107-114]. Eine Veränderung von PAP (VAS) ≥ 10 mm ist etabliert und anerkannt [107-108,112], um eine klinisch relevante Veränderung des Schmerzes nachzuweisen, ebenso validiert ist der höhere Cut-off für eine Veränderung von PAP (VAS) ≥ 20 mm [111,114].

Im Dossier werden die nachfolgenden Auswertungen des Endpunktes dargestellt:

- Veränderung des PAP (VAS) von Baseline bis Woche 24 und Woche 52
- Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm von Baseline bis Woche 24 und Woche 52
- Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm von Baseline bis Woche 24 und Woche 52

Patient's Global Assessment of Disease Activity (PtGADA) (VAS)

Eine weitere Einzelkomponente verschiedener patientenrelevanter Konstrukte zur Beurteilung der Krankheitsaktivität ist die globale Einschätzung der Krankheitsaktivität durch den Patienten (PtGADA). Dieser Endpunkt spiegelt die aus Sicht des Patienten wahrgenommene Krankheitsaktivität wider und gibt somit Aufschluss über den Krankheitszustand und damit auch über die damit assoziierte Lebensqualität.

Die Schwere der Symptome werden auf einer VAS von 0 - 100 mm erfasst, wobei ein Wert von 0 mm bedeutet, dass der Patient keine Symptome hat bzw. ein Wert von 100 mm bedeutet, dass der Patient die stärksten vorstellbaren Symptome verspürt. Daher zeigt ein geringer Wert eine niedrige Krankheitsaktivität an. Im Studienverlauf spiegelt eine Abnahme des Wertes eine Verminderung der Krankheitsaktivität wider, die auch mit einer Verbesserung der Lebensqualität assoziiert ist.

In der Literatur werden als MCID für die globale Einschätzung der Krankheitsaktivität durch den Patienten (VAS) unterschiedliche Werte angegeben, dabei reichen die MCIDs von 6 mm bis 20 mm [111-112,114]. Eine Veränderung von PtGADA (VAS) ≥ 10 mm ist etabliert und anerkannt [112], um eine klinisch relevante Veränderung nachzuweisen. Ebenso validiert ist der höhere Cut-off für eine Veränderung von PtGADA (VAS) ≥ 20 mm [111,113-114].

Im vorliegenden Dossier werden die nachfolgenden Auswertungen des Endpunktes dargestellt:

- Veränderung des PtGADA (VAS) von Baseline bis Woche 24 und Woche 52
- Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 10 mm von Baseline bis Woche 24 und Woche 52
- Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm von Baseline bis Woche 24 und Woche 52

Der PtGADA ist patientenrelevant, da der Krankheitszustand eines Patienten durch den Patienten selbst beurteilt wird. Im Verlauf einer Studie kann mittels PtGADA der Gesundheitszustand eines Patienten, welcher mit der Lebensqualität korreliert, beurteilt werden [41-42,114].

Health Assessment Questionnaire-Disability Index (HAQ-DI)

Der Patientenfragebogen HAQ-DI ist ein generisches Selbstbeurteilungsinstrument, der die krankheitsassoziierten Einschränkungen der physischen Funktionalität der Patienten, beides krankheitsspezifische Symptome der RA und somit patientenrelevant, erfasst. In einer Untersuchung von Gossec et al. konnte gezeigt werden, dass die funktionelle Beeinträchtigung aus Patientensicht die zweitbedeutendste Belastung darstellt [95]. Der validierte HAQ-DI, erstmals 1980 publiziert [81,115], besteht aus den folgenden acht Domänen:

- Anziehen/Ankleiden und Körperpflege (2 Fragen)
- Aufstehen (2 Fragen)
- Essen (3 Fragen)
- Gehen (2 Fragen)
- Hygiene (3 Fragen)
- Erreichbarkeit von Gegenständen (2 Fragen)
- Greifen (3 Fragen)
- Allgemeine tägliche Aktivitäten (3 Fragen)

Die Bewertung erfolgt durch eine 4-Punkte-Likert-Skala von null bis drei (0 = ohne Schwierigkeiten, 1 = mit einigen Schwierigkeiten, 2 = mit vielen Schwierigkeiten, 3 = unfähig zur Durchführung). Zudem wurden die berichtete Verwendung von speziellen Hilfsmitteln oder Vorrichtungen und/oder der Bedarf an Unterstützung durch eine andere Person bei der Durchführung bewertet.

Über die Mittelwerte der einzelnen Domänen wird der *Functional Disability Index* berechnet. Ein geringer Wert zeigt eine bessere physische Funktionalität bzw. eine Abnahme des Wertes spiegelt eine Verbesserung der physischen Funktionalität wider. In der Literatur werden als Cut-off für eine klinisch relevante Verbesserung der körperlichen Funktionen Werte von 0,2 bis 0,3 angegeben [33,109-110,116]. Ein MCID von HAQ-DI $\geq 0,22$ reflektiert eine klinisch relevante Verbesserung der körperlichen Funktionen. Dieser Schwellenwert wurde in zahlreichen Studien mit RA-Patienten validiert [33,109]. Ebenso dargestellt wurde der höhere Cut-off von HAQ-DI $\geq 0,3$, der von der FDA gefordert wurde.

Im vorliegenden Dossier werden die folgenden Auswertungen dargestellt:

- Veränderung von Baseline bis Woche 24 und Woche 52
- Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$ von Baseline bis Woche 24 und Woche 52
- Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$ von Baseline bis Woche 24 und Woche 52

Der HAQ-DI gibt Auskunft über die krankheitsbedingten körperlichen Einschränkungen im Alltag, die vom Patienten selbst bewertet werden und steht somit im Zusammenhang mit der Krankheitslast als Maß für die *Morbidität*. Eine Verschlechterung des HAQ-DI ist direkt mit Einschränkungen im täglichen Leben verbunden, weshalb von einem negativen Einfluss auf die gesundheitsbezogene Lebensqualität ausgegangen werden kann.

Dauer der Morgensteifigkeit (Duration of Morning Joint Stiffness, MJS)

Bei der Morgensteifigkeit (MJS) handelt es sich um ein typisches Begleitsymptom der RA und die Dauer der Morgensteifigkeit korreliert mit der körperlichen Einschränkung des Patienten und dessen Lebensqualität [45]. Morgensteifigkeit beschreibt eine Einschränkung der Beweglichkeit der Gelenke, die direkt nach dem Aufwachen von den Patienten wahrgenommen wird. Sie führt zu einer starken Einschränkung der üblichen morgendlichen Tätigkeiten wie beispielsweise dem Aufstehen und der Körperpflege. Aus Patientensicht ist die Dauer der Morgensteifigkeit ein extrem beeinträchtigendes Krankheitssymptom, welches starken Einfluss auf die Lebensqualität, den Funktionsstatus und die Arbeitsfähigkeit ausübt und welches ein wichtiges und klinisch relevantes Element der Krankheitsaktivität darstellt [117-119]. RA-Patienten beschreiben die Dauer der Morgensteifigkeit als wichtige Konsequenz der Erkrankung und sehen sie als relevantes Symptom, welches sich von anderen Symptomen der RA wie Schmerz unterscheidet. Viele der Patienten beschreiben, dass die Steifigkeit der Gelenke einen großen Einfluss auf die Lebensqualität haben kann. Speziell die Mobilität, Balance, Geschicklichkeit und der Bewegungsradius seien negativ durch die RA-bedingte Steifigkeit der Gelenke beeinflussbar [120-122].

Eine Umfrage unter 518 Rheumatologen und mit 750 RA-Patienten in elf europäischen Ländern zeigte, dass 92% der Patienten die Morgensteifigkeit als mittelschwer bis schwer beeinträchtigend beschreiben und 82% der Patienten angeben, dass sie einen erheblichen negativen Einfluss auf ihre Lebensqualität ausübt. 95% der befragten Rheumatologen stimmten diesem Ergebnis zu [123]. Es handelt sich bei der Dauer der Morgensteifigkeit demnach um einen patientenrelevanten Endpunkt.

Khan et al., 2009 nehmen für den patientenberichteten Endpunkt Morgensteifigkeit (Dauer) – *Morning Joint Stiffness (MJS)* – eine Einteilung in Keine: 0 Minuten, mild: 1 – 30 Minuten, mittelschwer: 31 – 60 Minuten und schwer: > 60 Minuten vor, die als Rationale zur Kategorisierung in eine der Endpunktkategorien herangezogen werden kann [124].

In der Studie RA-BEAM (JADV) erfolgte die Erhebung der Dauer der Morgensteifigkeit durch den Patienten mit Hilfe eines elektronischen *Patient Reported Outcome* (ePRO)-Tablets. Der Patient dokumentierte die Dauer der Morgensteifigkeit der von der RA betroffenen Gelenke (in Minuten) am Tag vor der jeweiligen Studienvisite.

Im vorliegenden Dossier wird die Veränderung der Dauer der Morgensteifigkeit von Baseline bis Woche 24 und Woche 52 dargestellt.

EQ-5D VAS

Mittels EQ-5D, einem krankheitsübergreifenden und validierten Fragebogen, kann die gesundheitsbezogene Lebensqualität eines Patienten ermittelt werden [46]. Dieser 5-Komponenten-Fragebogen besteht aus den EQ-5D *Utility Items*, wozu die Dimensionen Mobilität, Fähigkeit sich selbst zu versorgen, alltägliche Aktivitäten, Schmerz/körperliche Beschwerden, Ängste/Niedergeschlagenheit gehören, sowie der EQ-5D VAS.

Die EQ-5D VAS bewertet den allgemeinen Gesundheitszustand eines Patienten und kategorisiert diesen auf einer Skala von null „schlechtester vorstellbarer Gesundheitszustand“ („*Worst Imaginable Health State*“) bis 100 „bester vorstellbarer Gesundheitszustand“ („*Best Imaginable Health State*“). Eine Zunahme des Wertes spiegelt somit eine Verbesserung des Gesundheitszustandes des Patienten wider [46,125-126].

Für das vorliegende Dossier wurde eine *Minimal Important Difference* (MID) von 10 mm betrachtet. Eine spezifisch in einer RA-Population validierte MID für die EQ-5D VAS konnte nicht identifiziert werden. Häufig wird in Nutzenbewertungen für die EQ-5D VAS eine MID zwischen 7 mm und 12 mm herangezogen [127-128]. Auch wenn diese MID in einer Krebspopulation hergeleitet wurde, so stellt sie doch die einzig verfügbare Referenz für die Herleitung einer MID für die EQ-5D VAS dar [129]. Daher wurde, in bester Annäherung daran, ein Responsekriterium von 10 mm für die hier durchgeführten Analysen gewählt.

Im vorliegenden Dossier werden die folgenden Auswertungen dargestellt:

- Veränderung der EQ-5D VAS von Baseline bis Woche 24 und Woche 52
- Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm von Baseline bis Woche 24 und Woche 52

Zusammenfassend ist der EQ-5D ein standardisierter und validierter Fragebogen, der indikationsunabhängig eine Einschätzung zu Gesundheitszustand und somit zur Lebensqualität des Patienten wiedergibt. Vor allem die VAS ist ein verlässliches Messinstrument und ist sensitiv für Veränderungen des Gesundheitszustandes im Krankheitsverlauf eines Patienten. Es handelt sich dabei um einen patientenrelevanten Endpunkt.

Schwerste Müdigkeit (Worst Tiredness)

Müdigkeit ist ein häufiges Begleitsymptom der RA und hat direkten Einfluss auf die Lebensqualität des Patienten.

Fatigue und Müdigkeit sind wiederkehrende und stark beeinträchtigende Probleme für RA-Patienten [47], und werden von Patienten, nach dem Schmerz, als belastendste Krankheitssymptome beschrieben [48], welche einen großen Einfluss auf die Lebensqualität der Patienten ausüben [49-50].

Während die Fatigue ein multi-dimensionales Symptomkonzept ist [51-54], ist die Müdigkeit eine Schlüsselkomponente dieses Symptomkonzeptes. Mittels einer numerischen 11-Punkte Skala (*Numeric Rating Scale*, NSR) wurde der Patient gebeten, unter Verwendung eines elektronischen *Patient Reported Outcome* (ePRO)-Tablets, seine Müdigkeit während der letzten 24 Stunden selbst einzuschätzen. Bei einem Wert von 0 liegt keine Müdigkeit vor, während bei einem Wert von 10 eine Müdigkeit „so schlecht wie nur vorstellbar“ vorliegt. Eine Abnahme des Wertes stellt somit eine Verbesserung des Symptoms Müdigkeit bzw. der Lebensqualität dar.

Im vorliegenden Dossier wird die Veränderung der schwersten Müdigkeit von Baseline von Baseline bis Woche 24 und Woche 52 dargestellt. Müdigkeit als Symptom beeinträchtigt den Patienten in seinen alltäglichen Aktivitäten und mindert folglich seine Lebensqualität. Es ist somit ein patientenrelevanter Endpunkt [2].

Lebensqualität

SF-36 v2 akut

Der SF-36 ist ein generischer und international anerkannter Fragebogen zur Einschätzung der Lebensqualität durch den Patienten [55]. Dieses Messinstrument, das für die RA validiert und in zahlreichen Studien verwendet wurde [56-57], bildet die folgenden acht Dimensionen mit insgesamt 36 Fragen der allgemeinen Gesundheit ab, die wiederum in der physischen (*Physical Component Score*, PCS) und mentalen (*Mental Component Score*, MCS) Summenskala zusammengefasst werden [130-131]:

PCS

- Körperliche Funktionsfähigkeit (10 Items)
- Körperliche Rollenfunktion (4 Items)
- Körperliche Schmerzen (2 Items)
- Allgemeiner Gesundheitszustand (5 Items)

MCS

- Vitalität (4 Items)
- Soziale Funktionsfähigkeit (2 Items)
- Emotionale Rollenfunktion (3 Items)
- Mentaler Gesundheitszustand (5 Items)

Die Berechnung der SF-36 Summenskalen wird von Ware et al. (2014) detailliert beschrieben [132]. Ein hoher Wert zeigt einen besseren physischen oder mentalen Gesundheitsstatus bzw. eine Zunahme des Wertes spiegelt eine Verbesserung der physischen oder mentalen Komponente des Gesundheitsstatus wider.

In der Literatur werden als MCID für den SF-36 unterschiedliche Werte angegeben, dabei reichen diese von 2,5 Punkte, über 3 Punkte bis 5 Punkte auf der physischen und mentalen Summenskala [133-137]. Für dieses Nutzendossier wird die präspezifizierte und im *Clinical Study Report* (CSR) verwendete *Minimal Important Difference* (MID) von 5 Punkten verwendet.

Im vorliegenden Dossier werden die folgenden Auswertungen zur Beurteilung der Lebensqualität herangezogen:

- Veränderung des SF-36 v2 akut PCS von Baseline bis Woche 24 und Woche 52
- Anteil an Patienten mit SF-36 v2 akut PCS ≥ 5 von Baseline bis Woche 24 und Woche 52
- Veränderung des SF-36 v2 akut MCS von Baseline bis Woche 24 und Woche 52
- Anteil an Patienten mit SF-36 v2 akut MCS ≥ 5 von Baseline Woche 24 und Woche 52
- Veränderung aller SF-36 v2 akut Domänen von Baseline bis Woche 24 und Woche 52

Die Endpunkte der Nutzendimension *Gesundheitsbezogene Lebensqualität* fallen gemäß der „Allgemeinen Methoden“ des IQWiG unter die Kategorie *schwerwiegende (bzw. schwere) Symptome, Folgekomplikationen und Nebenwirkungen* [6].

Sicherheit

In diesem Dossier werden im Rahmen der Nutzenbewertung die Gesamtrate der Todesfälle, unerwünschter Ereignisse (UE), schwerwiegender unerwünschte Ereignisse (SUE), Studienabbruch/Therapieabbruch aufgrund unerwünschter Ereignisse sowie Studienabbruch aufgrund nicht ausreichender Wirksamkeit oder *Rescue*-Therapy und unerwünschte Ereignisse von besonderem Interesse (*Adverse Events of Special Interest*, AESI) berücksichtigt. Nach § 35a SGB V wird der patientenrelevante Effekt unter anderem als Verringerung von Nebenwirkungen definiert [73]. Die AM-NutzenV hebt die Patientenrelevanz dieses Endpunkts hervor [3]. Alle Arzneimittel üben neben erwünschten auch unerwünschte Wirkungen auf den menschlichen Körper aus. Die Bedeutung unerwünschter Arzneimittelwirkungen ist abhängig von ihrer Schwere, der Häufigkeit sowie der Reversibilität ihrer Symptome. Die Verringerung von unerwünschten Wirkungen ist ein wichtiger Aspekt der therapiebedingten *Morbidität* und ein vom IQWiG anerkannter patientenrelevanter Endpunkt [6].

UE werden mittels *electronic Case Report Form* (eCRF) erfasst und mit Hilfe der SOC oder eines *Preferred Term* (PT) unter Verwendung des *Medical Dictionary for Regulatory Activities* (MedDRA) Version 18.0 kodiert.

Insgesamt wurden zur Beurteilung der Sicherheit folgende patientenrelevanten Endpunkte erhoben:

- Todesfälle
- SUE
- UE
- Schwere UE (ohne ausschließliche Laborereignisse)
- Studienabbruch/Therapieabbruch aufgrund von UE
- UE von besonderem Interesse
 - Infektionen
 - Schwere Infektionen
 - Kardiovaskuläre UE
 - Malignität
 - Depressionen: QIDS-SR16
 - Allergische Reaktionen/Hypersensitivität
 - Gastrointestinale Perforation

Todesfälle

Die Verlängerung des Überlebens ist aus Patientensicht zweifellos der wichtigste Endpunkt. Auch die AM-NutzenV hebt die Patientenrelevanz dieses Endpunkts hervor [3].

Schwerwiegende unerwünschte Ereignisse (SUE)

Nach § 35a SGB V wird der patientenrelevante Effekt unter anderem als Verringerung von Nebenwirkungen [73]. Die AM-NutzenV hebt die Patientenrelevanz dieses Endpunkts hervor [3]. Ein schwerwiegendes unerwünschtes Ereignis, zudem übereinstimmend mit der ICH E2A Leitlinie [138], war jedes unerwünschte Ereignis, das eines der folgenden Konsequenzen zur Folge hatte:

- Tod
- Lebensbedrohliche Umstände (unmittelbares Risiko zu Sterben)
- Erst- oder verlängerte Hospitalisierung
- Anhaltende oder erhebliche Behinderung/Invalidität
- Angeborene Anomalie/Geburtsdefekt
- Durch den Prüfarzt aus einem anderen Grund als erheblich erachtet.

Unerwünschte Ereignisse (UE)

Alle in der vorliegenden Nutzenbewertung aufgeführten unerwünschten Ereignisse waren unerwünschte Ereignisse, die nach mindestens einer Dosis von Baricitinib aufgetreten waren (*Treatment-Emergent Adverse Event*, TEAE).

Schwere Unerwünschte Ereignisse (schwere UE, ohne ausschließliche Laborereignisse)

Alle in der vorliegenden Nutzenbewertung aufgeführten schweren unerwünschten Ereignisse wurden ohne Laborereignisse ausgewertet, in deren Zusammenhang keine Symptome beschrieben wurden oder dies angenommen werden musste und die somit asymptomatisch waren.

Dabei handelte es sich unter der Behandlung mit Baricitinib um die folgenden Ereignisse:

- Bluthochdruck, keine assoziierten UE (1 Patient)
- Lymphozytose an verschiedenen Zeitpunkten (inkl. Baseline), keine Symptome, keine assoziierten UE, keine Leukämie in Knochenmarksbiopsie (1 Patient)
- Erhöhte Kreatinphosphokinase (*Creatin Phosphokinase*, CPK > 3.000 U/I) nach anstrengendem Training im Fitnessstudio, keine kardialen oder muskulären Ereignisse, Patient hat Therapie in RA-BEYOND weitergeführt (1 Patient)
- Thrombozytose, bereits vorbestehend, keine thrombotischen Ereignisse (1 Patient)
- Hypercholesterinämie (2 Patienten)

Unter der Behandlung mit Adalimumab traten folgende Ereignisse auf:

- Alanin-Aminotransferase (ALT)/ Aspartat-Aminotransferase (AST) erhöht (1 Patient)

Diese schweren UE waren ausschließlich Laborereignisse ohne assoziierte Symptomatik und daher für den Patienten nicht zu spüren, also nicht patientenrelevant. Sollte eine mit diesen Laborereignissen in Zusammenhang stehende Symptomatik aufgetreten sein, dann wurde diese Symptomatik innerhalb der unerwünschten Ereignisse erfasst.

UE von besonderem Interesse

Unerwünschte Ereignisse von besonderem Interesse waren in der Studie RA-BEAM (JADV) gelistet als:

- Infektionen
- Schwere Infektionen
- Kardiovaskuläre UE
- Malignität
- Depressionen: QIDS-SR16
- Allergische Reaktionen/Hypersensitivität
- Gastrointestinale Perforation

Infektionen und schwere Infektionen

Infektionen wurden über PTs der SOC “Infektionen und parasitäre Erkrankungen” und der SOC „Untersuchungen“ definiert. Insgesamt wurden folgende behandlungsbedingte Infektionen erhoben:

- alle Infektionen
 - alle PTs der SOC “Infektionen und parasitäre Erkrankungen”
- schwere Infektionen
 - alle PTs der SOC “Infektionen und parasitäre Erkrankungen”, die als schwer kategorisiert sind
- Infektionen, die eine therapeutische Intervention nach sich zogen (Antibiotika, antivirale Arzneimittel, Antimykotika)
- Herpes Zoster
- Tuberkulose
- virale Hepatitis
- potentiell opportunistische Infektionen

Kardiovaskuläre UE

Kardiovaskuläre UE wurden gemäß “*Major Adverse Cardiovascular Events*” (MACE) und anderer kardiovaskulärer Ereignisse, die einer Adjudizierung bedurften, erhoben:

- MACE
 - kardiovaskulär bedingter Todesfall
 - Myokardinfarkt (MI)
 - Schlaganfall
- andere kardiovaskuläre Ereignisse, die einer Adjudizierung bedurften
 - Krankenhausaufenthalt wegen instabiler Angina
 - Krankenhausaufenthalt wegen Herzinsuffizienz
 - schwere Arrhythmien
 - Reanimation nach plötzlichem Herztod
 - kardiogener Schock
 - Koronarinterventionen (koronare Bypassoperation oder perkutane Koronarinterventionen)
- nicht kardiovaskuläre Todesfälle

Malignität

Maligne Erkrankungen wurden über die Standardisierte MedDRA-Abfrage (*Standardised MedDRA Query*, SMQ) SMQ 20000194 identifiziert.

Depressionen: 16-Item Quick Inventory of Depressive Symptomatology-Self-Report (QIDS-SR16)

Der QIDS-SR16 ist ein aus 16 Items bestehender Patientenfragebogen, der die neun DSM-IV (*Diagnostic and Statistical Manual of Mental Disorders-IV*) Diagnosesymptomgebiete, die für eine depressive Episode sprechen, erfasst. Die neun Gebiete enthalten:

1. Traurige Stimmung
2. Konzentration
3. Kritik des Selbstbildes
4. Suizidgedanken
5. Interessen
6. Energie/Ermüdung
7. Schlafstörungen
8. Ab- oder Zunahme des Appetits oder des Gewichts
9. Psychomotorische Bewegung, Verlangsamung, Unruhe

Depressionen betreffen einen großen Teil der RA-Patienten. Manche Studien zeigen, dass depressive Symptome in mindestens 25% der Patienten vorhanden sind. Zusätzlich leiden viele Patienten vermehrt an Angststörungen. Depressionen sind mit einem reduzierten Gesundheitszustand und einer geringeren Lebensqualität sowie einem höheren Schmerzniveau und einer stärkeren Ausprägung der Fatigue verbunden [139-141].

Allergische Reaktionen/Hypersensitivität

Behandlungsbedingte allergische Reaktionen und Hypersensitivität wurden im Kontext von anaphylaktischen Reaktionen erhoben und mittels SMQ 20000021 identifiziert. Anaphylaktische Reaktionen werden als schwerwiegende, schnell auftretende allergische Reaktionen, die sogar zum Tod führen können, definiert [142-143].

Gastrointestinale Perforation

Behandlungsbedingte unerwünschte Ereignisse, die potentiell im Zusammenhang mit gastrointestinalen Perforationen standen wurden mittels der SMQ 20000107 identifiziert.

Die individuelle Therapiestrategie sollte neben der Krankheitsaktivität auch weitere Faktoren wie radiologische Progression, Begleiterkrankungen, Sicherheitsaspekte sowie Teilhabe (z. B. Erhalt der Arbeitsfähigkeit und Einbezogenheit in das soziale Umfeld) berücksichtigen [5]. Die Beeinflussung von Nebenwirkungen, sowohl in Bezug auf Häufigkeit als auch auf Art und Schwere als ein Aspekt der therapiebedingten Morbidität, wird zudem vom IQWiG unter Bezugnahme auf § 35b SGB V zu den patientenrelevanten Endpunkten gezählt [6].

Auswertung der Endpunkte

Die Ergebnisse zu ausgewählten patientenrelevanten Endpunkten wurden entsprechend der vorliegenden Operationalisierung dargestellt und beschrieben. Hierbei wurden die im Studienbericht dargestellten Maßzahlen (präspezifizierte Analysen), wie z. B. adjustierte Mittelwerte mit Standardfehler bzw. Anteile und entsprechende Behandlungsunterschiede, berichtet.

Für kontinuierliche Endpunkte wurden tabellarisch die Anzahl der analysierten Patienten, Mittelwerte (MW) mit Standardabweichung (SD) zu Studienbeginn, Woche 24 und Studienende (Woche 52) sowie die Veränderung von Studienbeginn bis Woche 24 und Woche 52 durch den *Least Square Mean* (LSM) pro Behandlungsgruppe dargestellt. Der Behandlungseffekt wurde durch die *Least Square Mean Differenz* (LSMD) zwischen den Behandlungsgruppen mit 95%-Konfidenzintervall (KI) und p-Wert oder durch standardisierte Mittelwertdifferenzen (Hedges'g) quantifiziert.

Alle Vergleiche wurden für Baricitinib mit MTX als Kombinationstherapie minus der zVT Adalimumab mit MTX berechnet.

Für dichotome Endpunkte zeigt die tabellarische Ergebnisdarstellung die Anzahl der Patienten, die in die Analyse eingegangen sind sowie die Anzahl und den Anteil der Patienten mit einem Ereignis pro Behandlungsgruppe. Zusätzlich wurden für dichotome Zielvariablen das *Odds Ratio* (OR), das *Risk Ratio* (relatives Risiko, RR) und die *Absolute Risk Difference* (absolute Risikodifferenz, ARD) berechnet. Lagen präspezifizierte, adjustierte Analysemodelle vor, so wurden diese angewandt. Ansonsten waren die Analysen unadjustiert (Tabelle 4-6). Alle Effektmaße sind mit den jeweiligen Intervallschätzern angegeben.

Das RR und das OR wurden für Baricitinib mit MTX als Kombinationstherapie gegenüber der zVT Adalimumab mit MTX berechnet. Die ARD wurde für Baricitinib mit MTX als Kombinationstherapie minus der zVT Adalimumab mit MTX berechnet.

Die Interpretation der Ergebnisse für dichotome Endpunkte lautet wie folgt:

Bei negativ besetzten Ereignissen bedeuten die Effektschätzer RR bzw. OR < 1 einen Vorteil für Baricitinib und ein RR bzw. OR > 1 einen Vorteil für die zVT Adalimumab. Für den Effektschätzer ARD bedeuten Ergebnisse < 0 einen Vorteil für Baricitinib und Ergebnisse > 0 einen Vorteil für die zVT Adalimumab.

Bei positiv besetzten Ereignissen bedeuten Effektschätzer RR bzw. OR > 1 einen Vorteil für Baricitinib und RR bzw. OR < 1 einen Vorteil für die zVT Adalimumab. Für den Effektschätzer ARD bedeuten Ergebnisse > 0 einen Vorteil für Baricitinib und Ergebnisse < 0 einen Vorteil für die zVT Adalimumab.

Falls in beiden Behandlungsarmen keine Ereignisse auftraten oder alle Patienten Ereignisse hatten, wurden RR, OR und ARD nicht berechnet.

Tabelle 4-6: Übersicht der Analyseansätze für dichotome Endpunkte der Nutzendimension „Sicherheit“ (unadjustierte Analysen)

Vierfeldertafel-Szenario	Schätzer	95% KI	p-Wert
Häufigkeiten = 0 in beiden Behandlungsgruppen	Keine Analysen		
Häufigkeit ≥ 1 in einer Behandlungsgruppe und 0 in der anderen Gruppe	RR -> Peto OR OR: + 0.5 ^b ARD	asymptot. KI asymptot. KI ^b exaktes KI ^a	Fisher's exakter Test Fisher's exakter Test Z-Test ^c
Häufigkeiten ≥ 1 in beiden Behandlungsgruppen, aber $\leq 1\%$ in mind. einer Zelle	RR -> Peto OR OR ARD	asymptot. KI exaktes KI exaktes KI ^a	Fisher's exakter Test Fisher's exakter Test Z-Test ^c
Häufigkeiten ≥ 1 in beiden Behandlungsgruppen, $> 1\%$ in allen Zellen, aber $\leq 5\%$ in mindestens einer Zelle	RR OR ARD	asymptot. KI exaktes KI exaktes KI ^a	Fisher's exakter Test Fisher's exakter Test Z-Test ^c
Häufigkeiten ≥ 1 in beiden Behandlungsgruppen und $> 5\%$ in allen Zellen	RR OR ARD	asymptot. KI asymptot. KI asymptot. KI	Chi ² -Test Chi ² -Test Z-Test ^c
a: Exakte, nicht bedingte KI-Grenzen (siehe EXACT Statement in SAS PROC FREQ). b: + 0,5 addiert zu allen Zelhäufigkeiten. c: Stetigkeits-adjustierter Chi ² -Test (Z-Test). Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [144-146].			

Genauere Informationen zu den durchgeführten und *a priori* definierten Haupt- und Sensitivitätsanalysen sind dem Abschnitt 4.2.5.4 zu entnehmen.

Genauere Informationen zu den angenommenen Verteilungen und verwendeten statistischen Modellen befinden sich in den Studiendokumenten und entsprechenden Kapiteln der statistischen Analysepläne, des *post-hoc* Analyseplans sowie in den dazugehörigen Auswertetabellen (Abschnitt 4.2) [7-10]. Für alle Analysen wurde SAS Version 9.2 oder höher verwendet.

4.2.5.3 Meta-Analysen

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

Für die statistische Auswertung sollen primär die Ergebnisse aus 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 Effekten¹ 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 sowohl anhand des Odds Ratios als auch des Relativen Risikos durchgeführt werden. In begründeten Ausnahmefällen können auch andere Effektmaße zum Einsatz kommen. Bei kategorialen Variablen soll ein geeignetes Effektmaß in Abhängigkeit vom konkreten Endpunkt und den verfügbaren Daten verwendet² 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 Heterogenität³ 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 Abschnitt 4.2.5.4) und klinische Faktoren, sogenannte Effektmodifikatoren (siehe Abschnitt 4.2.5.5).

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

¹ DerSimonian R, Laird N. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7(3):177-188.

² 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.

³ Higgins JPT, Thompson SG, Deeks JJ, Altman DG. Measuring inconsistency in meta-analyses. *BMJ* 2003;327(7414):557-560.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer einzelnen klinischen Studie erfolgt.

4.2.5.4 Sensitivitätsanalysen

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

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

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

Neben Cut-off-Werten und Effektmaßen sind auch die Verwendung von Imputationsmethoden oder Berechnungsmethoden (z. B. für Scores) typische Sensitivitätsanalysen. Insbesondere wenn das Verzerrungspotenzial hoch ist, sollte geprüft werden, ob eine Sensitivitätsanalyse die Aussagekraft der Ergebnisse verbessern kann.

Mithilfe von Sensitivitätsanalysen soll neben der Robustheit der Ergebnisse die Möglichkeit überprüft werden, dass methodische Faktoren die Schlussfolgerungen aus diesen Ergebnissen modifizieren. Sensitivitätsanalysen können sich zum einen auf die Variation metaanalytischer Methodik beziehen (wie z. B. die Verwendung eines Modells mit festen Effekten anstelle eines Modells mit zufälligen Effekten oder die Nicht-Berücksichtigung von Studien mit minderer Qualität). Zum anderen können sie die Robustheit der Resultate der Einzelstudie im Hinblick auf methodische Variationen, wie z. B. das Ersetzen fehlender Werte, untersuchen. Dabei ist man auf die Verfügbarkeit solcher Analysen aus der Einzelstudie angewiesen.

Die Durchführung von Analysen mit unterschiedlichen Populationen zählt zu „andere Effektmodifikatoren“ und stellt keine Sensitivitätsanalyse dar. Diese Analysen werden in Abschnitt 4.2.5.5 beschrieben.

Hauptanalysen zur Bewertung der primären und sekundären Zielkriterien

Kontinuierliche Endpunkte wurden mittels einer Kovarianzanalyse - *Analysis of Covariance* (ANCOVA) - analysiert. Das ANCOVA Modell beinhaltete die Faktoren Behandlung, Region, Gelenkerosionsstatus, sowie den Baseline-Wert des jeweiligen Endpunktes. Dichotome Endpunkte wurden mittels eines logistischen Regressionsmodells analysiert, falls dieses bereits im CSR angewandt wurde. Das logistische Regressionsmodell beinhaltete die Faktoren Behandlung, Region, sowie den Gelenkerosionsstatus. Falls im CSR dichotome Endpunkte nicht mittels eines adjustierten statistischen Modells verglichen wurden, wurden die Behandlungen unadjustiert miteinander verglichen (Tabelle 4-7).

A priori definierte Sensitivitätsanalysen

Es werden im Rahmen dieses Dossiers nur diejenigen Sensitivitätsanalysen beschrieben und dargestellt, die im CSR für den jeweiligen Endpunkt zum Zeitpunkt 52 Wochen präspezifiziert bzw. durchgeführt waren. Woche 52 stellt den Hauptanalysezeitpunkt dar, die früheren Zeitpunkte werden im Rahmen dieses Dossiers als Sensitivitätsanalysen angesehen. Tabelle 4-7 zeigt die dazugehörige Übersicht für jeden Endpunkt.

“Treatment-by-Region” Interaktion

Dem logistischen Regressionsmodell beziehungsweise dem ANCOVA Modell wurde ergänzend zur Hauptanalyse der Interaktionsterm *“Treatment-by-Region”* als Sensitivitätsanalyse hinzugefügt. Diese spezielle Sensitivitätsanalyse ist Bestandteil der Subgruppenanalysen.

“Treatment-by-Baseline” Interaktion

Dem ANCOVA Modell wurde ergänzend zur Hauptanalyse der Interaktionsterm *“Treatment-by-Baseline”* als Sensitivitätsanalyse hinzugefügt. Diese spezielle Sensitivitätsanalyse ist Bestandteil der Subgruppenanalysen.

“Mixed Model for Repeated Measures” Analyse

Die *“Mixed Model for Repeated Measures”* Analyse für kontinuierliche Endpunkte berücksichtigte als fixe kategoriale Effekte die Behandlung, Region, den Gelenkerosionsstatus, die Visite und die *“Treatment-by-Visit”* Interaktion sowie als fixe kontinuierliche Effekte den *Baseline Wert* und die *“Baseline-by-Visit”* Interaktion des entsprechenden Endpunktes.

Imputationsverfahren

Eines der Hauptprobleme fehlender Daten stellt die mögliche Verzerrung der interessierenden Parameter dar, die aus den Daten geschätzt werden sollen. Dies kann dann zu einer Verfälschung und Einschränkung der Aussagekraft von Untersuchungsergebnissen führen. Unterschiedliche statistische Analysemethoden zum Umgang mit fehlenden Daten können als Sensitivitätsanalyse zur Hauptanalyse genutzt werden, um die Robustheit der Ergebnisse zu überprüfen.

In Übereinstimmung mit anderen Phase-III-Studien zur RA [56,85,147-150], wurden folgende Imputationsverfahren zur Vervollständigung fehlender Daten angewandt:

1. “*Non-Responder*” Imputation (NRI) – kategorische Endpunkte

Patienten, die die Studie oder die Therapie zu einem beliebigen Zeitpunkt und aus beliebiger Ursache abbrechen, wurden ab dem Zeitpunkt des Abbruchs zur Auswertung kategorischer Endpunkte als “*Non-Responder*” bewertet. Patienten, die eine *Rescue*-Therapie erhielten, wurden ab diesem Zeitpunkt zur Auswertung kategorischer Endpunkte als “*Non Responder*” bewertet. Randomisierte Patienten, für die keine Daten zu einem bestimmten Erhebungszeitpunkt nach der Erhebung zu Baseline zur Verfügung stehen, wurden für diesen Zeitpunkt zur Auswertung kategorischer Endpunkte als “*Non-Responder*” bewertet.

2. “*Modified Last Observation Carried Forward*” (mLOCF) – kontinuierliche Endpunkte

Generall wurde für die Analyse aller kontinuierlichen Endpunkte die “*Modified Last Observation Carried Forward*” (mLOCF) Methode angewandt (sofern nicht explizit anders erwähnt). Für Patienten, die eine *Rescue*-Therapie erhielten, wurden die letzten erhobenen Daten vor oder zum Zeitpunkt der *Rescue*-Therapie zur Vervollständigung fehlender Werte und Analyse des Endpunktes verwendet. Für Patienten, die die Studie oder die Therapie zu einem beliebigen Zeitpunkt und aus beliebiger Ursache abbrechen, wurden die letzten erhobenen Daten vor Abbruch zur Vervollständigung fehlender Werte und Analyse des Endpunktes verwendet.

3. “*Modified Baseline Observation Carried Forward*” (mBOCF) – kontinuierliche Endpunkte

Weiterhin wurde zur Analyse kontinuierlicher Endpunkte (sofern nicht explizit anders erwähnt) die “*Modified Baseline Observation Carried Forward*” (mBOCF) Methode angewandt. Für Patienten, die die Studie oder die Therapie aufgrund eines unerwünschten Ereignisses, einschließlich Todesfälle abbrechen, wurden Daten der Erhebung zu Baseline zur Vervollständigung fehlender Werte und Analyse des Endpunktes verwendet. Für Patienten, die die Studie oder die Therapie wegen anderer Gründe (nicht aufgrund unerwünschter Ereignisse) abbrechen, wurden die letzten erhobenen Daten vor Abbruch zur Vervollständigung fehlender Werte und Analyse des Endpunktes verwendet. Für Patienten, die eine *Rescue*-Therapie erhielten, wurden die letzten erhobenen Daten vor oder zum Zeitpunkt der *Rescue*-Therapie zur Vervollständigung fehlender Werte und Analyse des Endpunktes verwendet. Für Patienten mit fehlendem Baseline-Wert oder keiner Veränderung nach Baseline (auch nachdem Werte nach den oben beschriebenen Bedingungen ersetzt wurden) wurde die Veränderung auf Null gesetzt.

Weitere Informationen befinden sich in den Studiendokumenten und entsprechenden Kapiteln der statistischen Analysepläne.

Tabelle 4-7: Übersicht statistischer Analysemethoden

RA-BEAM (JADV)	Statistische Analysemethoden	
	Hauptanalyse und Imputationsmethode	Sensitivitätsanalyse bzw. Imputationsmethode
Morbidität		
Veränderung der Anzahl der druckschmerzhaften Gelenke und geschwollenen Gelenke	ANCOVA (mLOCF)	-
Veränderung des DAS28-hsCRP	ANCOVA (mBOCF)	MMRM ANCOVA (mLOCF)
Anteil Patienten mit DAS28-hsCRP $\leq 3,2$	logistisches Regressionsmodell (NRI)	-
Anteil Patienten mit DAS28-hsCRP $< 2,6$	logistisches Regressionsmodell (NRI)	-
Veränderung des SDAI	ANCOVA (mLOCF)	-
Anteil Patienten mit SDAI $\leq 3,3$	logistisches Regressionsmodell (NRI)	-
Veränderung des CDAI	ANCOVA (mLOCF)	-
Anteil Patienten mit CDAI ≤ 10	logistisches Regressionsmodell (NRI)	-
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	logistisches Regressionsmodell (NRI)	-
Veränderung des PAP (VAS)	ANCOVA (mLOCF)	-
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	logistisches Regressionsmodell (NRI)	-
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	logistisches Regressionsmodell (NRI)	-
Veränderung des PtGADA (VAS)	ANCOVA (mLOCF)	-
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	logistisches Regressionsmodell (NRI)	-
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	logistisches Regressionsmodell (NRI)	-
Veränderung des HAQ-DI	ANCOVA (mBOCF)	MMRM ANCOVA (mLOCF)
Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$	logistisches Regressionsmodell (NRI)	-
Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$	logistisches Regressionsmodell (NRI)	-
Veränderung der Dauer der Morgensteifigkeit (min)	ANCOVA (mLOCF)	nichtparametrisches Modell (mLOCF)

RA-BEAM (JADV)	Statistische Analysemethoden	
Endpunkt	Hauptanalyse und Imputationsmethode	Sensitivitätsanalyse bzw. Imputationsmethode
Veränderung des EQ-5D VAS	ANCOVA (mLOCF)	-
Anteil Patienten mit einer Verbesserung des EQ-5D VAS ≥ 10 mm	logistisches Regressionsmodell (NRI)	-
Veränderung der schwersten Müdigkeit	ANCOVA (mLOCF)	-
Lebensqualität		
Veränderung des SF-36 v2 akut PCS	ANCOVA (mLOCF)	-
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5	logistisches Regressionsmodell (NRI)	-
Veränderung des SF-36 v2 akut MCS	ANCOVA (mLOCF)	-
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5	logistisches Regressionsmodell (NRI)	-
Veränderung der SF-36 Domänen	ANCOVA (mLOCF)	-
Sicherheit		
Todesfälle	unadjustierte Analysen	-
SUE	unadjustierte Analysen	-
UE	unadjustierte Analysen	-
Schwere UE (ohne ausschließliche Laborereignisse)	unadjustierte Analysen	-
Studienabbruch aufgrund von UE	unadjustierte Analysen	-
Therapieabbruch aufgrund von UE	unadjustierte Analysen	-
UE von besonderem Interesse (kategorische Endpunkte)	unadjustierte Analysen	-
UE von besonderem Interesse (kontinuierliche Endpunkte)	ANCOVA (mLOCF)	-
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].		

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 sollen, soweit sinnvoll, folgende Faktoren bezüglich einer möglichen Effektmodifikation berücksichtigt werden:

- Geschlecht
- Alter
- Krankheitsschwere bzw. –stadium
- Zentrums- und Ländereffekte

Sollten sich aus den verfügbaren Informationen Anzeichen 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.

Subgruppenanalysen

Nach den Anforderungen der Verfo wurden Interaktionstests für:

- Geschlecht (weiblich, männlich)
- Alter (< 65, ≥ 65 Jahre)
- Krankheitsschwere bzw. –stadium
- Zentrums- und Ländereffekte

sowie alle *a priori* geplanten Subgruppen durchgeführt, um zu überprüfen, ob der Therapieeffekt über die Subgruppen hinweg konstant ist oder ob es Hinweise für eine Effektmodifikation gibt (Abschnitt 4.3.1.3.4).

Subgruppenanalysen wurden nur für den Zeitpunkt 52 Wochen und für die Hauptanalysen durchgeführt.

Methodik

Kontinuierliche Endpunkte wurden mittels einer Kovarianzanalyse - ANCOVA - analysiert. Das ANCOVA Modell beinhaltet die Faktoren Behandlung, Subgruppe und die Interaktion „*Subgroup-by-Treatment*“.

Dichotome Endpunkte wurden mittels eines logistischen Regressionsmodells analysiert, mit den Faktoren Behandlung, Subgruppe und die Interaktion „*Subgroup-by-Treatment*“. Für dichotome Endpunkte bezieht sich die Subgruppenanalyse immer auf das RR.

Fälle, in denen keine Subgruppenanalysen durchgeführt wurden:

Falls die kleinste Subgruppenkategorie über die Behandlungsgruppen hinweg weniger als 5% der Gesamtpopulation ausmachte, wurde diese Kategorie mit einer geeigneten anderen Kategorie zusammengefasst. Falls dies nicht möglich war oder es nur insgesamt zwei Kategorien gab, wurden für dieses Subgruppenmerkmal keine Subgruppenanalysen durchgeführt.

Für folgende, präspezifizierte Subgruppenmerkmale wurde daher aus den oben genannten Gründen keine Subgruppenanalyse durchgeführt:

- Alter (< 75 Jahre, ≥ 75 Jahre):
Fallzahl zu gering in der Kategorie ≥ 75 Jahre (< 5%)
- Alter (< 85 Jahre, ≥ 85 Jahre):
Fallzahl zu gering in der Kategorie ≥ 85 Jahre (< 5%)
- Alter (< 65 Jahre, 65 bis < 75 Jahre, 75 bis < 85 Jahre, ≥ 85 Jahre):
Fallzahl zu gering in den beiden obersten Kategorien „75 bis < 85 Jahre“ und „≥85 Jahre“ (< 5%)

- Nierenfunktion (beeinträchtigt (*Estimated Glomerular Filtration Rate* (eGFR) < 60 ml/min/1,73 m²) oder nicht beeinträchtigt (eGFR ≥ 60 ml/min/1,73 m²): Fallzahl zu gering in der Kategorie „beeinträchtigt (eGFR < 60 ml/min/1,73 m²)“ (< 5%)

Für dichotome Endpunkte wurden keine Subgruppenanalysen durchgeführt, wenn in der kleinsten Subgruppenkategorie (ggf. nach Zusammenlegung von Subgruppenkategorien) weniger als 10 Ereignisse auftraten. Weitere Details hierzu können im Post-hoc Analyseplan nachgelesen werden.

Zusätzlich wurden für folgende präspezifizierte Subgruppenmerkmale aus anderen Gründen keine Subgruppenanalysen durchgeführt:

- Gewicht zu Baseline (≤ Median, > Median):
Der Median basiert auf der Studienpopulation und ist deswegen im Kontext dieser Nutzenbewertung nicht unbedingt sinnvoll interpretierbar. Zusätzlich existieren andere Kategorisierungen, die auf aktuellen Gewichtsklassen basieren.
- *Body Mass Index* (BMI) zu Baseline (≤ Median, > Median):
Der Median basiert auf der Studienpopulation und ist deswegen im Kontext dieser Nutzenbewertung nicht unbedingt sinnvoll interpretierbar. Zusätzlich existieren andere Kategorisierungen, die auf aktuellen BMI-Klassen basieren.
- Ethnie (Asiatisch, Afroamerikanisch, Kaukasisch, Andere):
Dieses spezielle Subgruppenmerkmal wurde nicht analysiert, auch weil in den meisten Kategorien die Patientenzahlen zu gering waren (<5%). Stattdessen wurden die Kategorien Kaukasisch bzw. Andere gebildet, um den deutschen Versorgungskontext besser abbilden zu können.
- Region A (USA, Andere):
Dieses Subgruppenmerkmal ist für den Kontext dieser Nutzenbewertung nicht relevant.
- Region B (Japan, Andere):
Dieses Subgruppenmerkmal ist für den Kontext dieser Nutzenbewertung nicht relevant.

Für folgende Endpunkte wurden wegen insgesamt zu geringer Ereigniszahlen keinerlei Subgruppenanalysen durchgeführt:

- Mortalität (Die Anzahl der Todesfälle wurde im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.)
- Schwere UE (ohne ausschließliche Laborereignisse)
Für diesen Endpunkt wurde für Subpopulation A3 keine Subgruppenanalysen durchgeführt.
- Schwere Infektionen
- Kardiovaskuläre Ereignisse
- Malignität

- Allergische Reaktionen/Hypersensitivität
- Gastrointestinale Perforation.

Es wurden alle Endpunkte (soweit möglich) für folgende Subgruppen analysiert und dargestellt:

- Geschlecht (männlich, weiblich)
- Alter (< 65 Jahre, ≥ 65 Jahre)
- Gewicht zu Baseline (< 60 kg, ≥ 60 bis < 100 kg, ≥ 100 kg)
Subpopulationen A2 und A3: Aufgrund geringer Fallzahlen wurden die beiden oberen Kategorien zusammengefasst (< 60 kg, ≥ 60 kg).
- BMI zu Baseline (< 25 kg/m², ≥ 25 bis < 30 kg/m², ≥ 30 kg/m²)
Alle Populationen: Die präspezifizierten Kategorien < 18 kg/m² und ≥ 18 bis < 25 kg/m² wurden aufgrund geringer Fallzahlen zusammengefasst (< 25 kg/m²).
- Zusätzliches Subgruppenmerkmal: Ethnie (Kaukasisch, Andere)
- Region (US und Kanada, Zentral- und Südamerika und Mexiko, Ost-Europa, West-Europa, Asien ohne Japan, Japan, restliche Welt)
Subpopulationen A2 und A3: Aufgrund geringer Fallzahlen wurden die folgenden Kategorien gebildet: US/Kanada/restliche Welt, Zentral- und Südamerika/Mexiko, Europa, Asien.
- Spezifische Region I (Europa, Andere)
- Zusätzliches Subgruppenmerkmal: Spezifische Region II (Westeuropa, Andere)
Subpopulation A3: Aufgrund zu geringer Fallzahlen wurde die Subgruppenanalyse nicht durchgeführt (< 5% in der Kategorie Westeuropa).
- Erkrankungsdauer (< 1 Jahre, ≥ 1 bis < 5 Jahre, ≥ 5 bis < 10 Jahre, ≥ 10 Jahre)
- Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP > 5,1)
- Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität, mindestens 3 Gelenkerosionen)
- Seropositivität und Gelenkerosionsstatus (Seropositiv und mindestens 3 Gelenkerosionen, andere)
- Seropositivität (Rheumafaktor (RF) positiv und/oder oder *Anti Citrullinated Peptide/Protein Antibodies* (ACPA) positiv, RF und ACPA negativ)
Subpopulation A3: Subgruppenanalysen wurden nicht durchgeführt aufgrund zu geringer Fallzahl (< 5% in der Kategorie „RF und ACPA negativ“).
- Seropositivität (RF positiv, RF negativ)
- Seropositivität (ACPA positiv, ACPA negativ)
Die Kategorie „ACPA negativ“ enthält sowohl ACPA-negative Patienten als auch Patienten, deren ACPA-Status unklar war.
- Art der Begleitmedikation (nur MTX; MTX und andere cDMARDs)
Subpopulationen A2 und A3: Da alle Patienten in diesen Subpopulationen nur MTX erhielten, wurden hier keine Subgruppenanalysen durchgeführt.

- *Modified Total Sharp Score* (mTSS) zu Baseline (\leq Median, $>$ Median)
Da für diesen Endpunkt nur dieses eine Subgruppenmerkmal präspezifiziert war, wurden die dazugehörigen Subgruppenanalysen durchgeführt, obwohl der Median auf der Studienpopulation basiert und deswegen im Kontext dieser Nutzenbewertung nicht unbedingt sinnvoll interpretierbar ist.
- Therapie mit Kortikosteroiden zu Baseline (ja, nein)

Für die Interpretation der Ergebnisse der Subgruppenanalysen sollte Folgendes berücksichtigt werden:

Zur Durchführung einer Metaanalyse über verschiedene Studien hinweg gibt der G-BA Subgruppenanalysen für die einzelnen Studien vor. Grenzwerte für den Interaktions-p-Wert sind laut IQWiG-Methodenpapier 4.2 nur in Zusammenhang mit Metaanalysen erwähnt. Hier wird der Grenzwert von $< 0,2$ als Hinweis auf eine Effektmodifikation verstanden. Das Methodenpapier des IQWiG beschreibt in Bezug auf einzelne Studien, dass

„... im Allgemeinen post-hoc durchgeführte Subgruppenanalysen auf Studienebene kritisch zu interpretieren sind...“ [6].

Im vorliegenden Fall handelt es sich nur um eine Studie, sodass keine Metaanalyse durchgeführt werden kann. In diesem Kontext gibt es daher verschiedene Limitierungen für die Interpretation der Ergebnisse der Subgruppenanalysen auf Ebene dieser einzelnen Studie.

In Anlehnung an den Entwurf des IQWiG-Methodenpapier 5.0 [151], der ausschließlich einen Grenzwert von 0,05 als Beleg für eine Effektmodifikation berücksichtigt, werden im vorliegenden Dossier lediglich die Details von Subgruppenanalysen mit einem Interaktions-p-Wert von $< 0,05$ dargestellt.

Bei der Beurteilung der Ergebnisse von Subgruppenanalysen sollte ferner die, mit multiplem Testen in Zusammenhang stehende α -Fehler-Kumulierung, die die Zunahme der Wahrscheinlichkeit für Fehler 1. Art (falsch positive Testergebnisse) beschreibt, berücksichtigt werden. Auch sollten bei der Interpretation der Resultate kleine Fallzahlen innerhalb der Subgruppen berücksichtigt werden.

Andere Effektmodifikatoren

Heterogenitätsanalysen

Gemäß G-BA-Beratungsgespräch vom 10.12.2015 [2] sind im vorliegenden Anwendungsgebiet verschiedene Therapiesituationen zu unterscheiden, die in:

- Therapie-naive Patienten ohne vorherige Behandlung
- Patienten mit vorangegangenen Behandlungen mit einem oder mehreren **cDMARDs**, einschließlich Methotrexat, oder (Anwendungsgebiet A)
- Patienten mit vorangegangenen Behandlungen mit einem oder mehreren **bdMARDs** (Anwendungsgebiet B)

zu differenzieren sind.

Die Zielpopulation wurde aufgrund der verschiedenen, durch den G-BA festgelegten Therapiesituationen in zwei Anwendungsgebiete unterschieden, die nachfolgend als Anwendungsgebiet A (Modul 4 A) und Anwendungsgebiet B (Modul 4 B) bezeichnet werden (siehe Modul 2/3; Abschnitt 3.1).

Die Bewertung des medizinischen Nutzens und des medizinischen Zusatznutzens von Baricitinib als Monotherapie oder Kombinationstherapie zur Behandlung der mittelschweren bis schweren aktiven rheumatoiden Arthritis bei erwachsenen Patienten, die auf eine vorangegangene Behandlung mit einem oder mehreren DMARDs unzureichend angesprochen oder diese nicht vertragen haben, erfolgt vergleichend für:

Die Zielpopulation im Anwendungsgebiet A wurde vom G-BA [2] in folgende Subpopulationen unterteilt:

Subpopulation A1

Patienten, bei denen **keine ungünstigen Prognosefaktoren**¹ vorliegen und die unzureichend auf eine vorangegangene Behandlung mit **einem** cDMARD (inklusive MTX) ansprachen.

Subpopulation A2

Patienten **mit ungünstigen Prognosefaktoren**¹, die unzureichend auf eine vorangegangene Behandlung mit **einem** cDMARD (inklusive MTX) ansprachen.

¹ Beispiele für ungünstige Prognosefaktoren:

- Nachweis von Autoantikörpern (z. B. Rheumafaktoren, hohe Spiegel von Antikörpern gegen citrullinierte Peptid Antigene)
- Hohe Krankheitsaktivität (nachgewiesen durch DAS- bzw. DAS28-Bewertungssystem, geschwollene Gelenke, Parameter der Akute-Phase-Reaktion wie z. B. C-Reaktives Protein, Erythrozytensedimentationsrate)
- Frühes Auftreten von Gelenkerosionen.

Quelle: [2].

Subpopulation A3

Patienten, die unzureichend auf eine vorangegangene Behandlung mit **mehreren** cDMARDs (inklusive MTX) ansprachen.

Bei der Bewertung des Zusatznutzens gemäß Fragestellung und gemäß G-BA-Beratungsgespräch vom 10.12.2015 [2] müssen daher im vorliegenden Anwendungsgebiet *post-hoc* Analysen der definierten Subpopulationen durchgeführt werden, die von der ursprünglichen Studienpopulation (mITT) abweichen. Aus diesem Grund wurde eine mögliche Heterogenität bezüglich des Effekts in den verschiedenen Populationen untersucht. Ist eine solche Heterogenität nicht vorhanden, so ist der über alle Patienten hinweg geschätzte Gesamteffekt sinnvoll interpretierbar und auch auf die verschiedenen Subpopulationen übertragbar [152-153].

Methodik

Die Untersuchung der möglichen Heterogenität zwischen den Ergebnissen beider Subpopulationen geschah mittels Interaktionstests.

Für Haupt- bzw. Sensitivitätsanalysen, für die bereits adjustierte Modelle verwendet wurden, wurden die folgenden Faktoren zu den jeweiligen Modellen hinzugefügt:

- ANCOVA bzw. logistische Regression: Es wurden die Faktoren „Subpopulation“ und die Interaktion „*Subpopulation-by-Treatment*“ hinzugefügt.
- MMRM: Es wurden die Faktoren „Subpopulation“, sowie die Interaktionen „*Subpopulation-by-Treatment*“ und „*Subpopulation*Treatment*Visit*“.

Für dichotome Endpunkte, die in den Hauptanalysen unadjustiert analysiert wurden, wurde die Heterogenität wie folgt evaluiert: Logistische Regression mit den Faktoren Behandlung, Subpopulation und der Interaktion „*Subpopulation-by-Treatment*“.

Die Beurteilung der Heterogenität basierte auf dem Interaktions p-Wert „*Subpopulation-by-Treatment*“. Für dichotome Endpunkte bezieht sich die Heterogenitätsanalyse immer auf das RR. Im Rahmen dieses Dossiers wurden bei Vorliegen eines p-Werts von $< 0,05$ die Ergebnisse näher diskutiert.

In den folgenden Fällen wurden keine Heterogenitätstests für dichotome Endpunkte durchgeführt:

- Keine Patienten mit Ereignissen in mindestens einer der beiden Subpopulationen
- Die Anzahl der Patienten mit Ereignissen war zu gering, d. h. $< 1\%$ in mindestens einer der beiden Behandlungsgruppen (mITT).

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 möglich¹. 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“², „Multiple-Treatment-Meta-Analysen“³ oder auch „Netzwerk-Meta-Analysen“⁴, 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 ist⁵.

Da das Ergebnis eines indirekten Vergleichs maßgeblich von der Auswahl des Brückenkomparators bzw. der Brückenkomparatoren abhängen kann, ist die Wahl des Brückenkomparators bzw. der Brückenkomparatoren zu begründen. Dies gilt insbesondere dann, wenn eine Beschränkung auf ein oder mehrere Brückenkomparatoren vorgenommen wird, obwohl Daten zu anderen Therapieoptionen, die ebenfalls als Brückenkomparatoren in Frage kommen, vorliegen. Insgesamt ist es notwendig, die zugrunde liegende Methodik genau und reproduzierbar zu beschreiben und die Annahme der Konsistenz zu untersuchen⁶.

¹ 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.

² Lu G, Ades AE. Combination of direct and indirect evidence in mixed treatment comparisons. Stat Med 2004; 23(20): 3105-3124.

³ Caldwell DM, Ades AE, Higgins JP. Simultaneous comparison of multiple treatments: combining direct and indirect evidence. BMJ 2005; 331(7521): 897-900.

⁴ Salanti G, Higgins JPT, Ades AE, Ioannidis JPA. Evaluation of networks of randomized trials. Stat Methods Med Res 2008; 17(3): 279-301.

⁵ B. Schöttker, D. Lühmann, D. Boukhemair, and H. Raspe. Indirekte Vergleiche von Therapieverfahren. Schriftenreihe Health Technology Assessment Band 88, DIMDI, Köln, 2009.

⁶ 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: A1147.

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

- *Benennung des Brückenkomparators bzw. der Brückenkomparatoren und Begründung für die Auswahl.*
- *Genauere 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.*

Für die vorliegende Fragestellung wurde eine direkte Vergleichsstudie der Evidenzstufe Ib identifiziert. Die Bewertung wurde auf Grundlage dieser Evidenzgüte durchgeführt und auf indirekte Vergleiche verzichtet.

4.3 Ergebnisse zum medizinischen Nutzen und zum medizinischen Zusatznutzen

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

Abschnitt 4.3.2 enthält weitere Unterlagen anderer Evidenzstufen, sofern diese aus Sicht des pharmazeutischen Unternehmers zum Nachweis des Zusatznutzens erforderlich sind. Diese Unterlagen teilen sich wie folgt auf:

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

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

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

- *Es sollen alle RCT, die der Zulassungsbehörde im Zulassungsdossier übermittelt wurden und deren Studienberichte im Abschnitt 5.3.5 des Zulassungsdossiers enthalten sind, aufgeführt werden. Darüber hinaus sollen alle RCT, für die der pharmazeutische Unternehmer Sponsor ist oder war oder auf andere Weise finanziell beteiligt ist oder war, aufgeführt werden.*
- *Benennen Sie in der nachfolgenden Tabelle nur solche RCT, die ganz oder teilweise innerhalb des in diesem Dokument beschriebenen Anwendungsgebiets durchgeführt wurden. Fügen Sie dabei für jede Studie eine neue Zeile ein.*

Folgende Informationen sind in der Tabelle 4-8 darzulegen: Studienbezeichnung, Angabe „Zulassungsstudie ja/nein“, Angabe über die Beteiligung (Sponsor ja/nein), Studienstatus (abgeschlossen, abgebrochen, laufend), Studiendauer und Therapiearme. Orientieren Sie sich dabei an der beispielhaften Angabe in der ersten Tabellenzeile.

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

Studie	Zulassungsstudie (ja/nein)	Sponsor (ja/nein)	Status (abgeschlossen/ abgebrochen/ laufend)	Studien- dauer	Therapiearme
I4V-MC-JADA	ja	ja	abgeschlossen	128 Wochen ^a	<u>Teil 1</u> a) Baricitinib 1, 2, 4, 8 mg b) Placebo <u>Teil 2</u> Baricitinib 2, 4, 8 mg <u>Teil 3</u> Baricitinib 4, 8 mg <u>Teil 4</u> Baricitinib 4 mg
I4V-MC-JADC	ja	ja	abgeschlossen	24 Wochen ^a	a) Baricitinib 4 mg b) Baricitinib 7 mg c) Baricitinib 10 mg d) Placebo mit Wechsel auf Baricitinib 7 mg e) Placebo mit Wechsel auf Baricitinib 10 mg
I4V-JE-JADN	ja	ja	abgeschlossen	64 Wochen ^a	<u>Teil 1</u> a) Baricitinib 1, 2, 4, 8 mg b) Placebo <u>Teil 2</u> Baricitinib 4, 8 mg
I4V-MC-JADV (RA-BEAM)	ja	ja	abgeschlossen	52 Wochen ^a	a) Baricitinib 4 mg b) Adalimumab 40 mg c) Placebo
I4V-MC-JADW (RA-BEACON)	ja	ja	abgeschlossen	24 Wochen ^a	a) Baricitinib 2, 4 mg b) Placebo
I4V-MC-JADX (RA-BUILD)	ja	ja	abgeschlossen	24 Wochen ^a	a) Baricitinib 2, 4 mg b) Placebo
I4V-MC-JADZ (RA-BEGIN)	ja	ja	abgeschlossen	52 Wochen ^a	a) Baricitinib 4 mg b) MTX 10-20 mg c) Placebo
I4V-MC-JADY (RA-BEYOND)	ja	ja	laufend	4 Jahre ^a	a) Baricitinib 2, 4 mg
I4V-CR-JAGS (RA-BALANCE)	ja	ja	laufend	52 Wochen ^a	a) Baricitinib 4 mg b) Placebo
a: Behandlungsdauer. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [154].					

Geben Sie an, welchen Stand die Information in Tabelle 4-8 hat, d. h. zu welchem Datum der Studienstatus abgebildet wird. Das Datum des Studienstatus soll nicht mehr als 3 Monate vor dem für die Einreichung des Dokuments maßgeblichen Zeitpunkt liegen.

Der Studienstatus in Tabelle 4-8 wird zum Zeitpunkt 02.02.2017 abgebildet.

Geben Sie in der nachfolgenden Tabelle an, welche der in Tabelle 4-9 genannten Studien nicht für die Nutzenbewertung herangezogen wurden. Begründen Sie dabei jeweils die Nichtberücksichtigung. Fügen Sie für jede Studie eine neue Zeile ein.

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

Studienbezeichnung	Begründung für die Nichtberücksichtigung der Studie
I4V-MC-JADA	Vergleichstherapie Placebo
I4V-MC-JADC	Vergleichstherapie Placebo
I4V-JE-JADN	Vergleichstherapie Placebo
I4V-MC-JADZ (RA-BEGIN)	Patientenpopulation
I4V-MC-JADW (RA-BEACON)	Vergleichstherapie Placebo
I4V-MC-JADX (RA-BUILD)	Vergleichstherapie Placebo
I4V-MC-JADY (RA-BEYOND)	Vergleichstherapie Placebo
I4V-CR-JAGS (RA-BALANCE)	Vergleichstherapie Placebo
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [154].	

4.3.1.1.2 Studien aus der bibliografischen Literaturrecherche

Beschreiben Sie nachfolgend das Ergebnis der bibliografischen Literaturrecherche. 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 Dokumente unter Nennung des Ausschlussgrunds in Anhang 4-C.

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

Geben Sie im Flussdiagramm auch das Datum der Recherche an. Die Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

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

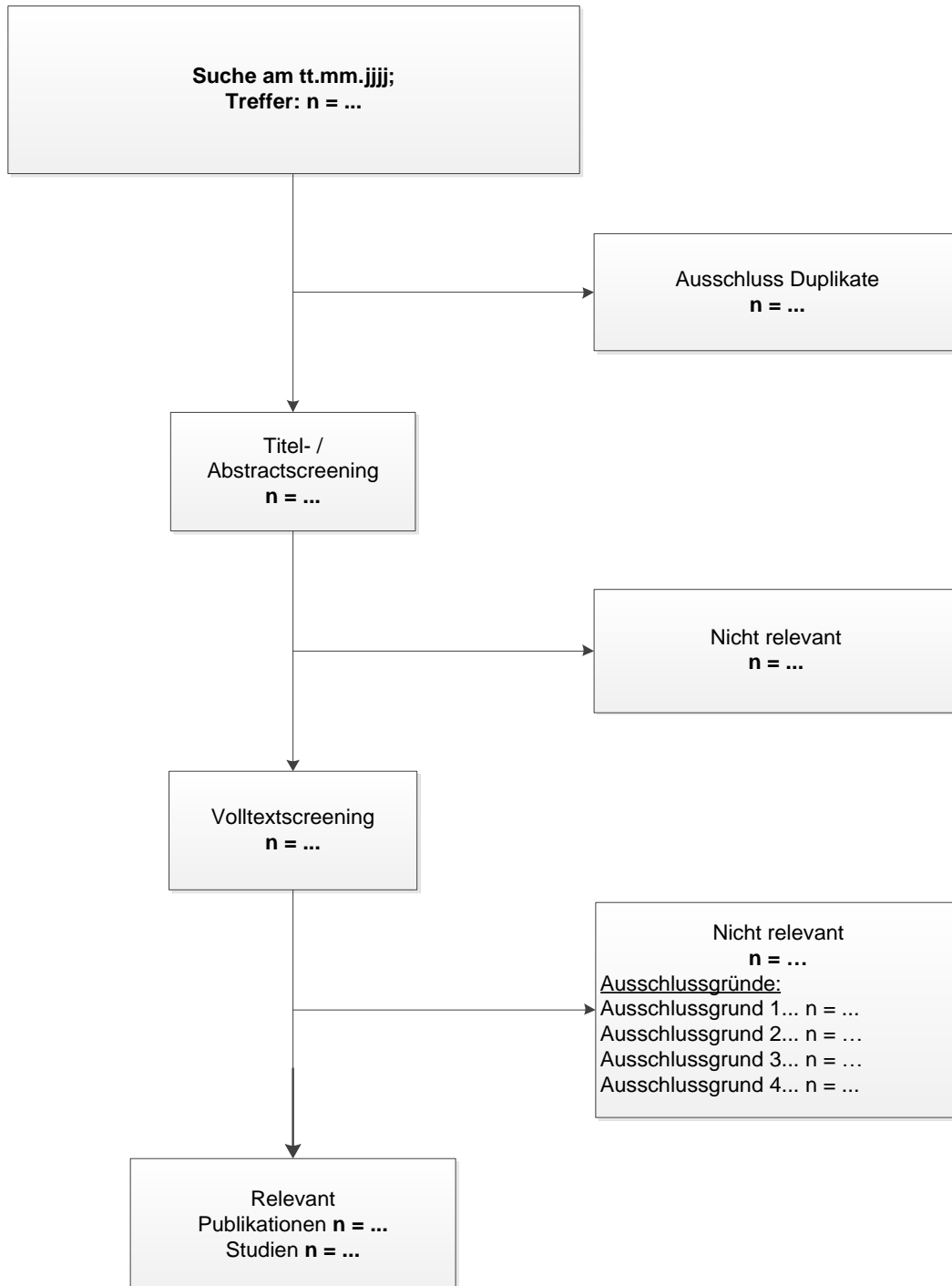


Abbildung 1: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel

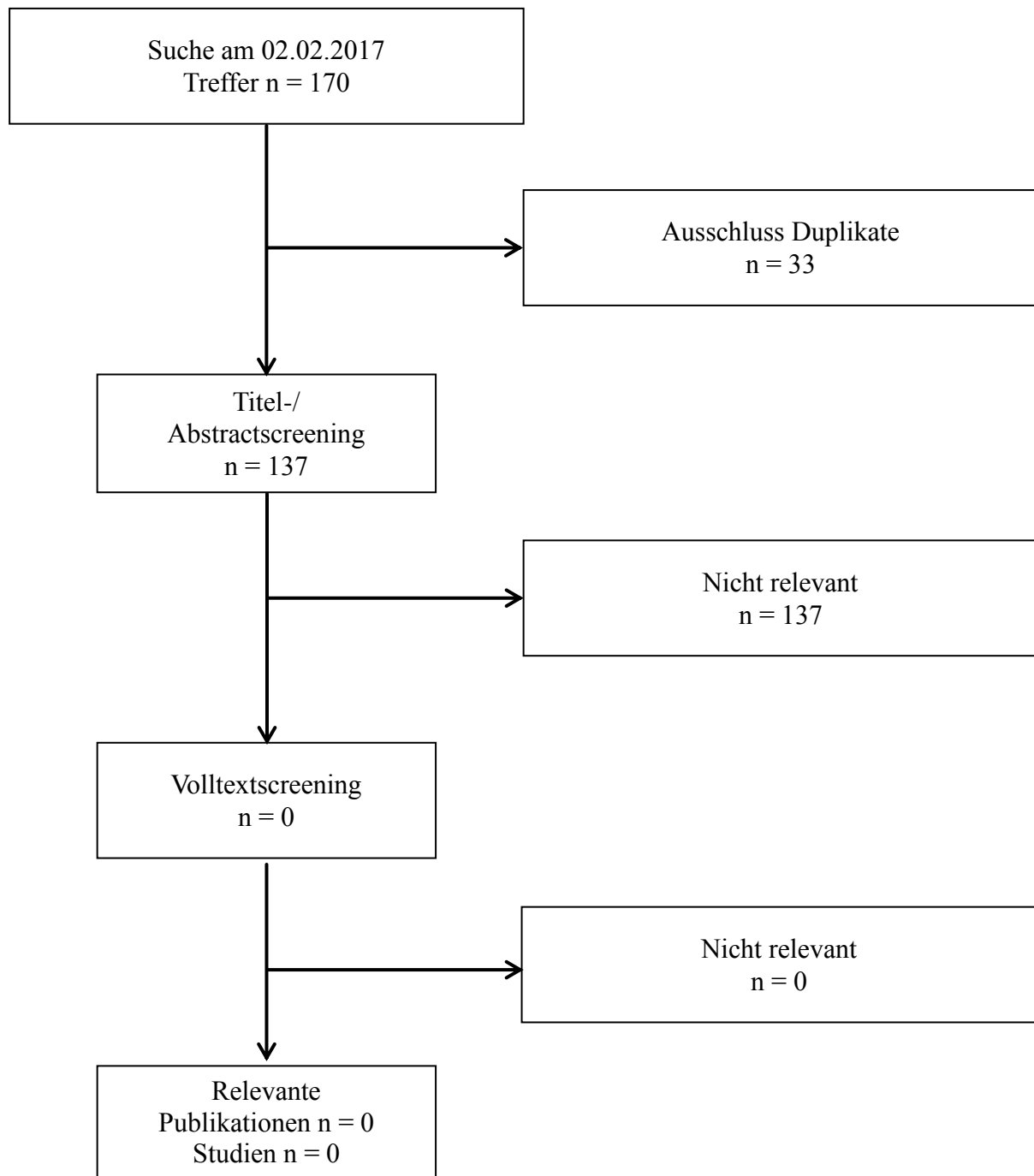


Abbildung 2: Flussdiagramm der bibliografischen Literaturrecherche – Suche nach randomisierten kontrollierten Studien mit dem zu bewertenden Arzneimittel (für die Subpopulationen A1, A2 und A3)

Es wurde eine Suchstrategie für die bibliografische Literaturrecherche nach RCT mit dem zu bewertenden Arzneimittel für die Subpopulationen A1, A2 sowie A3 angewendet und die Selektion gemäß der in Tabelle 4-3 definierten Ein- und Ausschlusskriterien vorgenommen. Die Ergebnisse der Selektionen waren für die Subpopulation A1, A2 und A3 gleich, deshalb wird an dieser Stelle auf eine Mehrfachdarstellung des Flussdiagramms der bibliographischen Literaturrecherche für die verschiedenen Populationen verzichtet.

Zur Identifizierung von relevanten Studien wurden neben internen Informationsquellen und Datenbanken der *Eli Lilly and Company* eine systematische Literaturrecherche sowie eine Studienregistersuche zur Informationsbeschaffung herangezogen.

Die systematische Literaturrecherche wurde in den Datenbanken Cochrane, MEDLINE und EMBASE über die Suchoberfläche auf der Internetseite von Ovid durchgeführt (Suchzeitpunkt: 02.02.2017).

Die bibliografische Literaturrecherche fand am 02.02.2017 statt und erzielte insgesamt 170 Treffer. Nach Ausschluss der Duplikate (n = 33) wurden Titel und Abstract der verbleibenden 137 Publikationen unter Berücksichtigung der Einschlusskriterien (Abschnitt 4.2.2) gesichtet.

Es wurden alle identifizierten Publikationen/Studien als nicht relevant eingestuft und somit wurde keine Publikation/Studie in die Bewertung eingeschlossen werden.

4.3.1.1.3 Studien aus der Suche in Studienregistern

Beschreiben Sie in der nachfolgenden Tabelle 4-10 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 4-8) und ob die Studie auch durch die bibliografische Literaturrecherche identifiziert wurde. Fügen Sie für jede Studie eine neue Zeile ein. Listen Sie die ausgeschlossenen Studien unter Nennung des Ausschlussgrunds in Anhang 4-D.

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

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

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

Studie	Identifikationsorte (Name des Studienregisters und Angabe der Zitate ^a)	Studie in Liste der Studien des pharmazeutischen Unternehmers enthalten (ja/nein)	Studie durch bibliografische Literaturrecherche identifiziert (ja/nein)	Status ^b (abgeschlossen/abgebrochen/laufend)
Subpopulation A1				
Es wurde keine relevante Studie identifiziert.				
Subpopulation A2/Subpopulation A3				
RA-BEAM (JADV)	<u>ClinicalTrials.gov</u> NCT01710358 [155] <u>ICTRP</u> NCT01710358 [156] <u>EU-CTR</u> 2012-002322-73 [157] <u>PharmNet.Bund</u> 2012-002322-73 [158]	ja	nein	abgeschlossen
<p>a: Zitat des Studienregistereintrags sowie, falls vorhanden, der im Studienregister aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.</p> <p>b: Information, ob eine es sich um eine abgeschlossene, abgebrochene oder laufende Studie handelt.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [155-158].</p>				

Geben Sie an, welchen Stand die Information in Tabelle 4-10 hat, d. h. zu welchem Datum die Recherche durchgeführt wurde. Das Datum der Recherche soll nicht mehr als 3 Monate vor dem für die Einreichung des Dossiers maßgeblichen Zeitpunkt liegen.

Der Studienstatus in Tabelle 4-10 wird zum Zeitpunkt 02.02.2017 abgebildet.

4.3.1.1.4 Resultierender Studienpool: RCT mit dem zu bewertenden Arzneimittel

Benennen Sie in der nachfolgenden Tabelle 4-11 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 (exklusive laufender Studien) für das zu bewertende Arzneimittel, auch im direkten Vergleich zur zweckmäßigen Vergleichstherapie. Führen Sie außerdem alle relevanten Studien einschließlich der verfügbaren Quellen in Abschnitt 4.6 auf. Alle durch die vorhergehenden Schritte identifizierten und in der Tabelle genannten Quellen der relevanten Studien sollen für die Bewertung dieser Studien herangezogen werden.

Folgende Informationen sind in der Tabelle 4-11 darzulegen: Studienbezeichnung, Studienkategorie und verfügbare Quellen. 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 4-11: Studienpool – RCT mit dem zu bewertenden Arzneimittel

Studie	Studienkategorie			verfügbare Quellen ^a		
	Studie zur Zulassung des zu bewertenden Arzneimittels (ja/nein)	gesponserte Studie ^b (ja/nein)	Studie Dritter (ja/nein)	Studienbericht (ja/nein (Zitat))	Registereintrag ^c (ja/nein (Zitat))	Publikation (ja/nein (Zitat))
Placebo-kontrolliert						
Nicht zutreffend.						
aktiv-kontrolliert, zweckmäßige Vergleichstherapie(n)						
Subpopulation A1						
Es wurde keine relevante Studie identifiziert.						
Subpopulation A2/Subpopulation A3						
RA-BEAM^d (JADV)	ja	ja	nein	ja [9]	ClinicalTrials.gov NCT01710358 [155] ICTRP NCT01710358 [156] EU-CTR 2012-002322-73 [157] PharmNet.Bund 2012-002322-73 [158]	ja [159]
<p>a: Bei Angabe „ja“ sind jeweils die Zitate der Quelle(n) (z. B. Publikationen, Studienberichte, Studienregistereinträge) mit anzugeben, und zwar als Verweis auf die in Abschnitt 4.7 genannte Referenzliste. Darüber hinaus ist darauf zu achten, dass alle Quellen, auf die in dieser Tabelle verwiesen wird, auch in Abschnitt 4.6 (Liste der eingeschlossenen Studien) aufgeführt werden.</p> <p>b: Studie, für die der Unternehmer Sponsor war.</p> <p>c: Zitat der Studienregistereinträge sowie, falls vorhanden, der in den Studienregistern aufgelisteten Berichte über Studiendesign und/oder -ergebnisse.</p> <p>d: Die Studie war außerdem Placebo-kontrolliert.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [9,155-159].</p>						

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-E zu hinterlegen.

Tabelle 4-12: 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
RA-BEAM (JADV)	RCT Phase-III doppelblind parallel Placebo-kontrolliert aktiv-kontrolliert multizentrisch	Erwachsene Patienten (≥ 18 Jahre) mit mittelschwerer bis schwerer aktiver rheumatoider Arthritis (RA), die ein unzureichendes Ansprechen, oder eine Unverträglichkeit gegenüber konventionellen Basistherapeutika (cDMARDs) gezeigt haben	Gesamtstudienteilnehmer: N=1.307 ITT: N=1.307 mITT: N=1.305 (2 der 1.307 Studienteilnehmer wurden nicht behandelt) Baricitinib 4 mg N = 487 Placebo N = 488 Adalimumab 40 mg N = 330	<i>Screening:</i> 3 bis 42 Tage <i>Behandlung:</i> Teil A: Woche 0-24 Teil B: Woche 25-52 <i>Follow-up</i> (Teil C): 28 Tage	Weltweit in 26 Ländern: Argentinien, Belgien, China, Deutschland, Frankreich, Griechenland, Japan, Kanada, Kroatien, Lettland, Litauen, Mexiko, Niederlande, Polen, Portugal, Rumänien, Russland, Schweiz, Slowakei, Slowenien, Spanien, Südafrika, Südkorea, Taiwan, Tschechische Republik, Ungarn, Vereinigtes Königreich, Vereinigte Staaten von	Primärer Endpunkt: <ul style="list-style-type: none">• Anteil Patienten mit ACR20 von Baseline bis Woche 12 im Vergleich zu Placebo Patientenrelevante Sekundäre Endpunkte: <ul style="list-style-type: none">• Anteil Patienten mit DAS28-hsCRP ≤ 3,2 und DAS28-hsCRP < 2,6 von Baseline bis Woche 12, Woche 24 und Woche 52• Anteil Patienten mit HAQ-DI ≥ 0,22 und ≥ 0,3 von Baseline bis Woche 12, Woche 24 und Woche 52• Prozentuelle Veränderung der individuellen Komponenten des ACR Core Set von Baseline von Baseline bis Woche 24 und Woche 52• Veränderung des DAS28-hsCRP von Baseline bis Woche 52• Veränderung des HAQ-DI von Baseline von Baseline bis Woche 24 und Woche 52• Veränderung des CDAI von Baseline von Baseline bis

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

Studie	Studiendesign <RCT, doppel-blind/ 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
					Amerika 10/2012 - 09/2015	Woche 24 und Woche 52 <ul style="list-style-type: none"> • Veränderung des SDAI von Baseline von Baseline bis Woche 24 und Woche 52 • Anteil Patienten mit CDAI $\leq 2,8$ von Baseline bis Woche 24 und Woche 52 • Anteil Patienten mit SDAI $\leq 3,3$ von Baseline bis Woche 24 und 52 • Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert) von Baseline Woche 24 und Woche 52 • Veränderung der Einzel- und Summenskalen des SF-36 v2 akut MCS und des SF-36 v2 akut PCS von Baseline von Baseline bis Woche 24 und Woche 52 • Veränderung des EQ-5D-5L von Baseline von Baseline bis Woche 24 und Woche 52 • Veränderung der Dauer der Morgensteifigkeit von Baseline von Baseline bis Woche 12, Woche 24 und Woche 52 in (ePRO Tablet) • Veränderung der schwersten

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

Studie	Studiendesign <RCT, doppel-blind/ 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
						Müdigkeit (NRS) von Baseline von Baseline bis Woche 52 (ePRO Tablet) weitere Endpunkte: <ul style="list-style-type: none"> • Anteil Patienten mit SDAI ≤ 11 von Baseline bis Woche 24 und Woche 52 • Anteil Patienten mit CDAI ≤ 10 von Baseline bis Woche 12, Woche 24 und Woche 52
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-9].						

Tabelle 4-13: Charakterisierung der Interventionen - RCT mit dem zu bewertenden Arzneimittel

Studie	Baricitinib (N = 487)	Adalimumab (N = 330)	Placebo (N = 480)
RA-BEAM (JADV)	<p><u>Teil A+B^a</u> Baricitinib 4 mg p.o., q.d. eine Tablette von Woche 0-52, Placebo-Injektion (s. c.) alle zwei Wochen von Woche 0-50 (Patienten mit Nierenfunktionsstörung (eGFR < 60 ml/min/1,73 m²) Baricitinib 2 mg p.o., q.d. eine Tablette (kein <i>Rescue</i> zu 4 mg Baricitinib erlaubt))</p> <p><u>Start der <i>Rescue</i>-Therapie</u> ab Woche 16 für Non-Responder Patienten (alle Arme): Patienten wechseln zu Baricitinib 4 mg (+ MTX-Begleitbehandlung)</p>	<p><u>Teil A+B^a</u> Adalimumab-Injektion (s. c.) alle zwei Wochen von Woche 0-50, Placebo p.o., q.d. eine Tablette von Woche 0-52</p> <p><u>Start der <i>Rescue</i>-Therapie:</u> ab Woche 16 für Non-Responder Patienten (alle Arme): Patienten wechseln zu Baricitinib 4 mg (+ MTX-Begleitbehandlung)</p>	<p><u>Teil A^a</u> Placebo p.o., q.d. eine Tablette von Woche 0-24, Placebo-Injektion (s. c.) alle zwei Wochen von Woche 0-50</p> <p><u>Teil B</u> Baricitinib 4 mg p.o., q.d. eine Tablette von Woche 25-52, Placebo-Injektion (s. c.) alle zwei Wochen von Woche 0-50</p> <p><u>Start der <i>Rescue</i>-Therapie</u> ab Woche 16 für Non-Responder Patienten (alle Arme): Patienten wechseln zu Baricitinib 4 mg (+ MTX-Begleitbehandlung)</p>
<p>a: Teil A der Studie RA-BEAM dauerte von Woche 0-24, Teil B dauerte von Woche 25-52. MTX-Begleitbehandlung für alle Arme: stabile MTX-Dosierung (7,5-25 mg) p.o. wöchentlich als Kapsel oder subkutan als Injektion. Die MTX-Dosis darf aus Sicherheitsgründen angepasst werden. Patienten, die nicht an der Extensionsstudie JADY teilnahmen, hatten eine <i>Follow-up</i> Visite (Teil C) ca. 28 Tage nach der letzten Studienmedikation.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-9].</p>			

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

Tabelle 4-14: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel – Demographische Charakteristika der Patienten

Studie	RA-BEAM (JADV) (N = 1.305)			
Behandlungsgruppe Merkmal, Statistisches Maß Gruppe	mITT	Subpopulation A1	Subpopulation A2	Subpopulation A3
Baricitinib, N ^a	487	Nicht zutreffend.	243	170
Adalimumab, N ^a	330		153	124
Demographische Charakteristika der Patienten				
Alter (Jahre), MW (SD)				
Baricitinib	53,5 (12,2)	Nicht zutreffend.	54,3 (12,0)	52,5 (11,7)
Adalimumab	52,9 (12,3)		53,7 (12,0)	53,0 (12,4)
Altersgruppe (Jahre), n (%)				
Baricitinib		Nicht zutreffend.		
< 65	384 (78,9)		187 (77,0)	141 (82,9)
≥ 65	103 (21,1)		56 (23,0)	29 (17,1)
< 75	471 (96,7)		235 (96,7)	168 (98,8)
≥ 75	16 (3,3)		8 (3,3)	2 (1,2)
< 85	487 (100)		243 (100)	170 (100)
≥ 85	0		0	0
Adalimumab		Nicht zutreffend.		
< 65	274 (83,0)		127 (83,0)	102 (82,3)
≥ 65	56 (17,0)		26 (17,0)	22 (17,7)
< 75	323 (97,9)		151 (98,7)	119 (96,0)
≥ 75	7 (2,1)		2 (1,3)	5 (4,0)
< 85	329 (99,7)		152 (99,3)	124 (100)
≥ 85	1 (0,3)		1 (0,7)	0
Geschlecht, n (%)				
Baricitinib		Nicht zutreffend.		
männlich	112 (23,0)		64 (26,3)	32 (18,8)
weiblich	375 (77,0)	179 (73,7)	138 (81,2)	
Adalimumab		Nicht zutreffend.		
männlich	79 (23,9)		48 (31,4)	20 (16,1)
weiblich	251 (76,1)	105 (68,6)	104 (83,9)	

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

Studie	RA-BEAM (JADV) (N = 1.305)			
Behandlungsgruppe Merkmal, Statistisches Maß Gruppe	mITT	Subpopulation A1	Subpopulation A2	Subpopulation A3
Baricitinib, N ^a	487	Nicht zutreffend.	243	170
Adalimumab, N ^a	330		153	124
Demographische Charakteristika der Patienten				
Ethnie, n (%)				
Baricitinib				
indianische Ureinwohner Amerikas/ Ureinwohner Alaskas	19 (3,9)	Nicht zutreffend.	12 (4,9)	2 (1,2)
asiatisch	143 (29,4)		39 (16,0)	65 (38,2)
afroamerikanisch	2 (0,4)		1 (0,4)	0
Ureinwohner Hawaiiis/Pazifik	0		0	0
kaukasisch	312 (64,1)		186 (76,5)	101 (59,4)
Sonstige Angabe	11 (2,3)		5 (2,1)	2 (1,2)
Adalimumab				
indianische Ureinwohner Amerikas/ Ureinwohner Alaskas	18 (5,5)	Nicht zutreffend.	11 (7,2)	5 (4,0)
asiatisch	101 (30,6)		31 (20,3)	38 (30,6)
afroamerikanisch	4 (1,2)		4 (2,6)	0
Ureinwohner Hawaiiis/Pazifik	0		0	0
kaukasisch	204 (61,8)		107 (69,9)	78 (62,9)
Sonstige Angabe	3 (0,9)		0	3 (2,4)
Region, n (%)				
Baricitinib				
Zentral-/Südamerika und Mexiko	143 (29,4)	Nicht zutreffend.	92 (37,9)	38 (22,4)
Vereinigte Staaten von Amerika und Kanada	40 (8,2)		28 (11,5)	6 (3,5)
Japan	93 (19,1)		32 (13,2)	48 (28,2)
Osteuropa	85 (17,5)		42 (17,3)	37 (21,8)
Westeuropa	29 (6,0)		20 (8,2)	6 (3,5)
Asien (ohne Japan)	48 (9,9)		6 (2,5)	16 (9,4)
Restliche	49 (10,1)		23 (9,5)	19 (11,2)

Studie	RA-BEAM (JADV) (N = 1.305)			
Behandlungsgruppe Merkmal, Statistisches Maß Gruppe	mITT	Subpopulation A1	Subpopulation A2	Subpopulation A3
Baricitinib, N ^a	487	Nicht zutreffend.	243	170
Adalimumab, N ^a	330		153	124
Demographische Charakteristika der Patienten				
Adalimumab				
Zentral-/Südamerika und Mexiko	96 (29,1)	Nicht zutreffend.	56 (36,6)	32 (25,8)
Vereinigte Staaten von Amerika und Kanada	26 (7,9)		18 (11,8)	2 (1,6)
Japan	63 (19,1)		28 (18,3)	29 (23,4)
Osteuropa	58 (17,6)		22 (14,4)	30 (24,2)
Westeuropa	20 (6,1)		11 (7,2)	8 (6,5)
Asien (ohne Japan)	33 (10,0)		2 (1,3)	9 (7,3)
Restliche	34 (10,3)		16 (10,5)	14 (11,3)
Gewicht (kg), MW (SD)				
Baricitinib	70,25 (17,57)	Nicht zutreffend.	74,19 (17,88)	66,49 (16,15)
Adalimumab	69,77 (16,14)		71,40 (16,06)	68,99 (15,66)
<p>a: Die Anzahl der Patienten in der mITT Population entspricht nicht der Summe aus Subpopulation A2 und Subpopulation A3, da in Subpopulation A3 30,2 % (127/421 Patienten) der Patienten neben einer Begleitbehandlung mit MTX ein weiteres cDMARD erhalten haben. Da der Anteil Patienten in Subpopulation A3, die das Einschlusskriterium Prüfintervention erfüllen, somit unter 80 % liegt [6], erfolgte eine Neuberechnung der Analysen der Subpopulation A3 unter Ausschluss der Patienten, die neben einer Begleitbehandlung mit MTX ein weiteres cDMARD erhalten hatten.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [160].</p>				

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

Studie	RA-BEAM (JADV) (N = 1.305)			
Behandlungsgruppe Merkmal, Statistisches Maß Gruppe	mITT	Subpopulation A1	Subpopulation A2	Subpopulation A3
Baricitinib, N^a	487	Nicht zutreffend.	243	170
Adalimumab, N^a	330		153	124
Krankheitsgeschichte der Patienten				
Auftreten erster Symptome (Jahre)				
Baricitinib, N	475	Nicht zutreffend.	235	170
MW (SD)	10,3 (8,8)		9,2 (9,4)	12,0 (7,9)
Adalimumab, N	328		153	122
MW (SD)	9,6 (8,5)		8,3 (9,2)	11,5 (7,7)
Auftreten erster Symptome (Jahre) nach Gruppen, n (%)				
Baricitinib, N	475	Nicht zutreffend.	235	170
< 1	28 (5,9)		23 (9,8)	2 (1,2)
≥ 1 bis < 5	140 (29,5)		80 (34,0)	35 (20,6)
≥ 5	307 (64,6)		132 (56,2)	133 (78,2)
Adalimumab, N	328		153	122
< 1	22 (6,7)		18 (11,8)	3 (2,5)
≥ 1 bis < 5	104 (31,7)	60 (39,2)	26 (21,3)	
≥ 5	202 (61,6)	75 (49,0)	93 (76,2)	
Erstdiagnose (Jahre), MW (SD)				
Baricitinib	8,7 (8,6)	Nicht zutreffend.	7,7 (8,9)	10,4 (8,0)
Adalimumab	8,3 (7,9)		6,8 (7,9)	10,2 (7,9)
Kortikosteroidtherapie, n (%)				
Baricitinib		Nicht zutreffend.		
ja	275 (56,5)		127 (52,3)	101 (59,4)
nein	212 (43,5)		116 (47,7)	69 (40,6)
Adalimumab				
ja	201 (60,9)	93 (60,8)	75 (60,5)	
nein	129 (39,1)	60 (39,2)	49 (39,5)	
Dosierung Kortikosteroide (mg/Tag)				
Baricitinib, N	275	Nicht zutreffend.	127	101
MW (SD)	5,9 (2,6)		5,9 (2,5)	5,5 (2,7)
Adalimumab, N	201		93	75
MW (SD)	6,0 (2,4)		6,4 (2,4)	5,8 (2,2)

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

Studie	RA-BEAM (JADV) (N = 1.305)			
Behandlungsgruppe Merkmal, Statistisches Maß Gruppe	mITT	Subpopulation A1	Subpopulation A2	Subpopulation A3
Baricitinib, N ^a	487	Nicht zutreffend.	243	170
Adalimumab, N ^a	330		153	124
Krankheitsgeschichte der Patienten				
Anzahl früherer cDMARD Therapien, n (%)				
Baricitinib				
0	0		0	0
1	243 (49,9)		243 (100)	0
2	138 (28,3)		0	93 (54,7)
≥ 3	106 (21,8)		0	77 (45,3)
Adalimumab				
0	0	Nicht zutreffend.	0	0
1	153 (46,4)		153 (100)	0
2	105 (31,8)		0	77 (62,1)
≥ 3	72 (21,8)		0	47 (37,9)
Art der derzeitigen cDMARD Therapie, n (%)				
Baricitinib				
MTX	413 (84,8)		243 (100)	170 (100)
MTX+ andere cDMARD	74 (15,2)	Nicht zutreffend.	0	0
Nur cDMARD	0		0	0
keine	0		0	0
Adalimumab				
MTX	277 (83,9)		153 (100)	124 (100)
MTX+ andere cDMARD	53 (16,1)	Nicht zutreffend.	0	0
Nur cDMARD	0		0	0
keine	0		0	0
Anzahl derzeitiger cDMARD Therapien, n (%)				
Baricitinib				
0	0		0	0
1	413 (84,8)	Nicht zutreffend.	243 (100)	170 (100)
2	73 (15,0)		0	0
≥ 3	1 (0,2)		0	0

Studie	RA-BEAM (JADV) (N = 1.305)			
Behandlungsgruppe Merkmal, Statistisches Maß Gruppe	mITT	Subpopulation A1	Subpopulation A2	Subpopulation A3
Baricitinib, N ^a	487	Nicht zutreffend.	243	170
Adalimumab, N ^a	330		153	124
Krankheitsgeschichte der Patienten				
Adalimumab				
0	0	Nicht zutreffend.	0	0
1	277 (83,9)		153 (100)	124 (100)
2	53 (16,1)		0	0
≥ 3	0		0	0
<p>a: Die Anzahl der Patienten in der mITT Population entspricht nicht der Summe aus Subpopulation A2 und Subpopulation A3, da in Subpopulation A3 30,2 % (127/421 Patienten) der Patienten neben einer Begleitbehandlung mit MTX ein weiteres cDMARD erhalten haben. Da der Anteil Patienten in Subpopulation A3, die das Einschlusskriterium Prüfintervention erfüllen, somit unter 80 % liegt [6], erfolgte eine Neuberechnung der Analysen der Subpopulation A3 unter Ausschluss der Patienten, die neben einer Begleitbehandlung mit MTX ein weiteres cDMARD erhalten hatten.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [160].</p>				

Tabelle 4-16: Charakterisierung der Studienpopulationen – RCT mit dem zu bewertenden Arzneimittel - Klinische Charakteristika der Patienten

Studie	RA-BEAM (JADV) (N = 1.305)			
Behandlungsgruppe Merkmal, Statistisches Maß Gruppe	mITT	Subpopulation A1	Subpopulation A2	Subpopulation A3
Baricitinib, N ^a	487	Nicht zutreffend.	243	170
Adalimumab, N ^a	330		153	124
Klinische Charakteristika der Patienten				
Anzahl der druckschmerzhaften Gelenke (TJC, basierend auf 28 Gelenken), MW (SD)				
Baricitinib	14,0 (6,6)	Nicht zutreffend.	14,7 (6,7)	13,0 (6,3)
Adalimumab	13,9 (6,9)		14,8 (6,8)	13,0 (6,6)
Anzahl der geschwollenen Gelenke (SCJ, basierend auf 28 Gelenken), MW (SD)				
Baricitinib	11,1 (5,0)	Nicht zutreffend.	10,9 (5,0)	11,0 (4,8)
Adalimumab	11,2 (5,6)		11,8 (6,0)	10,6 (5,1)
Anzahl der druckschmerzhaften Gelenke (TJC, basierend auf 68 Gelenken), MW (SD)				
Baricitinib	23,4 (13,0)	Nicht zutreffend.	24,7 (13,5)	21,0 (11,5)
Adalimumab	23,4 (13,7)		25,3 (14,3)	21,4 (11,8)
Anzahl der geschwollenen Gelenke (SJC, basierend auf 66 Gelenken), MW (SD)				
Baricitinib	15,0 (8,2)	Nicht zutreffend.	14,7 (7,9)	15,0 (8,3)
Adalimumab	15,4 (9,1)		16,4 (10,6)	14,2 (6,3)
HAQ-DI, MW (SD)				
Baricitinib, N	486	Nicht zutreffend.	242	170
MW (SD)	1,57 (0,68)		1,68 (0,71)	1,41 (0,64)
Adalimumab, N	329	Nicht zutreffend.	152	124
MW (SD)	1,59 (0,70)		1,63 (0,69)	1,60 (0,72)
PhGADA (VAS), MW (SD)				
Baricitinib, N	483	Nicht zutreffend.	241	168
MW (SD)	65,7 (16,9)		66,7 (17,3)	62,9 (15,9)
Adalimumab, N	327	Nicht zutreffend.	151	123
MW (SD)	65,3 (16,6)		65,2 (17,2)	65,6 (15,8)
PtGADA, MW (SD)				
Baricitinib, N	486	Nicht zutreffend.	242	170
MW (SD)	63,1 (21,2)		64,8 (22,0)	61,1 (20,9)
Adalimumab, N	329	Nicht zutreffend.	152	124
MW (SD)	63,7 (21,2)		64,3 (21,7)	62,2 (20,5)

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

Studie	RA-BEAM (JADV) (N = 1.305)			
Behandlungsgruppe Merkmal, Statistisches Maß Gruppe	mITT	Subpopulation A1	Subpopulation A2	Subpopulation A3
Baricitinib, N ^a	487	Nicht zutreffend.	243	170
Adalimumab, N ^a	330		153	124
Klinische Charakteristika der Patienten				
PAP (VAS), MW (SD)				
Baricitinib, N	486	Nicht zutreffend.	242	170
MW (SD)	61,8 (21,8)		64,0 (22,5)	58,5 (21,4)
Adalimumab, N	329	Nicht zutreffend.	152	124
MW (SD)	61,0 (22,7)		62,9 (23,2)	58,9 (22,0)
hsCRP (mg/l), MW (SD)				
Baricitinib	22,20 (22,85)	Nicht zutreffend.	22,46 (23,90)	19,47 (19,94)
Adalimumab	21,78 (20,83)		23,31 (23,55)	20,11 (17,48)
DAS28-hsCRP (mg/L), MW (SD)				
Baricitinib, N	486	Nicht zutreffend.	242	170
MW (SD)	5,76 (0,92)		5,83 (0,93)	5,61 (0,91)
Adalimumab, N	329	Nicht zutreffend.	152	124
MW (SD)	5,76 (0,94)		5,88 (0,95)	5,65 (0,88)
DAS28-hsCRP Kategorie, n (%)				
Baricitinib, N	486	Nicht zutreffend.	242	170
≤ 3,2	2 (0,4)		0	2 (1,2)
>3,2 bis ≤ 5,1	117 (24,1)		54 (22,3)	51 (30,0)
> 5,1	367 (75,5)		188 (77,7)	117 (68,8)
Adalimumab, N	329	Nicht zutreffend.	152	124
≤ 3,2	0		0	0
>3,2 bis ≤ 5,1	80 (24,3)		32 (21,1)	33 (26,6)
> 5,1	249 (75,7)		120 (78,9)	91 (73,4)
DAS28-ESR				
Baricitinib, N	486	Nicht zutreffend.	242	170
n (%)	6,45 (0,94)		6,51 (0,92)	6,29 (0,94)
Adalimumab, N	329	Nicht zutreffend.	152	124
n (%)	6,43 (0,96)		6,56 (0,94)	6,34 (0,91)
DAS28-ESR Kategorie, n (%)				
Baricitinib, N	486	Nicht zutreffend.	242	170
≤ 3,2	0		0	0
> 3,2 bis ≤ 5,1	38 (7,8)		16 (6,6)	19 (11,2)
> 5,1	448 (92,2)		226 (93,4)	151 (88,8)

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

Studie	RA-BEAM (JADV) (N = 1.305)			
Behandlungsgruppe Merkmal, Statistisches Maß Gruppe	mITT	Subpopulation A1	Subpopulation A2	Subpopulation A3
Baricitinib, N^a	487	Nicht zutreffend.	243	170
Adalimumab, N^a	330		153	124
Klinische Charakteristika der Patienten				
Adalimumab, N	329	Nicht zutreffend.	152	124
≤ 3,2	0		0	0
> 3,2 bis ≤ 5,1	25 (7,6)		7 (4,6)	11 (8,9)
> 5,1	304 (92,4)		145 (95,4)	113 (91,1)
ESR (mm/h), MW (SD)				
Baricitinib	49,3 (25,8)	Nicht zutreffend.	49,0 (25,5)	46,7 (25,6)
Adalimumab	48,4 (25,6)	Nicht zutreffend.	50,6 (27,9)	47,7 (21,9)
Rheumafaktor (IU/ml), MW (SD)				
Baricitinib	242,7 (410,1)	Nicht zutreffend.	261,8 (377,0)	247,3 (501,9)
Adalimumab	246,0 (451,9)	Nicht zutreffend.	299,3 (506,5)	218,7 (416,6)
Status Rheumafaktor, n (%)				
Baricitinib		Nicht zutreffend.		
Negativ	48 (9,9)		23 (9,5)	17 (10,0)
Positiv	439 (90,1)	220 (90,5)	153 (90,0)	
Adalimumab		Nicht zutreffend.		
Negativ	29 (8,8)		20 (13,1)	5 (4,0)
Positiv	301 (91,2)	133 (86,9)	119 (96,0)	
ACPA (U/ml), MW (SD)				
Baricitinib	629,1 (1.053,7)	Nicht zutreffend.	579,7 (904,0)	782,9 (1.325,8)
Adalimumab	646,5 (1.118,4)	Nicht zutreffend.	701,4 (1.261,0)	644,4 (935,8)
Status ACPA, n (%)				
Baricitinib		Nicht zutreffend.		
Negativ	49 (10,1)		25 (10,3)	14 (8,2)
Positiv	427 (87,7)		213 (87,7)	153 (90,0)
Unbestimmt	11 (2,3)	5 (2,1)	3 (1,8)	
Adalimumab		Nicht zutreffend.		
Negativ	32 (9,7)		18 (11,8)	10 (8,1)
Positiv	295 (89,4)		135 (88,2)	112 (90,3)
Unbestimmt	3 (0,9)	0	2 (1,6)	

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

Studie	RA-BEAM (JADV) (N = 1.305)			
Behandlungsgruppe Merkmal, Statistisches Maß Gruppe	mITT	Subpopulation A1	Subpopulation A2	Subpopulation A3
Baricitinib, N ^a	487	Nicht zutreffend.	243	170
Adalimumab, N ^a	330		153	124
Klinische Charakteristika der Patienten				
Status Seropositivität, n (%)				
Baricitinib				
RF negativ/ ACPA negativ ^b	28 (5,7)	Nicht zutreffend.	16 (6,6)	7 (4,1)
RF positiv/ ACPA negativ ^b	32 (6,6)		14 (5,8)	10 (5,9)
RF negativ/ ACPA positiv	20 (4,1)		7 (2,9)	10 (5,9)
RF positiv/ ACPA positiv	407 (83,6)		206 (84,8)	143 (84,1)
Adalimumab				
RF negativ/ ACPA negativ ^b	14 (4,2)	Nicht zutreffend.	9 (5,9)	3 (2,4)
RF positiv/ ACPA negativ ^b	21 (6,4)		9 (5,9)	9 (7,3)
RF negativ/ ACPA positiv	15 (4,5)		11 (7,2)	2 (1,6)
RF positiv/ ACPA positiv	280 (84,8)		124 (81,0)	110 (88,7)
Modifizierter Gesamt Sharp-Score (mTSS)				
Baricitinib, N	473	Nicht zutreffend.	235	167
MW (SD)	42,46 (50,11)		39,31 (52,49)	47,09 (49,09)
Adalimumab, N	313	Nicht zutreffend.	145	119
MW (SD)	44,36 (50,93)		38,91 (47,19)	48,71 (55,06)
Knochenerosion Score				
Baricitinib, N	473	Nicht zutreffend.	235	167
MW (SD)	25,14 (28,26)		23,31 (29,27)	27,71 (27,85)
Adalimumab, N	313	Nicht zutreffend.	145	119
MW (SD)	26,36 (28,68)		23,94 (26,13)	28,45 (31,59)
Gelenkverengung Score				
Baricitinib, N	473	Nicht zutreffend.	235	167
MW (SD)	17,32 (23,15)		16,00 (24,43)	19,38 (22,46)
Adalimumab, N	313	Nicht zutreffend.	145	119
MW (SD)	18,01 (23,75)		14,97 (22,45)	20,26 (25,09)

Studie	RA-BEAM (JADV) (N = 1.305)			
Behandlungsgruppe Merkmal, Statistisches Maß Gruppe	mITT	Subpopulation A1	Subpopulation A2	Subpopulation A3
Baricitinib, N ^a	487	Nicht zutreffend.	243	170
Adalimumab, N ^a	330		153	124
Klinische Charakteristika der Patienten				
Erosionsstatus, n (%)				
Baricitinib, N	486	Nicht zutreffend.	242	170
1-2 + Seropositivität	115 (23,7)		59 (24,4)	39 (22,9)
3 oder mehr	371 (76,3)		183 (75,6)	131 (77,1)
Adalimumab, N	327	Nicht zutreffend.	151	123
1-2 + Seropositivität	82 (25,1)		40 (26,5)	29 (23,6)
3 oder mehr	245 (74,9)		111 (73,5)	94 (76,4)
<p>a: Die Anzahl der Patienten in der mITT Population entspricht nicht der Summe aus Subpopulation A2 und Subpopulation A3, da in Subpopulation A3 30,2 % (127/421 Patienten) der Patienten neben einer Begleitbehandlung mit MTX ein weiteres cDMARD erhalten haben. Da der Anteil Patienten in Subpopulation A3, die das Einschlusskriterium Prüfintervention erfüllen, somit unter 80 % liegt [6], erfolgte eine Neuberechnung der Analysen der Subpopulation A3 unter Ausschluss der Patienten, die neben einer Begleitbehandlung mit MTX ein weiteres cDMARD erhalten hatten.</p> <p>b: Die Gruppe ACPA negativ enthält Patienten mit negativen (≤ 7 U/ml) und unklaren (> 7 and ≤ 10 U/ml) Werten.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [160].</p>				

Beschreiben Sie die Studien zusammenfassend. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

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

RA-BEAM (JADV)

Zur Bewertung des medizinischen Zusatznutzens von Baricitinib im Vergleich zu Adalimumab wurde für das Anwendungsgebiet Patienten mit mittelschwerer bis schwerer, aktiver RA, die ein unzureichendes Ansprechen oder eine Unverträglichkeit gegenüber Basistherapeutika (konventionelle DMARDs) gezeigt haben, die pivotale Studie RA-BEAM (JADV) herangezogen.

Die Studie RA-BEAM (JADV) ist eine multizentrische, (3:3:2)-randomisierte, doppelblinde, Placebo- und aktiv-kontrollierte Phase-III-Studie, die weltweit in 26 Ländern durchgeführt wurde und eine Behandlungszeit von 52 Wochen hatte. Die Studie diente der Untersuchung der Wirksamkeit und Sicherheit von Baricitinib im Vergleich zu Adalimumab bei Patienten, die unzureichend auf MTX angesprochen haben.

Die Studie bestand nach der Patientenselektion (*Screening*) aus zwei Studienphasen (Teil A und Teil B) und einer vierwöchigen *Follow-up*-Phase (Teil C). Part A war eine Placebo- und aktiv-kontrollierte Phase, die 24 Wochen dauerte. In Teil B wechselten Patienten unter Placebo zu einer aktiven Therapie mit Baricitinib, sodass diese Phase nur noch aktiv-kontrolliert war. Teil B dauerte von Woche 25 bis Woche 52.

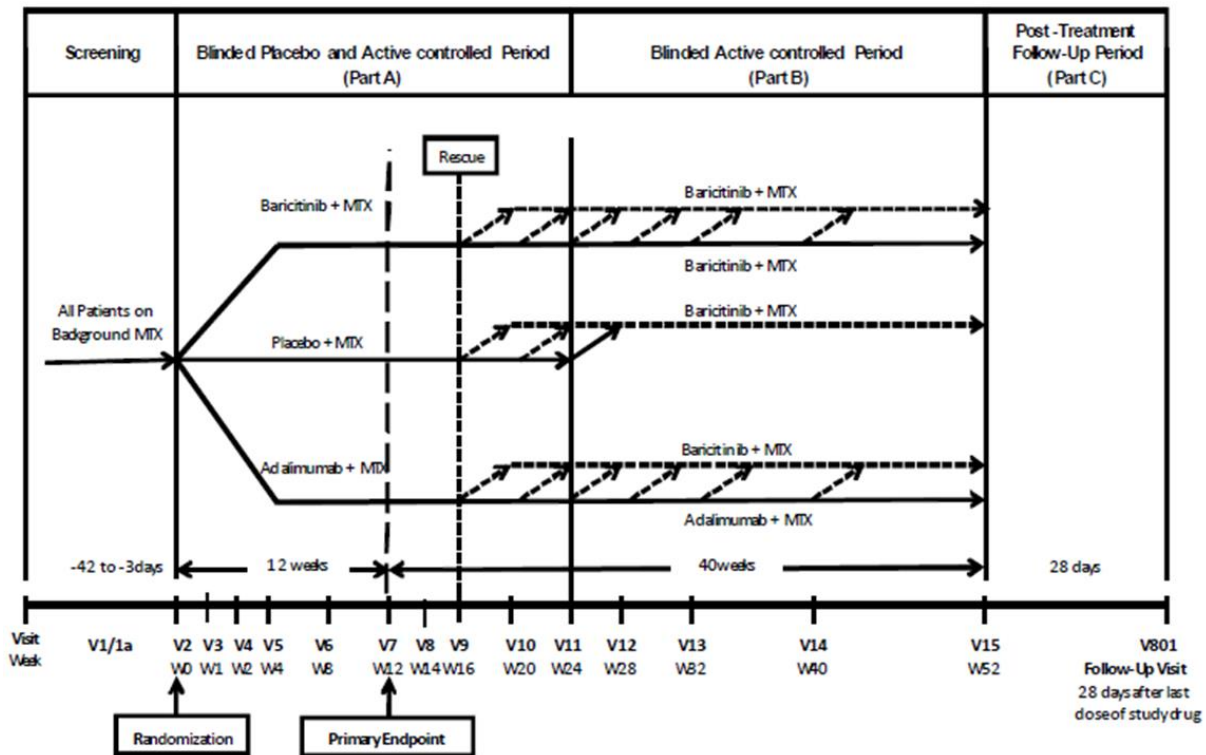
Beim *Screening* wurden Patienten gemäß der Erfüllung der Einschlusskriterien selektiert. Es wurden erwachsene männliche und weibliche Patienten mit einer mittelschweren bis schweren aktiven RA eingeschlossen, die die ACR/EULAR 2010 Kriterien zur Klassifizierung einer rheumatoiden Arthritis erfüllten. Alle Patienten mussten MTX für mindestens 12 Wochen vor Baseline erhalten, davon 8 Wochen mit konstanter Dosis und diese Dosis während des gesamten Studienverlaufes kontinuierlich fortführen (einschließlich der Patienten im Placebo-Behandlungsarm). Im Fall von Unverträglichkeit konnte die Dosis von MTX reduziert werden. Patienten, die zuvor bereits mit bDMARDs behandelt wurden, wurden von der Studie ausgeschlossen. Die *Follow-up*-Phase diente der Erhebung sicherheitsbezogener Parameter nach Absetzen der Studienmedikation.

Insgesamt wurden 1.307 Patienten zu den drei Behandlungsarmen randomisiert. Die Randomisierung wurde stratifiziert nach Region und Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität oder mindestens 3 Gelenkerosionen). Die Zuordnung in die Behandlungsgruppen erfolgte mittels IVRS (*Interactive Voice-Response System*). Zwei Patienten wurden jedoch niemals mit einer Studienmedikation behandelt, sodass nur 1.305 Patienten analysiert wurden (mITT Population) (Baricitinib: n = 487; Adalimumab: n = 330; Placebo: n = 488).

Die Patienten des Baricitinib-Behandlungsarms erhielten von Woche 0 bis Woche 52 einmal täglich 4 mg Baricitinib als Tablette oral verabreicht und alle zwei Wochen eine subkutane Placebo-Injektion (Woche 0 bis Woche 50) (Abbildung 3). Patienten, bei denen eine Nierenfunktionsstörung ($eGFR < 60 \text{ ml/min/1,73 m}^2$) vorlag, erhielten 2 mg Baricitinib täglich. Patienten im Adalimumab-Vergleichsarm erhielten von Woche 0 bis Woche 50 alle zwei Wochen 40 mg Adalimumab als subkutane Injektion und einmal täglich eine orale Placebogabe von Woche 0 bis Woche 52. Patienten unter Placebo erhielten einmal täglich eine orale Placebogabe und alle zwei Wochen eine subkutane Placebo-Injektion von Woche 0 bis Woche 24. Ab Woche 24 erfolgte für die Patienten im Placebo-Behandlungsarm eine Therapie mit Baricitinib.

Alle Patienten erhielten eine Begleitbehandlung mit einer stabilen MTX-Dosierung (7,5 mg bis 25 mg einmal wöchentlich oral in Kapselform oder als subkutane Injektion). Bei Unverträglichkeit der MTX-Dosis konnte diese verringert werden. Die Dosierung von MTX erfolgte entsprechend der medizinischen Praxis, den Empfehlungen der deutschen S1-Leitlinie der Deutschen Gesellschaft für Rheumatologie e.V. (DGRh) zur sequenziellen medikamentösen Therapie der rheumatoiden Arthritis und gemäß Fachinformation [1,5].

Für Patienten, die nicht auf ihre Behandlung ansprachen (*Non-Responder*), erfolgte ab Woche 16 eine *Rescue*-Therapie mit Baricitinib. Patienten, die seit Studienbeginn bereits mit Baricitinib behandelt wurden, erhielten weiterhin Baricitinib, um die Verblindung aufrechten erhalten zu können. Bei erneutem Nicht-Ansprechen unter der *Rescue*-Therapie wurden die Patienten von der Studie ausgeschlossen.



Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [9].

Abbildung 3: Studiendesign der BEAM-Studie (RA-JADV)

Für die statistische Auswertung wurde die mITT Population herangezogen. In dieser Population erhielten alle randomisierten Patienten, mit Ausnahme von 2 Patienten (jeweils ein Patient im Placebo- und Baricitinib-Behandlungsarm aufgrund des Nichterfüllens der Einschlusskriterien für die Studie), mindestens einmal die ihnen zugeordnete Behandlung. Für die Auswertung der Sicherheit wurde die *Safety*-Population verwendet, die definiert war als alle randomisierten Patienten die mindestens eine Studienmedikation erhalten haben und die die Studie nicht aufgrund von *Lost-to-Follow-up* zum ersten Post-Baseline-Visit abgebrochen haben. Die mITT Population entsprach der *Safety*-Population (Baricitinib: n = 487; Adalimumab: n = 330; Placebo: n = 488).

In der Studie RA-BEAM (JADV) wurden Daten zur *Mortalität*, *Morbidität*, *Lebensqualität* sowie *Sicherheit* erhoben. Der primäre Endpunkt war als *ACR20 Response* von Baseline bis Woche 12 definiert.

Die demografischen und krankheitsspezifischen Baseline-Charakteristika der mITT Population bzw. der Subpopulationen A2 und A3 waren zwischen den Behandlungsgruppen vergleichbar (Tabelle 4-14 bis Tabelle 4-16). Über alle Populationen und Behandlungsgruppen hinweg lag das durchschnittliche Alter bei 52,3 Jahren bis 54,3 Jahre und das durchschnittliche Gewicht zwischen 66,31 kg und 74,19 kg. Der Frauenanteil in der Studie RA-BEAM (JADV) war wesentlich höher als der Männeranteil und lag zwischen 68,6% und 82,5%, was der natürlichen Geschlechterverteilung der RA entspricht. Zwischen 51,6% und 76,5% der Patienten waren kaukasischer Ethnie.

Patienten aller Populationen und Behandlungsgruppen zeigten im Durchschnitt vor 8,3 Jahren bis 11,2 Jahren erste Symptome. Die Dauer der RA seit Diagnosestellung betrug 6,8 Jahren bis 9,7 Jahren. 56,5% bis 61,0% der Patienten erhielten eine Kortikosteroidtherapie und alle Patienten haben zuvor mindestens eine Therapie mit cDMARDs erhalten. In Subpopulation A2 erhielten alle Patienten eine Therapie mit einem cDMARD, während alle Patienten in Subpopulation A3 mit mindestens zwei cDMARDs vortherapiert wurden.

Alle Patienten hatten eine mittelschwere bis schwere RA. Die Anzahl der geschwollenen Gelenke, basierend auf dem 66/68 *Joint Count*-System, lag im Durchschnitt bei 14,6 bis 16,4. Patienten hatten zudem durchschnittlich 21,9 bis 25,3 druckschmerzhafte Gelenke (66/68 *Joint Count*-System). Der HAQ-DI lag bei allen Patienten vergleichbar bei 1,45 bis 1,68. Der DAS28-hsCRP lag im Durchschnitt bei 5,66 bis 5,88. 86,9% bis 94,9% aller Patienten waren Rheumafaktor-positiv.

Zwischen den Behandlungsarmen liegen keine relevanten Unterschiede in den Charakteristika der Studienpopulation vor. Der Unterschied zwischen den Subpopulationen A2 und A3 besteht erwartungsgemäß in der Krankheitsgeschichte der Patienten. Aufgrund des internationalen, multizentrischen Studiendesigns sind Unterschiede in Bezug auf die Ethnizität und Länderzugehörigkeit der Studienpopulation im Vergleich zur Zielpopulation im Anwendungsgebiet A der Nutzenbewertung anzunehmen. Die Studienpopulation der Studie RA-BEAM (JADV) entspricht jedoch in den anderen demografischen Baseline-Charakteristika (Alter, Geschlecht) sowie den krankheitsspezifischen Baseline-Charakteristika der Zielpopulation erwachsener Patienten mit mittelschwerer bis schwerer aktiver RA in Deutschland.

Die Studie RA-BEAM (JADV) wurde unter anderem in Deutschland, Europa und den USA durchgeführt. Aufgrund der Strukturgleichheit, v. a. bezüglich der klinischen Parameter, zwischen der Studienpopulation und der Zielpopulation im Anwendungsgebiet A wird davon ausgegangen, dass die in der Studie RA-BEAM (JADV) beobachteten klinischen Effekte auch in der medizinischen Versorgung unter Alltagsbedingungen auftreten und die Studienergebnisse somit auf den deutschen Versorgungskontext übertragbar sind.

Die Anzahl der Patienten in der mITT Population entspricht nicht der Summe aus Subpopulation A2 und Subpopulation A3, da in Subpopulation A3 30,2% (127/421 Patienten) der Patienten neben einer Begleitbehandlung mit MTX ein weiteres cDMARD erhalten haben. Da der Anteil Patienten in Subpopulation A3, die das Einschlusskriterium Prüfintervention erfüllen, somit unter 80% liegt [6], erfolgte eine Neuberechnung der Analysen der Subpopulation A3 unter Ausschluss der Patienten, die neben einer Begleitbehandlung mit MTX ein weiteres cDMARD erhalten hatten.

4.3.1.2.2 Verzerrungspotenzial auf Studienebene

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

Tabelle 4-17: Verzerrungspotenzial auf Studienebene – RCT mit dem zu bewertenden Arzneimittel

Studie	Adäquate Erzeugung der Randomisierungssequenz	Verdeckung der Gruppenzuteilung	Verblindung		Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial auf Studienebene
			Patient	Behandelnde Personen			
RA-BEAM (JADV)	ja	ja	ja	ja	ja	ja	niedrig

Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.
Quelle: [7-9].

Begründen Sie für jede Studie die abschließende Einschätzung.

Bei der Studie RA-BEAM (JADV) handelt es sich um eine multizentrische, randomisierte, doppelblinde, aktiv- und Placebo-kontrollierte Phase-III-Studie zur Untersuchung der Wirksamkeit und Sicherheit von Baricitinib. Das Verzerrungspotenzial wurde als niedrig eingeschätzt, da die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Studienabbrecher/Patienten mit Therapieanpassung

Eine indikationsspezifische Besonderheit ist, dass zur vorliegenden Fragestellung im Rahmen des Studiendesigns bei fehlendem Ansprechen eine Therapieanpassung (*Rescue*-Therapy) möglich war. Die Rahmenbedingungen einer solchen Therapieanpassung waren vorab festgelegt, frühestens 16 Wochen nach Randomisierung möglich und an vordefinierte Kriterien, die als fehlendes Ansprechen auf die randomisiert zugeteilte Therapie zu interpretieren waren, gebunden. Patienten mit Therapieanpassung wurden in der Studie auf verschiedene Weisen in den Analysen berücksichtigt. Insbesondere bei hohen Anteilen von Therapieanpassungen bzw. hohen Unterschieden in der Rate der Anpassung zwischen der Verum- und Kontrollgruppe sind Verzerrungen hinsichtlich des interessierenden Behandlungseffekts möglich. Diese wurden studienspezifisch untersucht und bei der Bewertung des Verzerrungspotenzials auf Endpunktebene berücksichtigt. Ebenso wie Therapieanpassungen wurden für die vorliegende Nutzenbewertung Studienabbrüche aufgrund fehlender Wirksamkeit behandelt.

Für die mITT Population der Studie RA-BEAM (JADV) ist die Anzahl der Patienten, die die Studie aufgrund fehlender Wirksamkeit abbrechen, in beiden Behandlungsarmen sehr gering. Unter Baricitinib brachen die Studie vier Patienten (0,8%) und unter Adalimumab ebenfalls vier Patienten (1,2%) aufgrund fehlender Wirksamkeit ab (Tabelle 4-18). Die entsprechenden Raten sind zwischen Baricitinib und Adalimumab vergleichbar. Die Raten der Patienten, die eine Therapieanpassung (*Rescue*-Therapie) erhielten waren unter Baricitinib 8,8% und unter Adalimumab 15,5% (Tabelle 4-18). Da keine entscheidenden Unterschiede ($\geq 10\%$) zwischen Baricitinib und Adalimumab festzustellen sind, ist nicht von Verzerrungen hinsichtlich des interessierenden Behandlungseffekts aufgrund von Studienabbruchern aufgrund fehlender Wirksamkeit und Patienten mit Therapieanpassung auszugehen.

Für die Subpopulation A2 der Studie RA-BEAM (JADV) ist die Anzahl der Patienten, die die Studie aufgrund fehlender Wirksamkeit abbrechen in beiden Behandlungsarmen sehr gering. Unter Baricitinib brachen die Studie zwei Patienten (0,8%) und unter Adalimumab drei Patienten (2,0%) aufgrund fehlender Wirksamkeit ab (Tabelle 4-18). Diese Raten sind zwischen Baricitinib und Adalimumab vergleichbar. Auch die Raten der Patienten, die eine Therapieanpassung (*Rescue*-Therapie) erhielten sind unter Baricitinib mit 8,6% und unter Adalimumab mit 15,7% vergleichbar (Tabelle 4-18). Da auch bezüglich der Subpopulation A2 keine entscheidenden Unterschiede ($\geq 10\%$) zwischen Baricitinib und Adalimumab festzustellen sind, ist nicht von Verzerrungen hinsichtlich des interessierenden Behandlungseffekts aufgrund von Studienabbruchern aufgrund fehl

Auch für die Subpopulation A3 der Studie RA-BEAM (JADV) ist die Anzahl der Patienten, die die Studie aufgrund fehlender Wirksamkeit abbrechen in beiden Behandlungsarmen sehr gering. Unter Baricitinib brachen die Studie ein Patient (0,6%) und unter Adalimumab kein Patient (0,0%) aufgrund fehlender Wirksamkeit ab (Tabelle 4-18). Diese Raten sind zwischen Baricitinib und Adalimumab vergleichbar. Auch die Raten der Patienten, die eine Therapieanpassung (*Rescue*-Therapie) erhielten sind unter Baricitinib mit 9,4% und unter Adalimumab mit 13,7% vergleichbar (Tabelle 4-18). Da auch bezüglich der Subpopulation A3 keine entscheidenden Unterschiede ($\geq 10\%$) zwischen Baricitinib und Adalimumab festzustellen sind, ist nicht von Verzerrungen hinsichtlich des interessierenden Behandlungseffekts aufgrund von Studienabbruchern aufgrund fehlender Wirksamkeit und Patienten mit Therapieanpassung auszugehen.

Tabelle 4-18: Verzerrungspotenzial auf Studienebene – Studienabbrecher/Patienten mit Therapieanpassung in RCT

Studie	Studienabbrecher aufgrund fehlender Wirksamkeit n (%)	Patienten mit Therapieanpassung (<i>Rescue</i> -Therapie) n (%)
RA-BEAM (JADV)		
Subpopulation A2		
Baricitinib (N=243)	2 (0,8)	21 (8,6)
Adalimumab (N=153)	3 (2,0)	24 (15,7)
Subpopulation A3		
Baricitinib (N=170)	1 (0,6)	16 (9,4)
Adalimumab (N=124)	0 (0,0)	17 (13,7)
mITT		
Baricitinib (N=487)	4 (0,8)	43 (8,8)
Adalimumab (N=330)	4 (1,2)	51 (15,5)
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-9,160].		

Insgesamt ist demnach nicht von verzerrenden Aspekten durch die Anzahl der Patienten, die die Studie aufgrund fehlender Wirksamkeit abbrechen und durch die Anzahl der Patienten mit Therapieanpassung auszugehen, die eine Herabstufung der Ergebnissicherheit auf Endpunktebene bzw. ein erhöhtes Verzerrungspotenzial rechtfertigen würden.

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 4-19: Matrix der Endpunkte in den eingeschlossenen RCT mit dem zu bewertenden Arzneimittel

<i>Nutzendimension</i>	<i>Mortalität^a</i>	<i>Morbidität</i>	<i>Lebensqualität</i>	<i>Sicherheit</i>
Studie RA-BEAM				
Erhoben anhand der Endpunkte	ja	ja	ja	ja
Morbidität				
Druckschmerzhaftes Gelenke und geschwollene Gelenke				
Veränderung der Anzahl der druckschmerzhaften Gelenke (basierend auf 28 Gelenken)	—	●	—	—
Veränderung der Anzahl der geschwollenen Gelenke (basierend auf 28 Gelenken)	—	●	—	—
DAS28-hsCRP				
Veränderung des DAS28-hsCRP (mBOCF)	—	●	—	—
Veränderung des DAS28-hsCRP (mLOCF)	—	●	—	—
Veränderung des DAS28-hsCRP (MMRM)	—	●	—	—
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	—	●	—	—
Anteil Patienten mit DAS28-hsCRP < 2,6	—	●	—	—
SDAI				
Veränderung des SDAI	—	●	—	—
Anteil Patienten mit SDAI ≤ 3,3	—	●	—	—
CDAI				
Veränderung des CDAI	—	●	—	—
Anteil Patienten mit CDAI ≤ 10	—	●	—	—
EULAR Responsekriterien („good Response“, DAS28-basiert)				
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	—	●	—	—
PAP (VAS)				
Veränderung des PAP (VAS)	—	●	—	—
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	—	●	—	—
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	—	●	—	—

<i>Nutzendimension</i>	<i>Mortalität^a</i>	<i>Morbidität</i>	<i>Lebensqualität</i>	<i>Sicherheit</i>
Studie RA-BEAM				
Erhoben anhand der Endpunkte	ja	ja	ja	ja
PtGADA (VAS)				
Veränderung des PtGADA (VAS)	—	●	—	—
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	—	●	—	—
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	—	●	—	—
HAQ-DI				
Veränderung des HAQ-DI (mBOCF)	—	●	—	—
Veränderung des HAQ-DI (mLOCF)	—	●	—	—
Veränderung des HAQ-DI (MMRM)	—	●	—	—
Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$	—	●	—	—
Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$	—	●	—	—
Dauer der Morgensteifigkeit (min)				
Veränderung der Dauer der Morgensteifigkeit (min)	—	●	—	—
EQ-5D VAS				
Veränderung der EQ-5D VAS	—	●	—	—
Anteil Patienten mit einer Verbesserung EQ-5D VAS ≥ 10 mm	—	●	—	—
Schwerste Müdigkeit				
Veränderung der schwersten Müdigkeit	—	●	—	—
Lebensqualität				
SF-36 v2 akut				
Veränderung des SF-36 v2 akut PCS	—	—	●	—
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5	—	—	●	—
Veränderung des SF-36 v2 akut MCS	—	—	●	—
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5	—	—	●	—
Veränderung der SF-36 v2 akut Domänen				
Körperliche Funktionsfähigkeit	—	—	●	—
Körperliche Rollenfunktion	—	—	●	—
Körperliche Schmerzen	—	—	●	—
Allgemeiner Gesundheitszustand	—	—	●	—

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

<i>Nutzendimension</i>	<i>Mortalität^a</i>	<i>Morbidität</i>	<i>Lebensqualität</i>	<i>Sicherheit</i>
Studie RA-BEAM				
Erhoben anhand der Endpunkte	ja	ja	ja	ja
Vitalität	—	—	●	—
Soziale Funktionsfähigkeit	—	—	●	—
Emotionale Rollenfunktion	—	—	●	—
Mentaler Gesundheitszustand	—	—	●	—
Sicherheit				
Todesfälle	●	—	—	●
SUE	—	—	—	●
UE	—	—	—	●
Schwere UE (ohne ausschließliche Laborereignisse)	—	—	—	●
Studienabbruch aufgrund von UE	—	—	—	●
Therapieabbruch aufgrund von UE	—	—	—	●
UE von besonderem Interesse				
Infektionen	—	—	—	●
Schwere Infektionen	—	—	—	●
Kardiovaskuläre UE	—	—	—	●
Malignität	—	—	—	●
Depressionen: QIDS-SR16	—	—	—	●
Allergische Reaktionen/Hypersensitivität	—	—	—	●
Gastrointestinale Perforation	—	—	—	●
<p>a: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [7-10].</p>				

4.3.1.3.1 Mortalität – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktebene in einer Tabelle zusammengefasst werden. Dann sollen die Ergebnisse der einzelnen Studien zu dem Endpunkt tabellarisch dargestellt und in einem Text zusammenfassend beschrieben werden. Anschließend sollen die Ergebnisse, wenn möglich und sinnvoll, in einer 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-20: Operationalisierung der Mortalität

Studie	Operationalisierung
RA-BEAM (JADV)	Nicht zutreffend ^a .
<p>a: Da Mortalität nicht als primärer oder sekundärer Endpunkt vorgesehen war und nicht im Rahmen einer Überlebenszeitanalyse untersucht wurde, wurde die Anzahl der Todesfälle im Rahmen der schwerwiegenden unerwünschten Ereignisse als Annäherung zur Beurteilung dieses patientenrelevanten Endpunktes ausgewertet (Abschnitt 4.3.1.3.4).</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [7-10].</p>	

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

Tabelle 4-21: Bewertung des Verzerrungspotenzials für Mortalität in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	Nicht zutreffend ^a .					
<p>a: Da Mortalität nicht als primärer oder sekundärer Endpunkt vorgesehen war und nicht im Rahmen einer Überlebenszeitanalyse untersucht wurde, wurde die Anzahl der Todesfälle im Rahmen der schwerwiegenden unerwünschten Ereignisse als Annäherung zur Beurteilung dieses patientenrelevanten Endpunktes ausgewertet (Abschnitt 4.3.1.3.4).</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [7-10].</p>						

Begründen Sie für jede Studie die abschließende Einschätzung.

Da Mortalität nicht als primärer oder sekundärer Endpunkt vorgesehen war und nicht im Rahmen einer Überlebenszeitanalyse untersucht wurde, wurde die Anzahl der Todesfälle im Rahmen der schwerwiegenden unerwünschten Ereignisse als Annäherung zur Beurteilung dieses patientenrelevanten Endpunktes ausgewertet (Abschnitt 4.3.1.3.4).

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

Tabelle 4-22: Ergebnisse für Mortalität aus RCT mit dem zu bewertenden Arzneimittel

Studie	Tabellarische Präsentation in geeigneter Form (Anforderungen siehe Erläuterung oben)
RA-BEAM (JADV)	Nicht zutreffend.
<p>a: Da Mortalität nicht als primärer oder sekundärer Endpunkt vorgesehen war und nicht im Rahmen einer Überlebenszeitanalyse untersucht wurde, wurde die Anzahl der Todesfälle im Rahmen der schwerwiegenden unerwünschten Ereignisse als Annäherung zur Beurteilung dieses patientenrelevanten Endpunktes ausgewertet (Abschnitt 4.3.1.3.4).</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [7-10].</p>	

Nicht zutreffend.

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2 Morbidität – RCT

4.3.1.3.2.1 Druckschmerzhafte Gelenke und geschwollene Gelenke – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-23: Operationalisierung der Anzahl druckschmerzhafter Gelenke und geschwollener Gelenke

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Baseline (Woche 0); Woche 24, Woche 52 (ANCOVA mit mLOCF)</p> <ul style="list-style-type: none"> • Änderung der Anzahl der druckschmerzhaften Gelenke und geschwollenen Gelenke (ANCOVA mit mLOCF) <p><u>Definition:</u> In der Studie RA-BEAM (JADV) wurden 66 (Schwellung) bzw. 68 (Druckschmerz) Gelenke untersucht und gemäß 2-Punkte-Skala als druckschmerzhaft oder nicht druckschmerzhaft bzw. als geschwollen oder nicht geschwollen beurteilt. Zu den auf Druckschmerzhaftigkeit und Schwellung zu untersuchenden Gelenken, die über den gesamten Körper verteilt sind, gehören u. a. das Kiefer-, Sternoklavikulär-, Schulterreck-, Schulter-, Ellenbogen-, Hand-, Daumengrund-, Fingermittel-, Fingerend-, Hüft-, Knie-, Knöchel- sowie Großzehengrundgelenk und die distalen Interphalangealgelenke der Füße sowie der Tarsus. Die beiden Hüftgelenke werden lediglich auf Druckschmerzhaftigkeit untersucht, da die Hüftgelenke konstitutionsbedingt nicht bei jedem Menschen im Hinblick auf Schwellung beurteilt werden können.</p> <p>Für die hier vorliegende Analyse wurde der 28 <i>Joint Count</i> verwendet.</p> <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].	

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

Tabelle 4-24: Bewertung des Verzerrungspotenzials für Anzahl druckschmerzhafter Gelenke und geschwollener Gelenke

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	ja	niedrig
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].						

Begründen Sie für jede Studie die abschließende Einschätzung.

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

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt Anzahl druckschmerzhafter Gelenke und geschwollener Gelenke als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Stellen Sie die Ergebnisse für den Endpunkt Anzahl druckschmerzhafter Gelenke und geschwollener Gelenke für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Veränderung der Anzahl der druckschmerzhaften Gelenke (basierend auf 28 Gelenken)

Tabelle 4-25: Ergebnisse für Veränderung der Anzahl der druckschmerzhaften Gelenke (basierend auf 28 Gelenken) aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges' g (95% KI)
Veränderung der Anzahl der druckschmerzhaften Gelenke (basierend auf 28 Gelenken)									
Woche 24									
Subpopulation A1 Nicht zutreffend.									
Subpopulation A2									
Baricitinib	243	243	14,7 (6,7)	239	4,4 (5,6)	239	-9,5 (0,47)	-0,8 (-2,1; 0,4); 0,196	-0,1 (-0,3; 0,1)
Adalimumab	153	153	14,8 (6,8)	153	5,3 (6,4)	153	-8,7 (0,56)		
Subpopulation A3									
Baricitinib	170	170	13,0 (6,3)	170	4,3 (5,5)	170	-8,7 (0,51)	0,1 (-1,2; 1,4); 0,885	0,0 (-0,2; 0,2)
Adalimumab	124	124	13,0 (6,6)	123	4,2 (5,3)	123	-8,8 (0,58)		
mITT									
Baricitinib	487	487	14,0 (6,6)	483	4,5 (5,7)	483	-9,1 (0,30)	-0,6 (-1,5; 0,2); 0,135	-0,1 (-0,2; 0,0)
Adalimumab	330	330	13,9 (6,9)	328	5,1 (6,2)	328	-8,5 (0,35)		
Heterogenität^b								p = 0,623	

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

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der Anzahl der druckschmerzhaften Gelenke (basierend auf 28 Gelenken)									
Woche 52									
Subpopulation A1 Nicht zutreffend.									
Subpopulation A2									
Baricitinib	243	243	14,7 (6,7)	239	4,0 (5,8)	239	-10,4 (0,50)	-0,8 (-2,0; 0,4); 0,184	-0,1 (-0,3; 0,1)
Adalimumab	153	153	14,8 (6,8)	153	5,0 (6,6)	153	-9,6 (0,58)		
Subpopulation A3									
Baricitinib	170	170	13,0 (6,3)	170	3,8 (5,4)	170	-8,7 (0,51)	-0,6 (-1,8; 0,6); 0,325	-0,1 (-0,3; 0,1)
Adalimumab	124	124	13,0 (6,6)	123	4,4 (5,5)	123	-8,1 (0,57)		
mITT									
Baricitinib	487	487	14,0 (6,6)	483	4,1 (5,8)	483	-9,5 (0,30)	-0,9 (-1,6; -0,1); 0,032	-0,1 (-0,3; 0,0)
Adalimumab	330	330	13,9 (6,9)	328	4,9 (6,4)	328	-8,7 (0,34)		
Heterogenität^b p = 0,806									
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.									
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.									
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.									
Quelle: [59].									

Für **Subpopulation A2** war die Anzahl der druckschmerzhaften Gelenke unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab verringert. Der Unterschied war statistisch nicht signifikant.

Für **Subpopulation A3** war die Anzahl der druckschmerzhaften Gelenke unter Adalimumab von Baseline bis Woche 24 im Vergleich zu Baricitinib verringert. Von Baseline bis Woche 52 war die Anzahl der druckschmerzhaften Gelenke unter Baricitinib im Vergleich zu Adalimumab verringert. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung der Anzahl der druckschmerzhaften Gelenke über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

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

Für die **mITT Population** war die Anzahl der druckschmerzhaften Gelenke unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab verringert. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant (Tabelle 4-25).

Veränderung der Anzahl der geschwollenen Gelenke (basierend auf 28 Gelenken)

Tabelle 4-26: Ergebnisse für Veränderung der Anzahl der geschwollenen Gelenke (basierend auf 28 Gelenken) aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges' g (95% KI)
Veränderung der Anzahl der geschwollenen Gelenke (basierend auf 28 Gelenken)									
Woche 24									
Subpopulation A1					Nicht zutreffend.				
Subpopulation A2									
Baricitinib	243	243	10,9 (5,0)	239	3,1 (4,4)	239	-7,3 (0,37)	-0,4 (-1,4; 0,6); 0,407	-0,1 (-0,3; 0,1)
Adalimumab	153	153	11,8 (6,0)	153	3,9 (5,5)	153	-6,9 (0,44)		
Subpopulation A3									
Baricitinib	170	170	11,0 (4,8)	170	3,2 (4,3)	170	-7,7 (0,42)	-0,6 (-1,7; 0,4); 0,244	-0,1 (-0,3; 0,1)
Adalimumab	124	124	10,6 (5,1)	123	3,5 (4,4)	123	-7,1 (0,48)		
mITT									
Baricitinib	487	487	11,1 (5,0)	483	3,3 (4,6)	483	-7,6 (0,24)	-0,5 (-1,2; 0,2); 0,159	-0,1 (-0,2; 0,0)
Adalimumab	330	330	11,2 (5,6)	328	3,8 (5,0)	328	-7,1 (0,28)		
Heterogenität^b								p = 0,236	

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges' g (95% KI)
Veränderung der Anzahl der geschwollenen Gelenke (basierend auf 28 Gelenken)									
Woche 52									
Subpopulation A1 Nicht zutreffend.									
Subpopulation A2									
Baricitinib	243	243	10,9 (5,0)	239	2,7 (4,6)	239	-8,0 (0,40)	-0,8 (-1,7; 0,2); 0,117	-0,1 (-0,3; 0,1)
Adalimumab	153	153	11,8 (6,0)	153	3,9 (5,8)	153	-7,2 (0,47)		
Subpopulation A3									
Baricitinib	170	170	11,0 (4,8)	170	2,9 (4,4)	170	-8,1 (5,7)	-0,6 (-1,6; 0,4); 0,270	-0,1 (-0,3; 0,1)
Adalimumab	124	124	10,6 (5,1)	123	3,3 (4,6)	123	-7,2 (5,4)		
mITT									
Baricitinib	487	487	11,1 (5,0)	483	2,9 (4,5)	483	-7,9 (0,24)	-0,9 (-1,5; -0,2); 0,007	-0,2 (-0,3; -0,0)
Adalimumab	330	330	11,2 (5,6)	328	3,8 (5,3)	328	-7,0 (0,28)		
Heterogenität^b								p = 0,814	
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.									
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.									
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.									
Quelle: [59].									

Für **Subpopulation A2** war die Anzahl der geschwollenen Gelenke unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab verringert. Der Unterschied war statistisch nicht signifikant.

Für **Subpopulation A3** war die Anzahl der geschwollenen Gelenke unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab verringert. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung der Anzahl der geschwollenen Gelenke über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

Für die **mITT Population** war die Anzahl der geschwollenen Gelenke unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab verringert. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant (Tabelle 4-26).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2.2 Disease Activity Score 28 High Sensitivity C-Reactive Protein (DAS28-hsCRP) – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-27: Operationalisierung des DAS28-hsCRP

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Baseline (Woche 0), Woche 24, Woche 52</p> <ul style="list-style-type: none"> • Veränderung des DAS28-hsCRP von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mBOCF/mLOCF, MMRM) • Anteil Patienten mit DAS28-hsCRP $\leq 2,6$ an Woche 24 und Woche 52 (NRI) • Anteil Patienten mit DAS28-hsCRP $\leq 3,2$ an Woche 24 und Woche 52 (NRI) <p><u>Definition:</u> Der DAS28-hsCRP ermöglicht die Beurteilung der Krankheitsaktivität und setzt sich wie folgt zusammen:</p> <ul style="list-style-type: none"> • Anzahl der druckschmerzhaften Gelenke (basierend auf der Untersuchung von 28 Gelenken) • Anzahl der geschwollenen Gelenke (basierend auf der Untersuchung von 28 Gelenken) • PtGADA (VAS) • hsCRP. <p>Die Krankheitsaktivität gemessen mittels DAS28 wird gemäß ACR und EULAR in verschiedene Stufen eingeteilt. Bei einem DAS28-Score von $< 2,6$ liegt eine Remission vor. Des Weiteren erfolgt die Einteilung der Krankheitsaktivität mittels DAS28 in die Stufen gering ($\geq 2,6$ bis $< 3,2$), mittelschwer ($\geq 3,2$ bis $\leq 5,1$) und hoch ($> 5,1$)</p> <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>	

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

Tabelle 4-28: Bewertung des Verzerrungspotenzials für DAS28-hsCRP in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	ja	niedrig
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].						

Begründen Sie für jede Studie die abschließende Einschätzung.

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt DAS28-hsCRP als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Stellen Sie die Ergebnisse für den Endpunkt DAS28-hsCRP für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Veränderung des DAS28-hsCRP

Tabelle 4-29: Ergebnisse für Veränderung des DAS28-hsCRP aus RCT mit dem zu bewertenden Arzneimittel (mBOCF)

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des DAS28-hsCRP								
Woche 24								
Subpopulation A1		Nicht zutreffend.						
Subpopulation A2								
Baricitinib	243	242	5,83 (0,93)	241	3,28 (1,38)	243	-2,45 (0,104)	-0,21 (-0,48; 0,07); 0,141
Adalimumab	153	152	5,88 (0,95)	152	3,52 (1,56)	153	-2,24 (0,125)	
Subpopulation A3								
Baricitinib	170	170	5,61 (0,91)	170	3,20 (1,38)	170	-2,36 (0,114)	-0,10 (-0,39; 0,19); 0,509
Adalimumab	124	124	5,65 (0,88)	123	3,28 (1,42)	124	-2,26 (0,129)	
mITT								
Baricitinib	487	486	5,76 (0,92)	485	3,28 (1,39)	487	-2,43 (0,065)	-0,21 (-0,39; -0,03); 0,022
Adalimumab	330	329	5,76 (0,94)	327	3,47 (1,52)	330	-2,22 (0,077)	
Heterogenität^b								p = 0,339
Woche 52								
Subpopulation A1		Nicht zutreffend.						
Subpopulation A2								
Baricitinib	243	242	5,83 (0,93)	241	3,21 (1,55)	243	-2,60 (0,130)	-0,15 (-0,46; 0,16); 0,332
Adalimumab	153	152	5,88 (0,95)	152	3,40 (1,67)	153	-2,45 (0,152)	
Subpopulation A3								
Baricitinib	170	170	5,61 (0,91)	170	3,06 (1,50)	170	-2,40 (0,144)	-0,32 (-0,65; 0,01); 0,058
Adalimumab	124	124	5,65 (0,88)	123	3,36 (1,52)	124	-2,08 (0,159)	

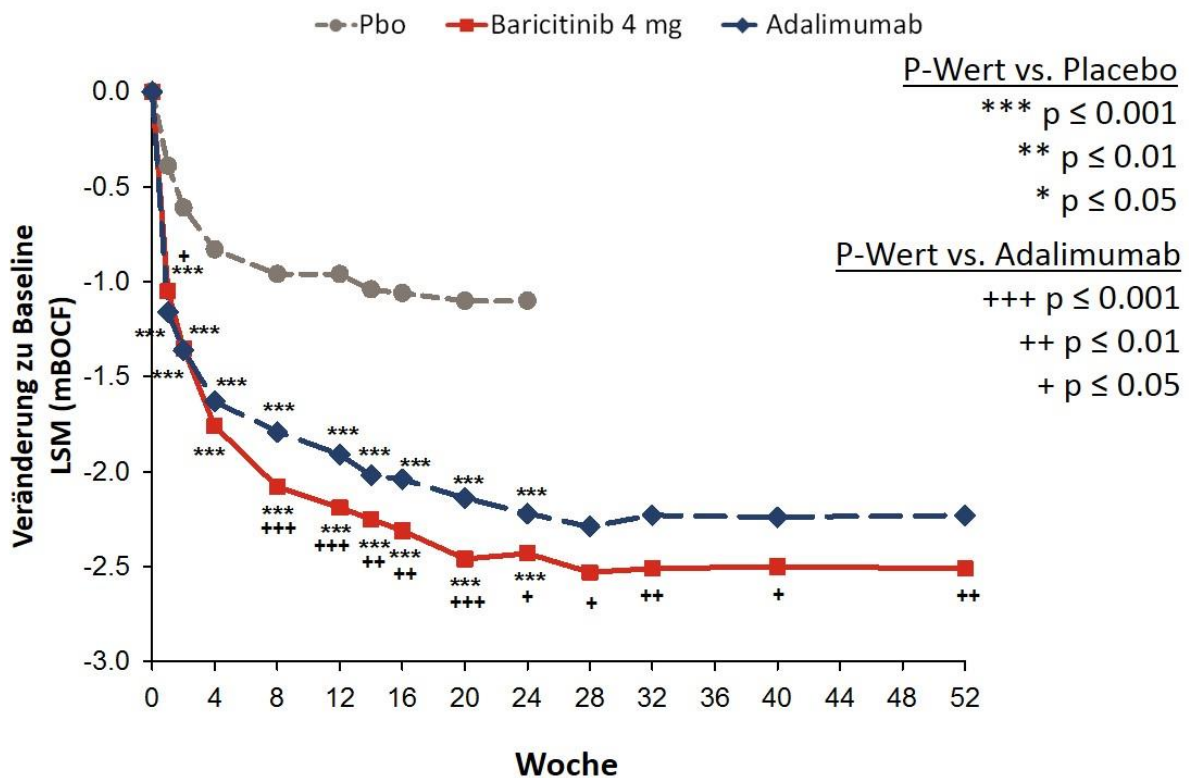
RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des DAS28-hsCRP								
mITT								
Baricitinib	487	486	5,76 (0,92)	485	3,17 (1,53)	487	-2,51 (0,079)	-0,29 (-0,49; -0,08); 0,007
Adalimumab	330	329	5,76 (0,94)	327	3,45 (1,64)	330	-2,23 (0,091)	
Heterogenität^b								p = 0,444
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mBOCF) berechnet.								
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mBOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.								
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.								
Quelle: [59].								

Für **Subpopulation A2** (mBOCF) konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Der Unterschied war statistisch nicht signifikant.

Für **Subpopulation A3** (mBOCF) konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung des DAS28-hsCRP (mBOCF) über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

Für die **mITT Population** (mBOCF) konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Von Baseline bis Woche 24 und Woche 52 war der Unterschied statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität (Tabelle 4-29).



Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [161].

Abbildung 4: Zeitlicher Verlauf des DAS28-hsCRP (mBOCF) über 52 Wochen (mITT)

Abbildung 4 zeigt den zeitlichen Verlauf des DAS28-hsCRP über 52 Wochen für die Behandlungsarme Baricitinib, Adalimumab und Placebo. Baricitinib ist Adalimumab ab Woche 8 statistisch signifikant überlegen ($p < 0,001$). Im weiteren Behandlungsverlauf bleibt die statistische signifikante Überlegenheit von Baricitinib gegenüber Adalimumab bis zum Behandlungsende (Woche 52) bestehen.

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

Tabelle 4-30: Ergebnisse für Veränderung des DAS28-hsCRP aus RCT mit dem zu bewertenden Arzneimittel (Sensitivitätsanalyse - mLOCF)

RA-BEAM (JADV)	N	Baseline		Woche 52		Veränderung zu Baseline		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des DAS28-hsCRP								
Woche 52								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	242	5,83 (0,93)	238	3,04 (1,38)	238	-2,79 (0,119)	-0,22 (-0,50; 0,07); 0,135
Adalimumab	153	152	5,88 (0,95)	152	3,29 (1,59)	152	-2,58 (0,139)	
Subpopulation A3								
Baricitinib	170	170	5,61 (0,91)	170	2,89 (1,36)	170	-2,55 (0,134)	-0,39 (-0,70; -0,08); 0,014
Adalimumab	124	124	5,65 (0,88)	123	3,29 (1,50)	123	-2,16 (0,149)	
mITT								
Baricitinib	487	486	5,76 (0,92)	482	3,00 (1,37)	482	-2,68 (0,072)	-0,35 (-0,54; -0,16); < 0,001
Adalimumab	330	329	5,76 (0,94)	327	3,35 (1,58)	327	-2,33 (0,084)	
Heterogenität^b p = 0,393								
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.								
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.								
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.								
Quelle: [59].								

Für **Subpopulation A2** (mLOCF) konnte unter Baricitinib von Baseline bis Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Der Unterschied war statistisch nicht signifikant.

Für **Subpopulation A3** (mLOCF) konnte unter Baricitinib von Baseline bis Woche 52 im Vergleich zu Adalimumab eine statistisch signifikante niedrigere Krankheitsaktivität beobachtet werden. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität.

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

Da der Heterogenitätstest für den Endpunkt Veränderung des DAS28-hsCRP (mLOCF) über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

Für die **mITT Population** (mLOCF) konnte unter Baricitinib von Baseline bis Woche 52 im Vergleich zu Adalimumab eine statistisch signifikante niedrigere Krankheitsaktivität beobachtet werden. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität (Tabelle 4-30).

Tabelle 4-31: Ergebnisse Veränderung des DAS28-hsCRP aus RCT mit dem zu bewertenden Arzneimittel (Sensitivitätsanalyse - MMRM)

RA-BEAM (JADV)	N	n	Veränderung zu Baseline	Baricitinib vs. Adalimumab
			LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des DAS28-hsCRP				
Woche 24				
Subpopulation A1			Nicht zutreffend.	
Subpopulation A2				
Baricitinib	243	210	-2,57 (0,09)	-0,15 (-0,40; 0,10); 0,239
Adalimumab	153	123	-2,42 (0,11)	
Subpopulation A3				
Baricitinib	170	152	-2,59 (0,10)	-0,11 (-0,38; 0,16); 0,431
Adalimumab	124	103	-2,48 (0,11)	
mITT				
Baricitinib	487	424	-2,59 (0,06)	-0,19 (-0,36; -0,02); 0,029
Adalimumab	330	265	-2,40 (0,07)	
Woche 52				
Subpopulation A1			Nicht zutreffend.	
Subpopulation A2				
Baricitinib	243	193	-2,77 (0,09)	-0,02 (-0,29; 0,25); 0,866
Adalimumab	153	107	-2,74 (0,11)	
Subpopulation A3				
Baricitinib	170	140	-2,85 (0,10)	-0,35 (-0,64; -0,06); 0,020
Adalimumab	127	94	-2,50 (0,12)	
mITT				
Baricitinib	487	391	-2,82 (0,06)	-0,23 (-0,41; -0,05); 0,014
Adalimumab	330	236	-2,60 (0,07)	
Heterogenität^b				p = 0,662

RA-BEAM (JADV)	N	n	Veränderung zu Baseline	Baricitinib vs. Adalimumab
			LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des DAS28-hsCRP				
<p>a: p-Werte, LSM, SE, LSMD und 95%-Konfidenzintervall wurden mittels eines MMRM-Modells berechnet.</p> <p>b: Interaktions-p-Wert mittels eines MMRM-Modells berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>				

Für **Subpopulation A2** (MMRM) konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Der Unterschied war statistisch nicht signifikant.

Für **Subpopulation A3** (MMRM) konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung des DAS28-hsCRP (MMRM) über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

Für die **mITT Population** (MMRM) konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine statistisch signifikante niedrigere Krankheitsaktivität beobachtet werden. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität (Tabelle 4-31).

Anteil Patienten mit DAS28-hsCRP \leq 3,2

Tabelle 4-32: Ergebnisse für Anteil Patienten mit DAS28-hsCRP \leq 3,2 aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI)^d; p-Wert	RR^a (95% KI)^d; p-Wert	ARD^b (95% KI)^c; p-Wert^c
Anteil Patienten mit DAS28-hsCRP \leq 3,2					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	125 (51,4)	1,18 (0,78; 1,79); 0,425	1,10 (0,90; 1,34); 0,351	3,73 (-6,38; 13,83); 0,536
Adalimumab	153	73 (47,7)			
Subpopulation A3					
Baricitinib	170	97 (57,1)	1,21 (0,75; 1,94); 0,436	1,07 (0,87; 1,32); 0,534	4,64 (-6,88; 16,16); 0,477
Adalimumab	124	65 (52,4)			
mITT					
Baricitinib	487	255 (52,4)	1,21 (0,91; 1,61); 0,194	1,08 (0,95; 1,24); 0,251	4,48 (-2,50; 11,46); 0,226
Adalimumab	330	158 (47,9)			
Heterogenität^e p = 0,1822					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	139 (57,2)	1,35 (0,89; 2,05); 0,156	1,14 (0,95; 1,38); 0,161	7,53 (-2,54; 17,60); 0,148
Adalimumab	153	76 (49,7)			
Subpopulation A3					
Baricitinib	170	93 (54,7)	1,14 (0,71; 1,83); 0,589	1,05 (0,85; 1,31); 0,640	3,09 (-8,46; 14,64); 0,637
Adalimumab	124	64 (51,6)			
mITT					
Baricitinib	487	271 (55,6)	1,35 (1,01; 1,79); 0,042	1,14 (1,00; 1,30); 0,059	7,46 (0,50; 14,43); 0,039
Adalimumab	330	159 (48,2)			
Heterogenität^e p = 0,6534					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit DAS28-hsCRP ≤ 3,2					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine niedrige Krankheitsaktivität als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine niedrige Krankheitsaktivität als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit DAS28-hsCRP ≤ 3,2 über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine niedrige Krankheitsaktivität als unter Adalimumab. Der Unterschied war statistisch nicht signifikant (Tabelle 4-32).

Anteil Patienten mit DAS28-hsCRP <2,6

Tabelle 4-33: Ergebnisse für Anteil Patienten mit DAS28-hsCRP < 2,6 aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI)^d; p-Wert	RR^a (95% KI)^d; p-Wert	ARD^b (95% KI)^c; p-Wert^c
Anteil Patienten mit DAS28-hsCRP < 2,6					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	82 (33,7)	1,09 (0,70; 1,69); 0,717	1,06 (0,80; 1,41); 0,663	1,07 (-8,45; 10,58); 0,913
Adalimumab	153	50 (32,7)			
Subpopulation A3					
Baricitinib	170	62 (36,5)	1,14 (0,69; 1,88); 0,613	1,07 (0,78; 1,46); 0,679	2,60 (-8,43; 13,63); 0,711
Adalimumab	124	42 (33,9)			
mITT					
Baricitinib	487	168 (34,5)	1,14 (0,84; 1,55); 0,394	1,09 (0,90; 1,33); 0,385	2,68 (-3,88; 9,24); 0,450
Adalimumab	330	105 (31,8)			
Heterogenität^e p = 0,1451					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	99 (40,7)	0,92 (0,60; 1,41); 0,701	0,96 (0,76; 1,22); 0,755	-1,74 (-11,72; 8,23); 0,754
Adalimumab	153	65 (42,5)			
Subpopulation A3					
Baricitinib	170	68 (40,0)	1,04 (0,64; 1,69); 0,867	1,03 (0,77; 1,37); 0,858	0,48 (-10,84; 11,81); 1,000
Adalimumab	124	49 (39,5)			
mITT					
Baricitinib	487	193 (39,6)	1,01 (0,76; 1,36); 0,929	1,01 (0,85; 1,20); 0,904	0,54 (-6,29; 7,37); 0,884
Adalimumab	330	129 (39,1)			
Heterogenität^e p = 0,6180					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit DAS28-hsCRP < 2,6					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 eine Remission als unter Adalimumab. Der Unterschied war statistisch nicht signifikant. Von Baseline bis Woche 52 erreichte unter Baricitinib ein geringerer Anteil der Patienten eine Remission als unter Adalimumab. Der Unterschied war ebenfalls statistisch nicht signifikant.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine Remission als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit DAS28-hsCRP < 2,6 über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine Remission als unter Adalimumab. Der Unterschied war statistisch nicht signifikant (Tabelle 4-33).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2.3 *Simplified Disease Activity Index (SDAI) – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-34: Operationalisierung des SDAI

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Baseline (Woche 0), Woche 24, Woche 52</p> <ul style="list-style-type: none"> • Veränderung des SDAI von Baseline von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mLOCF) • Anteil Patienten mit SDAI ≤ 3,3 an Woche 24 und Woche 52 (NRI) <p><u>Definition:</u> Der SDAI ist ein zusammengesetztes Konstrukt (mit Akut-Phase-Protein), das die Krankheitsaktivität der RA erfasst und wie folgt zusammengesetzt ist:</p> <ul style="list-style-type: none"> • Anzahl der druckschmerzhaften Gelenker (basierend auf der Untersuchung von 28 Gelenken) • Anzahl der geschwollenen Gelenken (basierend auf der Untersuchung von 28 Gelenken) • PtGADA (VAS) • PhGADA (VAS) • hsCRP <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>	

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

Tabelle 4-35: Bewertung des Verzerrungspotenzials für SDAI in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	ja	niedrig
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>						

Begründen Sie für jede Studie die abschließende Einschätzung.

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

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt SDAI als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

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

Veränderung des SDAI

Tabelle 4-36: Ergebnisse für Veränderung des SDAI aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des SDAI								
Woche 24								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	240	41,04 (12,84)	236	13,57 (12,51)	236	-25,67 (1,048)	-2,39 (-5,14; 0,36); 0,089
Adalimumab	153	150	42,02 (13,77)	150	16,48 (14,87)	150	-23,28 (1,248)	
Subpopulation A3								
Baricitinib	170	168	38,40 (12,05)	168	13,02 (12,18)	168	-25,14 (1,155)	-1,43 (-4,39; 1,52); 0,341
Adalimumab	124	123	38,30 (12,30)	122	14,05 (13,19)	122	-23,71 (1,309)	
mITT								
Baricitinib	487	482	40,30 (12,66)	478	13,73 (12,47)	478	-25,75 (0,667)	-2,43 (-4,28; -0,59); 0,010
Adalimumab	330	326	40,12 (13,37)	324	16,04 (14,39)	324	-23,31 (0,782)	
Heterogenität^b p = 0,270								

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

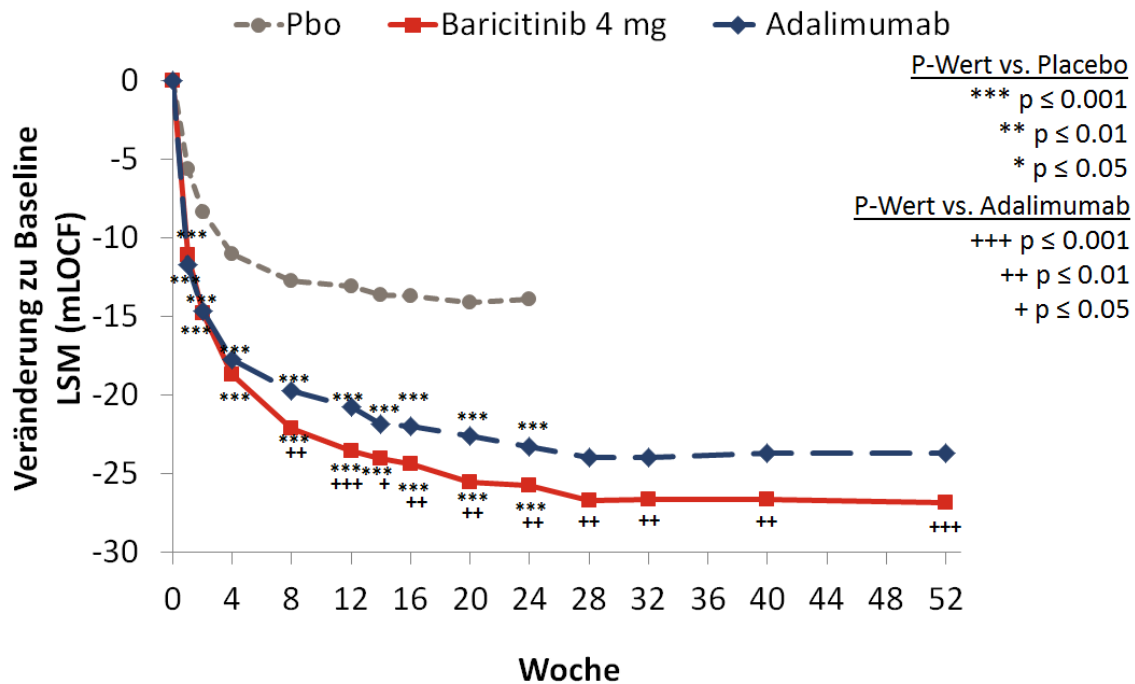
RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des SDAI								
Woche 52								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	240	41,04 (12,84)	236	12,43 (13,21)	236	-28,11 (1,167)	-2,72 (-5,51; 0,08); 0,057
Adalimumab	153	150	42,02 (13,77)	150	15,63 (15,86)	150	-25,39 (1,363)	
Subpopulation A3								
Baricitinib	170	168	38,40 (12,05)	168	11,75 (12,58)	168	-24,85 (1,272)	-2,54 (-5,48; 0,40); 0,090
Adalimumab	124	123	38,24 (12,33)	122	14,08 (14,28)	122	-22,31 (1,407)	
mITT								
Baricitinib	487	482	40,30 (12,66)	478	12,44 (13,04)	478	-26,86 (0,708)	-3,18 (-5,04; -1,32); < 0,001
Adalimumab	330	326	40,12 (13,37)	324	15,54 (15,57)	324	-23,68 (0,817)	
Heterogenität^b p = 0,980								
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.								
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.								
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.								
Quelle: [59].								

Für **Subpopulation A2** konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Der Unterschied war statistisch nicht signifikant.

Für **Subpopulation A3** konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung des SDAI über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

Für die **mITT Population** konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Von Baseline bis Woche 24 und Woche 52 war der Unterschied statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität (Tabelle 4-36).



Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [161].

Abbildung 5: Zeitlicher Verlauf des SDAI über 52 Wochen (mITT)

Abbildung 5 zeigt den zeitlichen Verlauf des SDAI über 52 Wochen für die Behandlungsarme Baricitinib, Adalimumab und Placebo. Baricitinib ist Adalimumab ab Woche 8 statistisch signifikant überlegen ($p < 0,01$). Im weiteren Behandlungsverlauf bleibt die statistische signifikante Überlegenheit von Baricitinib gegenüber Adalimumab bis zum Behandlungsende (Woche 52) bestehen.

Anteil Patienten mit SDAI \leq 3,3Tabelle 4-37: Ergebnisse für Anteil Patienten mit SDAI \leq 3,3 aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI)^d; p-Wert	RR^a (95% KI)^d; p-Wert	ARD^b (95% KI)^c; p-Wert^c
Anteil Patienten mit SDAI \leq 3,3					
Woche 24					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	45 (18,5)	1,38 (0,78; 2,45); 0,268	1,34 (0,85; 2,13); 0,210	3,49 (-3,99; 10,96); 0,413
Adalimumab	153	23 (15,0)			
Subpopulation A3					
Baricitinib	170	25 (14,7)	1,05 (0,54; 2,06); 0,877	1,02 (0,59; 1,78); 0,932	-0,62 (-8,90; 7,66); 1,000
Adalimumab	124	19 (15,3)			
mITT					
Baricitinib	487	78 (16,0)	1,25 (0,83; 1,89); 0,276	1,22 (0,86; 1,71); 0,261	2,38 (-2,55; 7,31); 0,371
Adalimumab	330	45 (13,6)			
Heterogenität^e p = 0,4740					
Woche 52					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	54 (22,2)	1,20 (0,72; 2,00); 0,496	1,16 (0,78; 1,71); 0,464	2,61 (-5,56; 10,79); 0,614
Adalimumab	153	30 (19,6)			
Subpopulation A3					
Baricitinib	170	40 (23,5)	1,63 (0,89; 2,98); 0,115	1,46 (0,90; 2,36); 0,125	7,40 (-1,69; 16,49); 0,143
Adalimumab	124	20 (16,1)			
mITT					
Baricitinib	487	110 (22,6)	1,34 (0,94; 1,91); 0,111	1,26 (0,96; 1,67); 0,101	4,71 (-0,85; 10,27); 0,113
Adalimumab	330	59 (17,9)			
Heterogenität^e p = 0,4194					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit SDAI ≤ 3,3					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine Remission als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Adalimumab von Baseline bis Woche 24 und ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 52 eine Remission. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit SDAI ≤ 3,3 über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine Remission als unter Adalimumab. Der Unterschied war statistisch nicht signifikant (Tabelle 4-37).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2.4 *Clinical Disease Activity Index (CDAI) – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-38: Operationalisierung des CDAI

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Baseline (Woche 0), Woche 24, Woche 52</p> <ul style="list-style-type: none"> • Veränderung des CDAI von Baseline von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mLOCF) • Anteil Patienten mit CDAI ≤ 10 an Woche 24 und Woche 52 (NRI) <p><u>Definition:</u> Der CDAI ist ein validiertes Konstrukt, mit dem das komplexe Krankheitsgeschehen der rheumatoiden Arthritis abgebildet werden soll. (ohne Akute-Phase-Protein). Dieser Score setzt sich aus den folgenden patientenrelevanten Einzelkomponenten zusammen:</p> <ul style="list-style-type: none"> • Anzahl der druckschmerzhaften Gelenke (basierend auf der Untersuchung von 28 Gelenken) • Anzahl der geschwollenen Gelenke (basierend auf der Untersuchung von 28 Gelenken) • PtGADA (VAS) • PhGADA (VAS) <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>	

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

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

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	ja	niedrig
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>						

Begründen Sie für jede Studie die abschließende Einschätzung.

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

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt CDAI als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

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

Veränderung des CDAI

Tabelle 4-40: Ergebnisse für Veränderung des CDAI aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des CDAI								
Woche 24								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	240	38,78 (12,20)	236	12,77 (12,08)	236	-24,16 (1,008)	-2,06 (-4,71; 0,58); 0,126
Adalimumab	153	150	39,71 (13,37)	150	15,37 (14,27)	150	-22,10 (1,199)	
Subpopulation A3								
Baricitinib	170	168	36,44 (11,34)	168	12,35 (11,51)	168	-23,92 (1,105)	-1,29 (-4,12; 1,54); 0,371
Adalimumab	124	123	36,28 (11,96)	122	13,23 (12,58)	122	-22,63 (1,253)	
mITT								
Baricitinib	487	482	38,06 (11,99)	478	13,03 (12,01)	478	-24,20 (0,638)	-2,06 (-3,83; -0,29); 0,023
Adalimumab	330	326	37,95 (12,95)	324	15,00 (13,68)	324	-22,15 (0,749)	
Heterogenität^b p = 0,334								

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

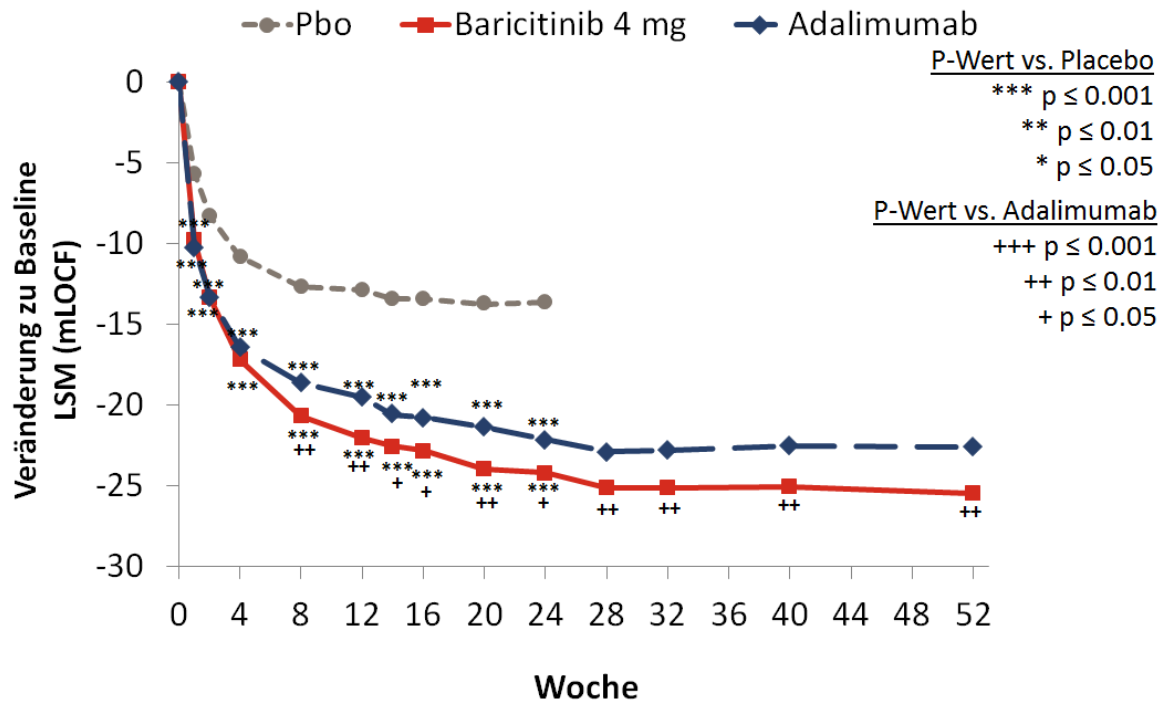
RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des CDAI								
Woche 52								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	240	38,78 (12,20)	236	11,55 (12,71)	236	-26,45 (1,116)	-2,70 (-5,37; -0,02); 0,048
Adalimumab	153	150	39,71 (13,37)	150	14,76 (15,06)	150	-23,75 (1,302)	
Subpopulation A3								
Baricitinib	170	168	36,44 (11,34)	168	10,98 (11,89)	168	-24,17 (1,206)	-2,07 (-4,86; 0,72); 0,145
Adalimumab	124	123	36,28 (11,96)	122	12,80 (13,42)	122	-22,10 (1,335)	
mITT								
Baricitinib	487	482	38,06 (11,99)	478	11,62 (12,56)	478	-25,46 (0,672)	-2,85 (-4,62; -1,09); 0,002
Adalimumab	330	326	37,95 (12,95)	324	14,44 (14,60)	324	-22,61 (0,775)	
Heterogenität^b p = 0,794								
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.								
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.								
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.								
Quelle: [59].								

Für **Subpopulation A2** konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Von Baseline bis Woche 52 war der Unterschied statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität.

Für **Subpopulation A3** konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung des CDAI über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

Für die **mITT Population** konnte unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab eine niedrigere Krankheitsaktivität beobachtet werden. Von Baseline bis Woche 24 und Woche 52 war der Unterschied statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität (Tabelle 4-40).



Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [161].

Abbildung 6: Zeitlicher Verlauf des CDAI über 52 Wochen (mITT)

Abbildung 6 zeigt den zeitlichen Verlauf des CDAI über 52 Wochen für die Behandlungsarme Baricitinib, Adalimumab und Placebo. Baricitinib ist Adalimumab ab Woche 8 statistisch signifikant überlegen ($p < 0,01$). Im weiteren Behandlungsverlauf bleibt die statistische signifikante Überlegenheit von Baricitinib gegenüber Adalimumab bis zum Behandlungsende (Woche 52) bestehen.

Anteil Patienten mit CDAI \leq 10

Tabelle 4-41: Ergebnisse für Anteil Patienten mit CDAI \leq 10 aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI)^d; p-Wert^c	RR^a (95% KI)^d; p-Wert^c	ARD^b (95% KI)^c; p-Wert^c
Anteil Patienten mit CDAI \leq 10					
Woche 24					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	125 (51,4)	1,18 (0,78; 1,79); 0,442	1,08 (0,88; 1,32); 0,458	3,73 (-6,38; 13,83); 0,536
Adalimumab	153	73 (47,7)			
Subpopulation A3					
Baricitinib	170	88 (51,8)	1,05 (0,65; 1,69); 0,846	1,01 (0,81; 1,25); 0,953	0,15 (-11,41; 11,72); 1,000
Adalimumab	124	64 (51,6)			
mITT					
Baricitinib	487	243 (49,9)	1,10 (0,83; 1,47); 0,508	1,04 (0,91; 1,20); 0,567	2,32 (-4,66; 9,30); 0,522
Adalimumab	330	157 (47,6)			
Heterogenität^d					p = 0,2619
Woche 52					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	142 (58,4)	1,42 (0,93; 2,17); 0,105	1,15 (0,96; 1,39); 0,130	8,76 (-1,29; 18,82); 0,097
Adalimumab	153	76 (49,7)			
Subpopulation A3					
Baricitinib	170	99 (58,2)	1,18 (0,73; 1,90); 0,494	1,05 (0,86; 1,29); 0,632	3,40 (-8,08; 14,87); 0,634
Adalimumab	124	68 (54,8)			
mITT					
Baricitinib	487	277 (56,9)	1,35 (1,01; 1,80); 0,041	1,14 (1,00; 1,29); 0,055	7,48 (0,52; 14,45); 0,038
Adalimumab	330	163 (49,4)			
Heterogenität^d					p = 0,6078

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert ^c	RR ^a (95% KI) ^d ; p-Wert ^c	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit CDAI ≤ 10					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine geringe Krankheitsaktivität als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine geringe Krankheitsaktivität als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit CDAI ≤ 10,0 über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine geringe Krankheitsaktivität als unter Adalimumab. Der Unterschied war statistisch nicht signifikant (Tabelle 4-41).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2.5 EULAR Responsekriterien („good Response“, DAS28-basiert) – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-42: Operationalisierung der EULAR Responsekriterien („good Response“, DAS28-basiert)

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Woche 24, Woche 52</p> <ul style="list-style-type: none"> Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert) an Woche 24 und Woche 52 (NRI) <p><u>Definition:</u> Mit Hilfe der validierten EULAR Responsekriterien wird das Therapieansprechen eines Patienten auf Grundlage der Veränderung und des Post-Baseline Levels des DAS28 beurteilt und klassifiziert einen Patienten somit als Responder oder Non-Responder. Die EULAR „good Response“ (DAS28-basiert) wird folgt kategorisiert:</p> <ul style="list-style-type: none"> Verbesserung DAS28-hsCRP von Baseline > 1,2 mit Post-Baseline Level DAS28-hsCRP ≤ 3,2 <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>	

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

Tabelle 4-43: Bewertung des Verzerrungspotenzials für EULAR Responsekriterien („good Response“, DAS28-basiert) in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	ja	niedrig
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>						

Begründen Sie für jede Studie die abschließende Einschätzung.

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

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt EULAR Responsekriterien („good Response“, DAS28-basiert) als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Stellen Sie die Ergebnisse für den Endpunkt EULAR Responsekriterien („good Response“, DAS28-basiert) für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Tabelle 4-44: Ergebnisse für Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert) aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	124 (51,0)	1,16 (0,77; 1,75); 0,484	1,09 (0,89; 1,33); 0,395	3,32 (-6,79; 13,42); 0,537
Adalimumab	153	73 (47,7)			
Subpopulation A3					
Baricitinib	170	93 (54,7)	1,18 (0,74; 1,89); 0,482	1,06 (0,85; 1,32); 0,609	3,90 (-7,65; 15,45); 0,555
Adalimumab	124	63 (50,8)			
mITT					
Baricitinib	487	250 (51,3)	1,19 (0,89; 1,58); 0,240	1,07 (0,93; 1,23); 0,323	4,06 (-2,92; 11,04); 0,285
Adalimumab	330	156 (47,3)			
Heterogenität^e p = 0,2775					
Woche 52					
Subpopulation A1 Nicht zutreffend,					
Subpopulation A2					
Baricitinib	243	137 (56,4)	1,30 (0,86; 1,97); 0,215	1,12 (0,93; 1,36); 0,234	6,71 (-3,38; 16,79); 0,215
Adalimumab	153	76 (49,7)			
Subpopulation A3					
Baricitinib	170	92 (54,1)	1,20 (0,75; 1,92); 0,449	1,08 (0,86; 1,36); 0,494	4,12 (-7,44; 15,67); 0,555
Adalimumab	124	62 (50,0)			

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)					
mITT					
Baricitinib	487	268 (55,0)	1,36 (1,02; 1,81); 0,034	1,15 (1,00; 1,31); 0,051	7,76 (0,79; 14,72); 0,032
Adalimumab	330	156 (47,3)			
Heterogenität^e					p = 0,9020
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 ein gutes Therapieansprechen als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 ein gutes Therapieansprechen als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit „good Response“ über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 ein gutes Therapieansprechen als unter Adalimumab. Der Unterschied war statistisch nicht signifikant (Tabelle 4-44).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2.6 Patient Assessment of Pain (PAP) Visual Analog Scale (VAS) – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-45: Operationalisierung von PAP (VAS)

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Baseline (Woche 0), Woche 24, Woche 52</p> <ul style="list-style-type: none"> • Veränderung des PAP (VAS) von Baseline von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mLOCF) • Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm von Baseline bis Woche 24 und Woche 52 (NRI) • Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm von Baseline bis Woche 24 und Woche 52 (NRI) <p><u>Definition:</u> Mittels visueller Analogskala (VAS; 0 - 100 mm) wird der Patient gebeten seinen Schmerz zu beurteilen, wobei ein Wert von 0 mm bedeutet, dass der Patient keine Schmerzen bzw. ein Wert von 100 mm bedeutet, dass der Patient massive/schwere Schmerzen hat. Eine Abnahme des Wertes spiegelt somit eine Verbesserung wider.</p> <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>	

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

Tabelle 4-46: Bewertung des Verzerrungspotenzials für PAP (VAS) in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	ja	niedrig
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>						

Begründen Sie für jede Studie die abschließende Einschätzung.

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

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt PAP (VAS) als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

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

Veränderung des PAP (VAS)

Tabelle 4-47: Ergebnisse für Veränderung des PAP (VAS) aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des PAP (VAS)								
Woche 24								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	242	64,0 (22,5)	238	29,1 (24,6)	238	-32,4 (1,85)	-4,4 (-9,2; 0,5); 0,076
Adalimumab	153	152	62,9 (23,2)	152	33,1 (25,8)	152	-28,0 (2,20)	
Subpopulation A3								
Baricitinib	170	170	58,5 (21,4)	170	24,1 (20,3)	170	-34,4 (1,94)	-3,9 (-8,9; 1,0); 0,121
Adalimumab	124	124	58,9 (22,0)	123	28,1 (22,1)	123	-30,5 (2,21)	
mITT								
Baricitinib	487	486	61,8 (21,8)	482	27,3 (23,1)	482	-33,6 (1,15)	-4,8 (-8,0; -1,6); 0,003
Adalimumab	330	329	61,0 (22,7)	327	31,8 (24,5)	327	-28,8 (1,35)	
Heterogenität^{a,b}								p = 0,003

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

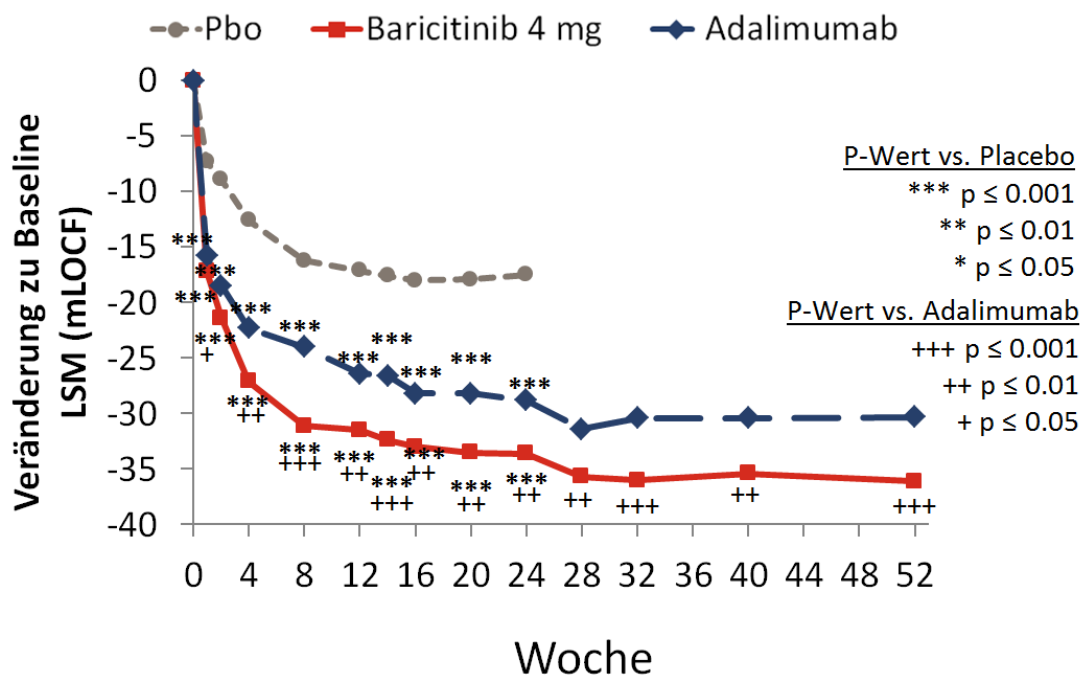
RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des PAP (VAS)								
Woche 52								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	242	64,0 (22,5)	238	27,7 (24,7)	238	-35,2 (2,08)	-4,6 (-9,6; 0,3); 0,067
Adalimumab	153	152	62,9 (23,2)	152	31,6 (25,9)	152	-30,6 (2,42)	
Subpopulation A3								
Baricitinib	170	170	58,5 (21,4)	170	20,9 (20,7)	170	-35,3 (2,20)	-5,9 (-11,0; -0,8); 0,023
Adalimumab	124	124	58,9 (22,0)	123	27,0 (24,6)	123	-29,4 (2,45)	
mITT								
Baricitinib	487	486	61,8 (21,8)	482	25,1 (23,4)	482	-36,1 (1,24)	-5,9 (-9,1; -2,6); < 0,001
Adalimumab	330	329	61,0 (22,7)	327	30,6 (25,7)	327	-30,3 (1,43)	
Heterogenität^{a,b} p = 0,661								
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.								
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.								
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.								
Quelle: [59].								

In **Subpopulation A2** wurde von den Patienten von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab weniger Schmerzen berichtet. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** wurde von den Patienten von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab weniger Schmerzen berichtet. Dieser Unterschied war zu Woche 52, jedoch nicht zu Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen.

An Woche 24 liegt der p-Wert des Heterogenitätstest für den Endpunkt Veränderung des PAP (VAS) unter 0,05, jedoch zeigen die Punktschätzer in die gleiche Richtung, mit sich weitestgehend überlappenden Konfidenzintervallen, und die Unterschiede zwischen den Subpopulationen sind medizinisch irrelevant. An Woche 52 ist der Heterogenitätstest unauffällig. Somit wird nicht von einer Modifikation des Effekts zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** wurde von den Patienten von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab statistisch signifikant weniger Schmerzen berichtet. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen (Tabelle 4-47).



Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [161].

Abbildung 7: Zeitlicher Verlauf des PAP (VAS) über 52 Wochen (mITT)

Abbildung 7 zeigt den zeitlichen Verlauf des PAP (VAS) über 52 Wochen für die Behandlungsarme Baricitinib, Adalimumab und Placebo. Baricitinib ist Adalimumab ab Woche 2 statistisch signifikant überlegen ($p < 0,05$). Im weiteren Behandlungsverlauf bleibt die statistische signifikante Überlegenheit von Baricitinib gegenüber Adalimumab bis zum Behandlungsende (Woche 52) bestehen.

Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm

Tabelle 4-48: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI)^d; p-Wert	RR^a (95% KI)^d; p-Wert	ARD^b (95% KI)^c; p-Wert^c
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm					
Woche 24					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	196 (82,0)	1,27 (0,76; 2,13); 0,358	1,06 (0,96; 1,17); 0,273	4,88 (-3,36; 13,13); 0,246
Adalimumab	153	118 (77,1)			
Subpopulation A3					
Baricitinib	170	136 (80,0)	1,12 (0,62; 2,00); 0,710	1,03 (0,92; 1,16); 0,621	1,14 (-8,25; 10,53); 0,884
Adalimumab	124	97 (78,9)			
mITT					
Baricitinib	487	395 (81,8)	1,30 (0,92; 1,85); 0,137	1,06 (0,99; 1,14); 0,111	4,65 (-1,06; 10,35); 0,109
Adalimumab	330	253 (77,1)			
Heterogenität^e					p = 0,4414
Woche 52					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	196 (82,0)	1,65 (1,00; 2,73); 0,051	1,14 (1,02; 1,27); 0,018	8,81 (0,26; 17,35); 0,044
Adalimumab	153	112 (73,2)			
Subpopulation A3					
Baricitinib	170	143 (84,1)	1,91 (1,04; 3,50); 0,038	1,14 (1,01; 1,28); 0,031	8,51 (-0,86; 17,88); 0,075
Adalimumab	124	93 (75,6)			
mITT					
Baricitinib	487	402 (83,2)	1,79 (1,26; 2,53); 0,001	1,13 (1,05; 1,22); 0,001	9,75 (3,93; 15,58); 0,001
Adalimumab	330	241 (73,5)			
Heterogenität^e					p = 0,9493

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 10 mm					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine klinisch relevante Schmerzreduktion.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine klinisch relevante Schmerzreduktion.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 10 mm über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine klinisch relevante Schmerzreduktion (Tabelle 4-48).

Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 20 mmTabelle 4-49: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 20 mm aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert ^c	RR ^a (95% KI) ^d ; p-Wert ^c	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 20 mm					
Woche 24					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	168 (70,3)	1,34 (0,86; 2,08); 0,194	1,08 (0,94; 1,25); 0,267	7,55 (-2,06; 17,15); 0,124
Adalimumab	153	96 (62,7)			
Subpopulation A3					
Baricitinib	170	119 (70,0)	1,20 (0,72; 1,99); 0,489	1,08 (0,92; 1,26); 0,344	2,52 (-8,25; 13,29); 0,702
Adalimumab	123	83 (67,5)			
mITT					
Baricitinib	487	336 (69,6)	1,31 (0,97; 1,77); 0,076	1,09 (0,99; 1,20); 0,090	6,15 (-0,48; 12,78); 0,068
Adalimumab	330	208 (63,4)			
Heterogenität^e					p = 0,1363
Woche 52					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	168 (70,3)	1,33 (0,85; 2,07); 0,207	1,12 (0,97; 1,29); 0,118	6,89 (-2,69; 16,48); 0,184
Adalimumab	153	97 (63,4)			
Subpopulation A3					
Baricitinib	170	129 (75,9)	2,00 (1,18; 3,40); 0,010	1,23 (1,05; 1,44); 0,012	11,65 (1,02; 22,29); 0,037
Adalimumab	123	79 (64,2)			
mITT					
Baricitinib	487	354 (73,3)	1,66 (1,22; 2,25); 0,001	1,18 (1,07; 1,30); 0,001	10,79 (4,23; 17,35); 0,001
Adalimumab	330	205 (62,5)			
Heterogenität^e					p = 0,3628

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert ^c	RR ^a (95% KI) ^d ; p-Wert ^c	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 20 mm					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine klinisch relevante Schmerzreduktion.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 20 mm über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine klinisch relevante Schmerzreduktion (Tabelle 4-49).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2.7 Patient's Global Assessment of Disease Activity (PtGADA) (VAS) – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-50: Operationalisierung des PtGADA (VAS)

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Baseline (Woche 0), Woche 24, Woche 52</p> <ul style="list-style-type: none"> • Veränderung des PtGADA (VAS) von Baseline von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mLOCF) • Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm Woche 24 und Woche 52 (NRI) • Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm von Baseline bis Woche 24 und Woche 52 (NRI) <p><u>Definition:</u> Spiegelt die aus Sicht des Patienten wahrgenommene Krankheitsaktivität wider. Mittels VAS von 0 - 100 mm kategorisiert der Patient die Schwere der Symptome, wobei ein Wert von 0 mm bedeutet, dass der Patient keine Symptome hat bzw. ein Wert von 100 mm bedeutet, dass der Patient massive/schweren Symptome verspürt. Es geringer Wert zeigt eine niedrige Krankheitsaktivität an. Im Studienverlauf spiegelt eine Abnahme des Wertes eine Verbesserung der Krankheitsaktivität und somit der Lebensqualität wider.</p> <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>	

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

Tabelle 4-51: Bewertung des Verzerrungspotenzials für PtGADA (VAS) in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	ja	niedrig
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>						

Begründen Sie für jede Studie die abschließende Einschätzung.

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

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt PtGADA (VAS) als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

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

Veränderung des PtGADA (VAS)

Tabelle 4-52: Ergebnisse für Veränderung des PtGADA (VAS) aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des PtGADA (VAS)								
Woche 24								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	242	64,8 (22,0)	238	30,2 (24,0)	238	-32,7 (1,82)	-4,7 (-9,4; 0,1); 0,054
Adalimumab	153	152	64,3 (21,7)	152	34,6 (25,3)	152	-28,1 (2,16)	
Subpopulation A3								
Baricitinib	170	170	61,1 (20,9)	170	26,5 (20,1)	170	-34,6 (24,6)	-2,4 (-7,2; 2,5); 0,343
Adalimumab	124	124	62,2 (20,5)	123	29,0 (22,6)	123	-33,4 (25,3)	
mITT								
Baricitinib	487	486	63,1 (21,2)	482	29,0 (22,5)	482	-33,1 (1,13)	-4,0 (-7,1; -0,8); 0,013
Adalimumab	330	329	63,7 (21,2)	327	33,2 (24,3)	327	-29,1 (1,33)	
Heterogenität^{a,b}								p = 0,007

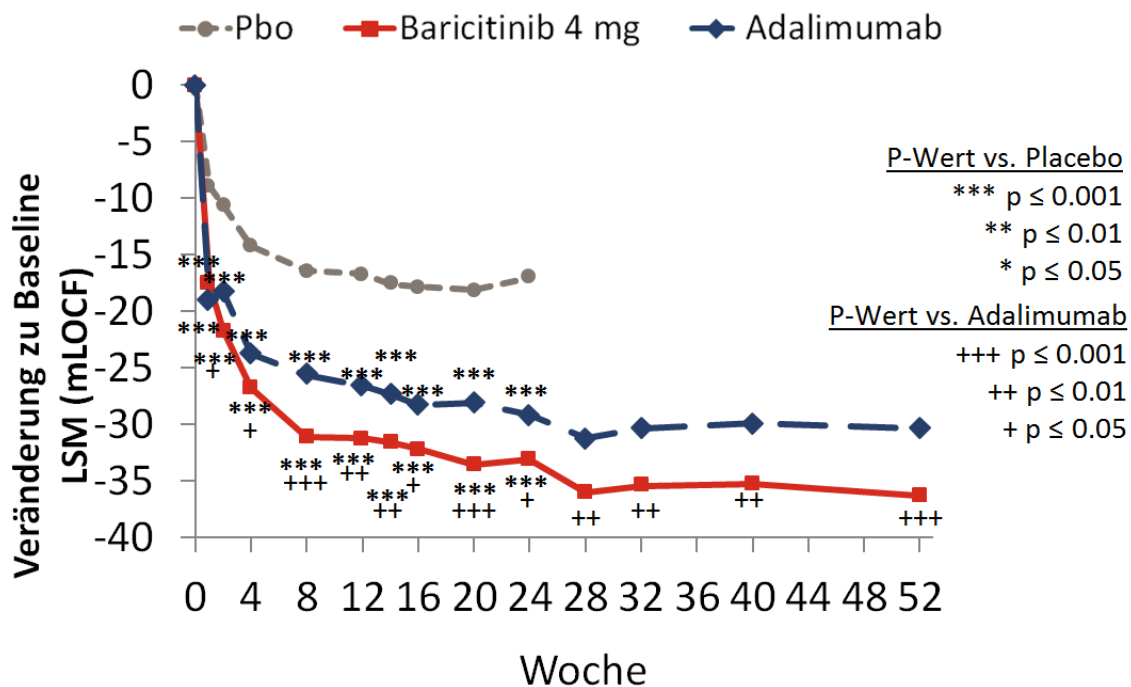
RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des PtGADA (VAS)								
Woche 52								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	242	64,8 (22,0)	238	28,3 (23,4)	238	-36,2 (2,01)	-6,0 (-10,8; -1,2); 0,015
Adalimumab	153	152	64,3 (21,7)	152	33,9 (25,9)	152	-30,2 (2,34)	
Subpopulation A3								
Baricitinib	170	170	61,1 (20,9)	170	22,8 (21,0)	170	-36,7 (2,23)	-5,2 (-10,4; -0,1); 0,046
Adalimumab	124	124	62,2 (20,5)	123	28,2 (24,6)	123	-31,5 (2,48)	
mITT								
Baricitinib	487	486	63,1 (21,2)	482	26,4 (22,6)	482	-36,3 (1,23)	-6,0 (-9,2; -2,8); < 0,001
Adalimumab	330	329	63,7 (21,2)	327	32,6 (25,9)	327	-30,3 (1,42)	
Heterogenität^{a,b} p = 0,899								
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.								
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.								
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.								
Quelle: [59].								

In **Subpopulation A2** war die Krankheitsaktivität aus Sicht der Patienten von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab niedriger. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität.

In **Subpopulation A3** war die Krankheitsaktivität aus Sicht der Patienten von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab niedriger. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität.

An Woche 24 liegt der p-Wert des Heterogenitätstest für den Endpunkt Veränderung des PtGADA (VAS) unter 0,05, jedoch zeigen die Punktschätzer in die gleiche Richtung, mit sich weitestgehend überlappenden Konfidenzintervallen, und die Unterschiede zwischen den Subpopulationen sind medizinisch irrelevant. An Woche 52 ist der Heterogenitätstest unauffällig. Somit wird nicht von einer Modifikation des Effekts zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** war die Krankheitsaktivität aus Sicht der Patienten von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab niedriger. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität (Tabelle 4-52).



Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [161].

Abbildung 8: Zeitlicher Verlauf des PtGADA (VAS) über 52 Wochen (mITT)

Abbildung 8 zeigt den zeitlichen Verlauf des PtGADA (VAS) über 52 Wochen für die Behandlungsarme Baricitinib, Adalimumab und Placebo. Baricitinib ist Adalimumab ab Woche 2 statistisch signifikant überlegen ($p < 0,05$). Im weiteren Behandlungsverlauf bleibt die statistische signifikante Überlegenheit von Baricitinib gegenüber Adalimumab bis zum Behandlungsende (Woche 52) bestehen.

Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 10 mm

Tabelle 4-53: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 10 mm aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI)^d; p-Wert	RR^a (95% KI)^d; p-Wert	ARD^b (95% KI)^c; p-Wert^c
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 10 mm					
Woche 24					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	193 (80,8)	1,14 (0,69; 1,90); 0,605	1,02 (0,92; 1,13); 0,727	3,63 (-4,69; 11,95); 0,443
Adalimumab	153	118 (77,1)			
Subpopulation A3					
Baricitinib	170	145 (85,3)	1,27 (0,67; 2,41); 0,469	1,04 (0,95; 1,14); 0,401	2,37 (-6,15; 10,89); 0,627
Adalimumab	124	102 (82,9)			
mITT					
Baricitinib	487	396 (82,0)	1,20 (0,84; 1,72); 0,306	1,04 (0,97; 1,11); 0,298	3,33 (-2,28; 8,93); 0,241
Adalimumab	330	258 (78,7)			
Heterogenität^e					
p = 0,1036					
Woche 52					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	192 (80,3)	1,33 (0,81; 2,20); 0,259	1,05 (0,94; 1,17); 0,396	5,82 (-2,72; 14,37); 0,211
Adalimumab	153	114 (74,5)			
Subpopulation A3					
Baricitinib	170	144 (84,7)	1,49 (0,79; 2,81); 0,222	1,05 (0,95; 1,15); 0,365	4,22 (-4,63; 13,07); 0,350
Adalimumab	124	99 (80,5)			
mITT					
Baricitinib	487	396 (82,0)	1,38 (0,97; 1,97); 0,071	1,05 (0,98; 1,13); 0,167	5,46 (-0,26; 11,19); 0,062
Adalimumab	330	251 (76,5)			
Heterogenität^e					
p = 0,9396					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 10 mm					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 10 mm über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Der Unterschied war statistisch nicht signifikant (Tabelle 4-53).

Anteil Patienten mit einer Verbesserung der PtGADA (VAS) \geq 20 mmTabelle 4-54: Ergebnisse für Anteil Patienten mit einer Verbesserung der PtGADA (VAS) \geq 20 mm aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) \geq 20 mm					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	169 (70,7)	1,27 (0,82; 1,98); 0,291	1,05 (0,91; 1,21); 0,483	6,01 (-3,51; 15,53); 0,222
Adalimumab	153	99 (64,7)			
Subpopulation A3					
Baricitinib	170	125 (73,5)	1,09 (0,64; 1,87), 0,740	1,05 (0,92; 1,20); 0,445	0,36 (-9,90; 10,62); 1,000
Adalimumab	124	90 (73,2)			
mITT					
Baricitinib	487	343 (71,0)	1,26 (0,93; 1,71); 0,140	1,08 (0,98; 1,18); 0,124	5,16 (-1,37; 11,70); 0,123
Adalimumab	330	216 (65,9)			
Heterogenität^e p = 0,0278					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	169 (70,7)	1,35 (0,86; 2,11); 0,187	1,09 (0,94; 1,25); 0,243	7,31 (-2,26; 16,88); 0,150
Adalimumab	153	97 (63,4)			
Subpopulation A3					
Baricitinib	170	133 (78,2)	1,51 (0,86; 2,63); 0,151	1,08 (0,95; 1,23); 0,238	5,88 (-4,17; 15,92); 0,270
Adalimumab	123	89 (72,4)			
mITT					
Baricitinib	487	352 (72,9)	1,43 (1,05; 1,95); 0,023	1,10 (1,00; 1,21); 0,043	7,63 (1,13; 14,14); 0,024
Adalimumab	330	214 (65,2)			
Heterogenität^e p = 0,9685					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) \geq 20 mm					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

An Woche 24 liegt der p-Wert des Heterogenitätstest für den Endpunkt Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 20 mm unter 0,05, jedoch zeigen die Punktschätzer in die gleiche Richtung, mit sich weitestgehend überlappenden Konfidenzintervallen, und die Unterschiede zwischen den Subpopulationen sind medizinisch irrelevant. An Woche 52 ist der Heterogenitätstest unauffällig. Somit wird nicht von einer Modifikation des Effekts zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant (Tabelle 4-54).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden.

Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2.8 Health Assessment Questionnaire-Disability Index (HAQ-DI) – RCT

Die Ergebnisdarstellung für jeden Endpunkt umfasst 3 Abschnitte. Zunächst soll für jede Studie das Verzerrungspotenzial auf Endpunktbene 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-55: Operationalisierung des HAQ-DI

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Baseline (Woche 0), Woche 24, Woche 52</p> <ul style="list-style-type: none"> • Veränderung des HAQ-DI von Baseline von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mBOCF/mLOCF; MMRM) • Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$ von Baseline bis Woche 24 und Woche 52 (NRI) • Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$ von Baseline bis Woche 24 und Woche 52 (NRI) <p><u>Definition:</u> Der HAQ-DI ist ein Selbstbeurteilungsinstrument, der die krankheitsassoziierten Einschränkungen der physischen Funktionalität der Patienten erfasst. Er besteht aus acht Domänen und die Bewertung erfolgt durch eine 4-Punkte-Likert-Skala (0 = ohne Schwierigkeiten, 1 = mit einigen Schwierigkeiten, 2 = mit vielen Schwierigkeiten, 3 = unfähig zur Durchführung). Über die Mittelwerte der einzelnen Domänen wird der <i>Functional Disability Index</i> berechnet. Ein geringer Wert zeigt eine bessere physische Funktionalität an. Eine Abnahme des Wertes spiegelt eine Verbesserung der physischen Funktionalität wider.</p> <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>	

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

Tabelle 4-56: Bewertung des Verzerrungspotenzials HAQ-DI in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	ja	niedrig
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>						

Begründen Sie für jede Studie die abschließende Einschätzung.

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

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt HAQ-DI als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

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

Veränderung des HAQ-DI

Tabelle 4-57: Ergebnisse für Veränderung des HAQ-DI aus RCT mit dem zu bewertenden Arzneimittel (mBOCF)

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des HAQ-DI								
Woche 24								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	242	1,68 (0,71)	241	0,88 (0,75)	243	-0,76 (0,045)	-0,12 (-0,23; 0,00); 0,050
Adalimumab	153	152	1,63 (0,69)	152	0,97 (0,74)	153	-0,65 (0,053)	
Subpopulation A3								
Baricitinib	170	170	1,41 (0,64)	170	0,75 (0,67)	170	-0,70 (0,050)	-0,07 (-0,20; 0,05); 0,252
Adalimumab	124	124	1,60 (0,72)	123	0,96 (0,72)	124	-0,63 (0,057)	
mITT								
Baricitinib	487	486	1,57 (0,68)	485	0,84 (0,70)	487	-0,72 (0,028)	-0,11 (-0,18; -0,03); 0,008
Adalimumab	330	329	1,59 (0,70)	327	0,96 (0,73)	330	-0,61 (0,033)	
Heterogenität^{a,b}								p = 0,394

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

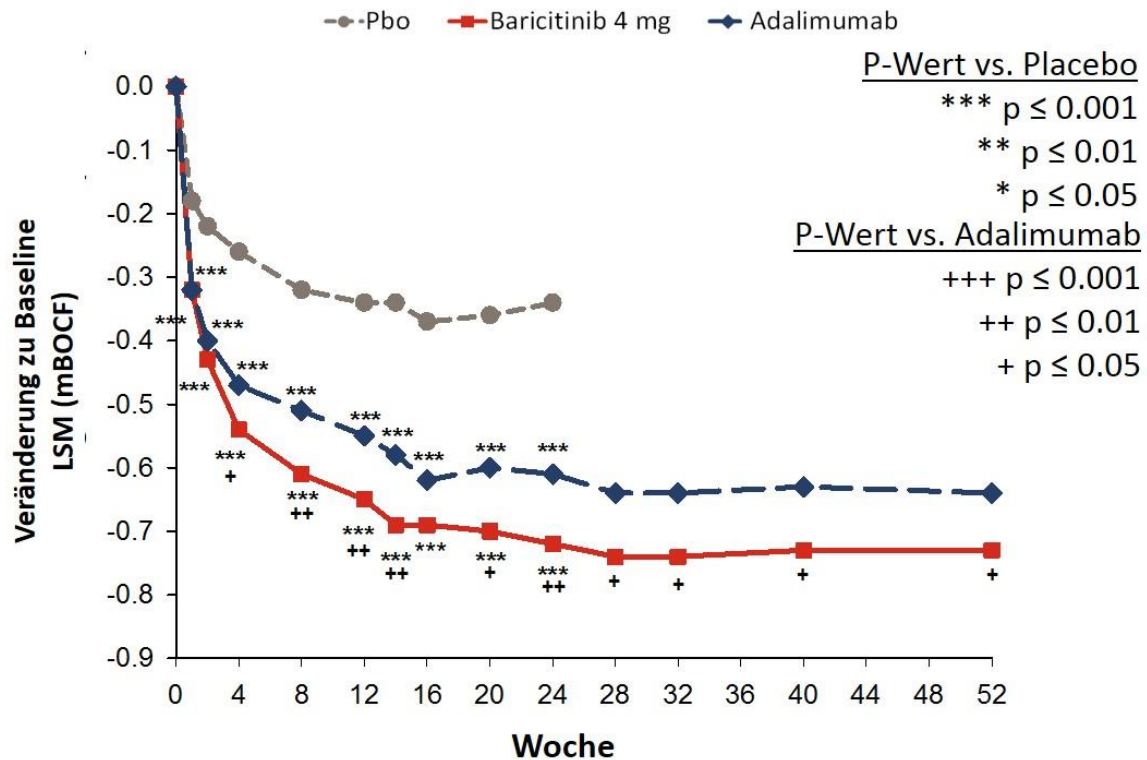
RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des HAQ-DI								
Woche 52								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	242	1,68 (0,71)	241	0,93 (0,75)	243	-0,76 (0,052)	-0,08 (-0,20; 0,05); 0,220
Adalimumab	153	153	1,63 (0,69)	152	0,97 (0,76)	153	-0,69 (0,061)	
Subpopulation A3								
Baricitinib	170	170	1,41 (0,64)	170	0,73 (0,68)	170	-0,73 (0,059)	-0,08 (-0,22; 0,06); 0,262
Adalimumab	124	124	1,60 (0,72)	123	0,91 (0,75)	124	-0,65 (0,066)	
mITT								
Baricitinib	487	486	1,57 (0,68)	485	0,85 (0,71)	487	-0,73 (0,031)	-0,09 (-0,17; -0,00); 0,039
Adalimumab	330	329	1,59 (0,70)	327	0,95 (0,76)	330	-0,64 (0,036)	
Heterogenität^{a,b} p = 0,996								
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mBOCF) berechnet.								
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mBOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.								
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.								
Quelle: [59].								

Für **Subpopulation A2** (mBOCF) waren die krankheitsassoziierten Einschränkungen der physischen Funktionalität von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab niedriger. Der Unterschied war statistisch nicht signifikant.

Für **Subpopulation A3** (mBOCF) waren die krankheitsassoziierten Einschränkungen der physischen Funktionalität von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab niedriger. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung des HAQ-DI über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

Für die **mITT Population** (mBOCF) waren die krankheitsassoziierten Einschränkungen der physischen Funktionalität von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab niedriger. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant (Tabelle 4-60).



Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [161].

Abbildung 9: Zeitlicher Verlauf des HAQ-DI (mBOCF) über 52 Wochen (mITT)

Abbildung 9 zeigt den zeitlichen Verlauf des HAQ-DI über 52 Wochen für die Behandlungsarme Baricitinib, Adalimumab und Placebo. Baricitinib ist Adalimumab ab Woche 4 statistisch signifikant überlegen ($p < 0,05$). Im weiteren Behandlungsverlauf bleibt die statistische signifikante Überlegenheit von Baricitinib gegenüber Adalimumab bis zum Behandlungsende (Woche 52) bestehen.

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

Tabelle 4-58: Ergebnisse für Veränderung des HAQ-DI aus RCT mit dem zu bewertenden Arzneimittel (Sensitivitätsanalyse - mLOCF)

RA-BEAM (JADV)	N	Baseline		Woche 52 MW (SD)		Veränderung zu Baseline		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des HAQ-DI								
Woche 52								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	242	1,68 (0,71)	238	0,89 (0,73)	238	-0,83 (0,051)	-0,10 (-0,22; 0,02); 0,111
Adalimumab	153	152	1,63 (0,69)	152	0,94 (0,74)	152	-0,73 (0,059)	
Subpopulation A3								
Baricitinib	170	170	1,41 (0,64)	170	0,68 (0,65)	170	-0,78 (0,058)	-0,11 (-0,24; 0,02); 0,109
Adalimumab	124	124	1,60 (0,72)	123	0,90 (0,74)	123	-0,67 (0,064)	
mITT								
Baricitinib	487	486	1,57 (0,68)	482	0,81 (0,68)	482	-0,77 (0,031)	-0,11 (-0,19; -0,03); 0,007
Adalimumab	330	329	1,59 (0,70)	327	0,93 (0,74)	327	-0,66 (0,035)	
Heterogenität^{a,b} p = 0,927								
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.								
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.								
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.								
Quelle: [59].								

Für **Subpopulation A2** (mLOCF) waren die krankheitsassoziierten Einschränkungen der physischen Funktionalität von Baseline bis Woche 52 unter Baricitinib im Vergleich zu Adalimumab niedriger. Der Unterschied war statistisch nicht signifikant.

Für **Subpopulation A3** (mLOCF) waren die krankheitsassoziierten Einschränkungen der physischen Funktionalität von Baseline bis Woche 52 unter Baricitinib im Vergleich zu Adalimumab niedriger. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung des HAQ-DI (mLOCF) über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

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

Für die **mITT Population** (mLOCF) waren die krankheitsassoziierten Einschränkungen der physischen Funktionalität von Baseline bis Woche 52 unter Baricitinib im Vergleich zu Adalimumab niedriger. Der Unterschied war statistisch signifikant (Tabelle 4-60).

Tabelle 4-59: Ergebnisse für Veränderung des HAQ-DI aus RCT mit dem zu bewertenden Arzneimittel (Sensitivitätsanalyse - MMRM)

RA-BEAM (JADV)	N	n	Veränderung zu Baseline	Baricitinib vs. Adalimumab
			LSM (SE)	LSMD ^a (95% KI); p-Wert
Veränderung des HAQ-DI				
Woche 24				
Subpopulation A1			Nicht zutreffend.	
Subpopulation A2				
Baricitinib	243	214	-0,81 (0,04)	-0,11 (-0,2; 0,0); 0,052
Adalimumab	153	126	-0,69 (0,05)	
Subpopulation A3				
Baricitinib	170	152	-0,74 (0,05)	-0,11 (-0,2; -0,0); 0,144
Adalimumab	124	105	-0,64 (0,05)	
mITT				
Baricitinib	487	430	-0,77 (0,02)	-0,12 (-0,2; -0,0); 0,002
Adalimumab	330	272	-0,66 (0,03)	
Woche 52				
Subpopulation A1			Nicht zutreffend.	
Subpopulation A2				
Baricitinib	243	198	-0,79 (0,04)	-0,06 (-0,2; 0,1); 0,354
Adalimumab	153	110	-0,73 (0,05)	
Subpopulation A3				
Baricitinib	170	141	-0,78 (0,05)	-0,09 (-0,2; 0,01); 0,216
Adalimumab	124	96	-0,69 (0,06)	
mITT				
Baricitinib	487	397	-0,78 (0,03)	-0,08 (-0,2; -0,0); 0,044
Adalimumab	330	241	-0,70 (0,03)	
Heterogenität^b				p = 0,360

RA-BEAM (JADV)	N	n	Veränderung zu Baseline	Baricitinib vs. Adalimumab
			LSM (SE)	LSMD ^a (95% KI); p-Wert
Veränderung des HAQ-DI				
<p>a: p-Werte, LSM, SE, LSMD und 95%-Konfidenzintervall wurden mittels eines MMRM-Modells berechnet.</p> <p>b: Interaktions-p-Wert mittels eines MMRM-Modells berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>				

Für **Subpopulation A2** (MMRM) war unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab die krankheitsassoziierten Einschränkungen der physischen Funktionalität niedriger. Der Unterschied war statistisch nicht signifikant.

Für **Subpopulation A3** (MMRM) war unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab die krankheitsassoziierten Einschränkungen der physischen Funktionalität niedriger. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung des HAQ-DI (MMRM) über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

Für die **mITT Population** (MMRM) war unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab die krankheitsassoziierten Einschränkungen der physischen Funktionalität statistisch signifikant niedriger. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung der Krankheitsaktivität (Tabelle 4-59).

Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$

Tabelle 4-60: Ergebnisse für Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$ aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI)^d; p-Wert	RR^a (95% KI)^d; p-Wert	ARD^b (95% KI)^c; p-Wert^c
Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	179 (73,7)	1,40 (0,90; 2,20); 0,138	1,10 (0,96; 1,26); 0,172	8,96 (-0,42; 18,34); 0,071
Adalimumab	153	99 (64,7)			
Subpopulation A3					
Baricitinib	170	122 (71,8)	1,23 (0,73; 2,05); 0,436	1,03 (0,89; 1,20); 0,682	3,22 (-7,39; 13,83); 0,605
Adalimumab	124	85 (68,5)			
mITT					
Baricitinib	487	353 (72,5)	1,48 (1,09; 2,00); 0,012	1,12 (1,02; 1,23); 0,021	8,55 (2,02; 15,07); 0,011
Adalimumab	330	211 (63,9)			
Heterogenität^e p = 0,8585					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	163 (67,1)	1,42 (0,92; 2,19); 0,116	1,11 (0,95; 1,29); 0,180	8,25 (-1,53; 18,04); 0,107
Adalimumab	153	90 (58,8)			
Subpopulation A3					
Baricitinib	170	116 (68,2)	1,39 (0,84; 2,31); 0,198	1,08 (0,91; 1,29); 0,371	5,33 (-5,68; 16,34); 0,383
Adalimumab	124	78 (62,9)			
mITT					
Baricitinib	487	329 (67,6)	1,50 (1,12; 2,02); 0,007	1,14 (1,03; 1,27); 0,016	9,37 (2,62; 16,13); 0,008
Adalimumab	330	192 (58,2)			
Heterogenität^e p = 0,9972					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der körperlichen Funktionen als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der körperlichen Funktionen als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$ über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der körperlichen Funktionen, als unter Adalimumab. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung der physischen Funktionalität (Tabelle 4-60).

Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$

Tabelle 4-61: Ergebnisse für Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$ aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI)^d; p-Wert	RR^a (95% KI)^d; p-Wert	ARD^b (95% KI)^c; p-Wert^c
Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$					
Woche 24					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	167 (68,7)	1,33 (0,86; 2,05); 0,197	1,08 (0,93; 1,25); 0,309	7,94 (-1,75; 17,63); 0,128
Adalimumab	153	93 (60,8)			
Subpopulation A3					
Baricitinib	170	109 (64,1)	1,09 (0,67; 1,77); 0,734	1,03 (0,86; 1,23); 0,746	1,21 (-9,93; 12,36); 0,902
Adalimumab	124	78 (62,9)			
mITT					
Baricitinib	487	325 (66,7)	1,35 (1,00; 1,80); 0,048	1,11 (1,00; 1,23); 0,060	7,04 (0,29; 13,78); 0,045
Adalimumab	330	197 (59,7)			
Heterogenität^e					p = 0,6860
Woche 52					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	153 (63,0)	1,30 (0,85; 2,00); 0,228	1,08 (0,92; 1,27); 0,327	6,10 (-3,82; 16,02); 0,246
Adalimumab	153	87 (56,9)			
Subpopulation A3					
Baricitinib	170	104 (61,2)	1,28 (0,79; 2,08); 0,315	1,08 (0,89; 1,32); 0,437	3,92 (-7,46; 15,30); 0,548
Adalimumab	124	71 (57,3)			
mITT					
Baricitinib	487	299 (61,4)	1,31 (0,98; 1,75); 0,067	1,10 (0,98; 1,24); 0,103	6,55 (-0,35; 13,44); 0,070
Adalimumab	330	181 (54,8)			
Heterogenität^e					p = 0,9068

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der körperlichen Funktionen, als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der körperlichen Funktionen, als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$ über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der körperlichen Funktionen, als unter Adalimumab. Der Unterschied war statistisch nicht signifikant (Tabelle 4-61).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2.9 Dauer der Morgensteifigkeit (min) – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-62: Operationalisierung der Dauer der Morgensteifigkeit (min)

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Baseline (Woche 0), Woche 24, Woche 52</p> <ul style="list-style-type: none"> Veränderung der Dauer der Morgensteifigkeit (min) von Baseline von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mLOCF) <p><u>Definition:</u> Um die Dauer der Morgensteifigkeit zu bestimmen, wird der Patient gebeten, das Anhalten der Morgensteifigkeit der von der RA betroffenen Gelenke in Minuten mit Hilfe eines elektronischen <i>Patient-Reported-Outcome</i> (ePRO)-Tablets am Tag der Studienvisite zu dokumentieren.</p> <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].	

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

Tabelle 4-63: Bewertung des Verzerrungspotenzials für Dauer der Morgensteifigkeit (min) in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	nein	niedrig
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].						

Begründen Sie für jede Studie die abschließende Einschätzung.

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt Dauer der Morgensteifigkeit (min) als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung. Die zur Messung verwendeten ePRO-Tablets wurden an die teilnehmenden Studienzentren ausgeteilt, nachdem bereits ein Teil der Studienpopulation eingeschlossen war. Daher existieren für einen Teil der Patienten (560/1305) keine Baseline-Daten zu diesem Endpunkt. Diese Patienten flossen daher nicht in die Auswertung dieses Endpunkts mit ein. Da für den Großteil der eingeschlossenen Patienten jedoch Baseline-Daten vorlagen und das Nichtvorliegen von Baseline-Daten nicht mit bestimmten Patientencharakteristika in Verbindung steht (*Missing Completely at Random*), wird nicht von einer Verzerrung des Endpunkts ausgegangen.

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

Veränderung der Dauer der Morgensteifigkeit (min)

Tabelle 4-64: Ergebnisse für Veränderung der Dauer der Morgensteifigkeit aus RCT mit dem zu bewertenden Arzneimittel (mLOCF)

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der Dauer der Morgensteifigkeit									
Woche 24									
Subpopulation A1 Nicht zutreffend.									
Subpopulation A2									
Baricitinib	243	132	133,0 (174,6)	130	48,91 (112,35)	130	-71,70 (13,287)	-54,3 (-91,08; -17,66); 0,004	-0,36 (-0,65; -0,08)
Adalimumab	153	78	126,92 (146,74)	78	103,62 (172,77)	78	-17,34 (16,351)		
Subpopulation A3									
Baricitinib	170	98	115,16 (118,38)	98	51,24 (88,60)	98	-55,46 (16,730)	-5,61 (-43,12; 31,90); 0,769	-0,03 (-0,33; 0,26)
Adalimumab	124	79	124,04 (157,41)	79	63,46 (124,96)	79	-49,85 (18,616)		
mITT									
Baricitinib	487	279	120,53 (143,52)	277	56,31 (120,76)	277	-54,75 (9,072)	-20,65 (-45,26; 3,95); 0,100	-0,14 (-0,32; 0,05)
Adalimumab	330	190	121,32 (150,37)	190	78,21 (142,51)	190	-34,09 (10,525)		
Heterogenität^{a,b}								p = 0,050	

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der Dauer der Morgensteifigkeit									
Woche 52									
Subpopulation A1 Nicht zutreffend.									
Subpopulation A2									
Baricitinib	243	132	133,0 (174,6)	130	57,57 (140,55)	130	-70,58 (14,891)	-41,49 (-80,56; -2,42); 0,038	-0,25 (-0,53; 0,03)
Adalimumab	153	78	126,92 (146,74)	78	97,41 (167,83)	78	-29,09 (18,140)		
Subpopulation A3									
Baricitinib	170	98	115,16 (118,38)	98	40,72 (87,31)	98	-68,03 (16,515)	-24,14 (-58,65; 10,37); 0,169	-0,15 (-0,44; 0,15)
Adalimumab	124	79	124,04 (157,41)	79	68,68 (141,76)	79	-43,89 (17,826)		
mITT									
Baricitinib	487	279	120,53 (143,52)	277	54,04 (128,00)	277	-61,36 (9,351)	-27,34 (-51,35; -3,32); 0,026	-0,18 (-0,36; 0,01)
Adalimumab	330	190	121,32 (150,37)	190	80,61 (147,70)	190	-34,03 (10,674)		
Heterogenität^b								p = 0,547	
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.									
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.									
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.									
Quelle: [59].									

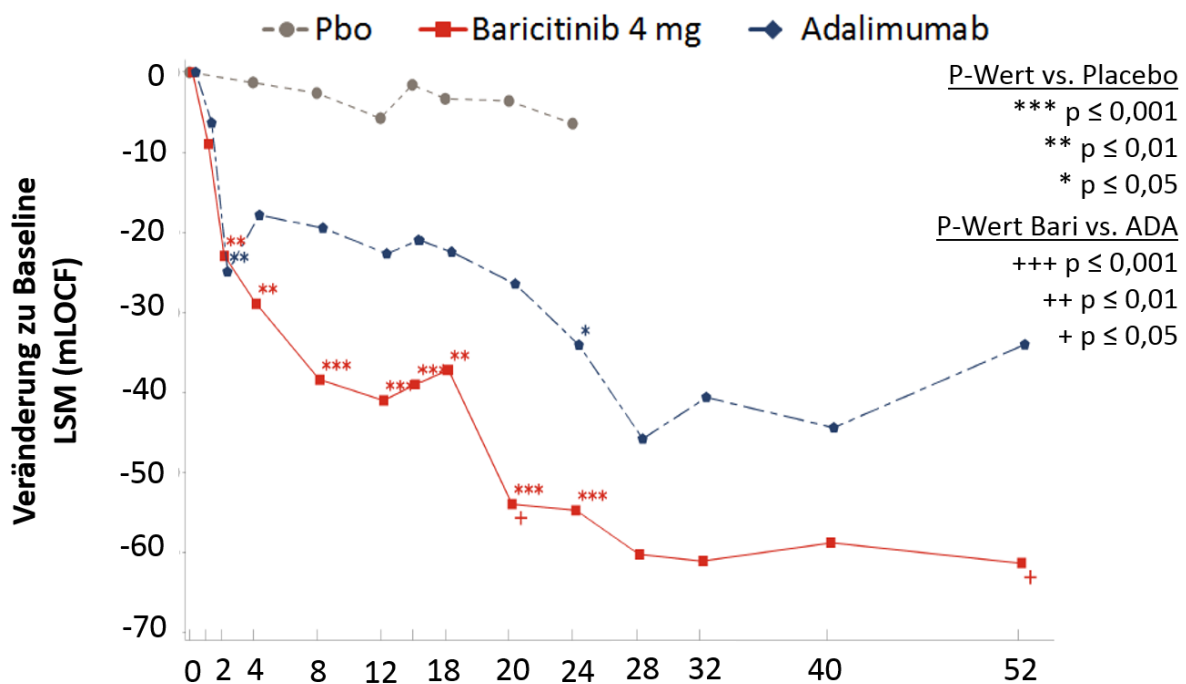
In der **Subpopulation A2** verkürzte sich die Dauer der Morgensteifigkeit unter Baricitinib von Baseline bis Woche 24 um 71,70 min und bis Woche 52 um 70,58 min, hingegen unter Adalimumab nur um 17,34 min bzw. um 29,09 min. Der Unterschied zwischen den Behandlungsarmen Baricitinib und Adalimumab betrug 54,30 min an Woche 24 und 41,49 min an Woche 52 mit statistisch signifikantem Effekt zugunsten von Baricitinib zu beiden Zeitpunkten. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verkürzung der Dauer der Morgensteifigkeit.

In der **Subpopulation A3** verkürzte sich die Dauer der Morgensteifigkeit unter Baricitinib von Baseline bis Woche 24 um 55,64 min und bis Woche 52 um 68,03 min und unter Adalimumab um 49,85 min bzw. 43,89 min. Der Unterschied zwischen den Behandlungsarmen Baricitinib und Adalimumab betrug 5,61 min an Woche 24 und 24,14 min an Woche 52 zugunsten von Baricitinib. Diese Unterschiede waren statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung der Dauer der Morgensteifigkeit über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte werden auf die Subpopulationen übertragen.

In der **mITT Population** verkürzte sich die Dauer der Morgensteifigkeit unter Baricitinib von Baseline bis Woche 24 um 54,75 min und bis Woche 52 um 61,36 min, hingegen unter Adalimumab nur um 34,09 min bzw. um 34,03 min. Der Unterschied zwischen den Behandlungsarmen Baricitinib und Adalimumab betrug 20,65 min zu Woche 24 und 27,34 min zu Woche 52 mit statistisch signifikantem Effekt zugunsten von Baricitinib zu Woche 52. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verkürzung der Dauer der Morgensteifigkeit (Tabelle 4-64).

In der Subpopulation A3 und in der mITT Population verkürzte sich die Dauer der Morgensteifigkeit im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter. Dies spricht für die anhaltende Wirkung von Baricitinib.



Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [161].

Abbildung 10: Zeitlicher Verlauf der Dauer der Morgensteifigkeit über 52 Wochen (mITT)

Abbildung 10 zeigt den zeitlichen Verlauf der Dauer der Morgensteifigkeit über 52 Wochen für die Behandlungsarme Baricitinib, Adalimumab und Placebo. Baricitinib ist Adalimumab zu Woche 24 und Woche 52 statistisch signifikant überlegen ($p < 0,05$).

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

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2.10 European Quality of Life-5 Dimensions (EQ-5D) (VAS) – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-65: Operationalisierung der EQ-5D VAS

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Baseline (Woche 0), Woche 24, Woche 52</p> <ul style="list-style-type: none"> • Veränderung der EQ-5D VAS von Baseline von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mLOCF) • Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm von Baseline bis Woche 24 und Woche 52 (NRI) <p><u>Definition:</u> Die EQ-5D VAS (0-100 mm) bewertet den allgemeinen Gesundheitszustand eines Patienten und kategorisiert diesen auf einer Skala von null „schlechtester vorstellbarer Gesundheitszustand“ bis 100 „bester vorstellbarer Gesundheitszustand“. Eine Zunahme des Wertes spiegelt somit eine Verbesserung des Gesundheitszustandes des Patienten wider.</p> <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].	

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

Tabelle 4-66: Bewertung des Verzerrungspotenzials für EQ-5D VAS in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	ja	niedrig
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].						

Begründen Sie für jede Studie die abschließende Einschätzung.

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt EQ-5D VAS als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Stellen Sie die Ergebnisse für den Endpunkt EQ-5D VAS für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Veränderung der EQ-5D VAS

Tabelle 4-67: Ergebnisse für Veränderung der EQ-5D VAS aus RCT mit dem zu bewertenden Arzneimittel (mLOCF)

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung der EQ-5D VAS								
Woche 24								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	242	50,5 (20,9)	238	68,3 (22,8)	237	15,8 (1,8)	6,4 (1,7; 11,2); 0,008
Adalimumab	153	152	48,8 (21,5)	149	61,7 (24,3)	148	9,3 (2,2)	
Subpopulation A3								
Baricitinib	170	170	51,0 (20,3)	169	69,2 (23,0)	169	17,5 (2,0)	5,4 (0,2; 10,6); 0,042
Adalimumab	124	124	51,7 (21,4)	121	64,8 (24,1)	121	12,1 (2,3)	
mITT								
Baricitinib	487	486	50,9 (20,1)	480	68,6 (22,6)	479	15,9 (1,2)	5,6 (2,4; 8,8); < 0,001
Adalimumab	330	329	50,3 (21,5)	321	62,8 (24,1)	320	10,3 (1,4)	
Heterogenität^{a,b} p = 0,019								

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung der EQ-5D VAS								
Woche 52								
Subpopulation A1 Nicht zutreffend.								
Subpopulation A2								
Baricitinib	243	242	50,5 (20,9)	238	70,2 (22,3)	237	19,6 (2,0)	8,3 (3,5; 13,1); < 0,001
Adalimumab	153	152	48,8 (21,5)	149	62,1 (25,5)	148	11,3 (2,3)	
Subpopulation A3								
Baricitinib	170	170	51,0 (20,3)	169	71,9 (23,8)	169	18,6 (2,4)	5,6 (0,1; 11,2); 0,046
Adalimumab	124	124	51,7 (21,4)	121	67,0 (25,6)	121	12,9 (2,7)	
mITT								
Baricitinib	487	486	50,9 (20,1)	480	71,0 (22,5)	479	19,1 (1,2)	7,4 (4,2; 10,7); < 0,001
Adalimumab	330	329	50,3 (21,5)	321	63,5 (25,7)	320	11,6 (1,4)	
Heterogenität^b p = 0,482								
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.								
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.								
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.								
Quelle: [59].								

In **Subpopulation A2** verbesserte sich der allgemeine Gesundheitszustand von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung des allgemeinen Gesundheitszustandes.

In **Subpopulation A3** verbesserte sich der allgemeine Gesundheitszustand von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung des allgemeinen Gesundheitszustandes.

An Woche 24 liegt der p-Wert des Heterogenitätstest für den Endpunkt Veränderung des EQ-5D (VAS) unter 0,05, jedoch zeigen die Punktschätzer in die gleiche Richtung, mit sich weitestgehend überlappenden Konfidenzintervallen, und die Unterschiede zwischen den Subpopulationen sind medizinisch irrelevant. An Woche 52 ist der Heterogenitätstest unauffällig. Somit wird nicht von einer Modifikation des Effekts zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** verbesserte sich der allgemeine Gesundheitszustand von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung des allgemeinen Gesundheitszustandes (Tabelle 4-67).

Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm

Tabelle 4-68: Ergebnisse für Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI)^d; p-Wert	RR^a (95% KI)^d; p-Wert	ARD^b (95% KI)^c; p-Wert^c
Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	151 (63,4)	1,65 (1,08; 2,51); 0,020	1,25 (1,04; 1,50); 0,020	12,44 (2,35; 22,53); 0,020
Adalimumab	153	76 (51,0)			
Subpopulation A3					
Baricitinib	170	113 (66,9)	1,27 (0,77; 2,09); 0,350	1,06 (0,90; 1,26); 0,462	4,05 (-7,10; 15,21); 0,532
Adalimumab	124	76 (62,8)			
mITT					
Baricitinib	487	309 (64,4)	1,42 (1,06; 1,90); 0,019	1,15 (1,02; 1,29); 0,020	8,30 (1,38; 15,22); 0,022
Adalimumab	330	180 (56,1)			
Heterogenität^e p = 0,1236					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	152 (63,9)	1,49 (0,97; 2,29); 0,067	1,19 (1,00; 1,41); 0,052	9,50 (-0,56; 19,56); 0,070
Adalimumab	153	81 (54,4)			
Subpopulation A3					
Baricitinib	170	119 (70,4)	1,60 (0,95; 2,67); 0,075	1,12 (0,95; 1,32); 0,193	8,43 (-2,62; 19,48); 0,164
Adalimumab	124	75 (62,0)			
mITT					
Baricitinib	487	319 (66,5)	1,54 (1,14; 2,07); 0,004	1,16 (1,04; 1,30); 0,009	9,76 (2,89; 16,63); 0,006
Adalimumab	330	182 (56,7)			
Heterogenität^e p = 0,7317					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit einer Verbesserung der EQ-5D VAS \geq 10 mm					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung des allgemeinen Gesundheitszustandes als unter Adalimumab. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine klinisch relevante Verbesserung des allgemeinen Gesundheitszustandes.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung des allgemeinen Gesundheitszustandes als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit einer Verbesserung des EQ-5D VAS \geq 10 mm über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung des allgemeinen Gesundheitszustandes als unter Adalimumab. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine klinisch relevante Verbesserung des allgemeinen Gesundheitszustandes (Tabelle 4-68).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.2.11 Schwerste Müdigkeit (*Worst Tiredness*) – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-69: Operationalisierung der schwersten Müdigkeit

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Baseline (Woche 0), Woche 24, Woche 52</p> <ul style="list-style-type: none"> Veränderung der schwersten Müdigkeit von Baseline von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mLOCF) <p><u>Definition:</u> Mittels einer numerischen 11-Punkte Skala (NRS) wird der Patient gebeten, unter Verwendung eine ePRO-Tablets, seine Müdigkeit während der letzten 24 Stunden selbst einzuschätzen. Bei einem Wert von 0 liegt keine Müdigkeit vor, während bei einem Wert von 10 die „schwerste nur denkbare“ Müdigkeit vorliegt.</p> <p>Eine Abnahme des Wertes spiegelt somit eine Verbesserung des Symptoms Müdigkeit wider.</p> <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].	

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

Tabelle 4-70: Bewertung des Verzerrungspotenzials für schwerste Müdigkeit in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	nein	niedrig
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].						

Begründen Sie für jede Studie die abschließende Einschätzung.

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

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt Müdigkeit als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung. Die zur Messung verwendeten ePRO-Tablets wurden an die teilnehmenden Studienzentren ausgeteilt, nachdem bereits ein Teil der Studienpopulation eingeschlossen war. Daher existieren für einen Teil der Patienten (560/1305) keine Baseline-Daten zu diesem Endpunkt. Diese Patienten flossen daher nicht in die Auswertung dieses Endpunkts mit ein. Da für den Großteil der eingeschlossenen Patienten jedoch Baseline-Daten vorlagen und das Nichtvorliegen von Baseline-Daten nicht mit bestimmten Patientencharakteristika in Verbindung steht (*Missing Completely at Random*), wird nicht von einer Verzerrung des Endpunkts ausgegangen.

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

Veränderung der schwersten Müdigkeit (*Worst Tiredness*)

Tabelle 4-71: Ergebnisse für Veränderung der schwersten Müdigkeit aus RCT mit dem zu bewertenden Arzneimittel (mLOCF)

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der schwersten Müdigkeit									
Woche 24									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	132	6,3 (2,4)	130	3,6 (2,6)	130	-2,69 (0,240)	-0,14 (-0,81; 0,53); 0,677	-0,05 (-0,33; 0,23)
Adalimumab	153	78	6,6 (2,3)	78	3,8 (2,3)	78	-2,55 (0,297)		
Subpopulation A3									
Baricitinib	170	98	5,9 (2,4)	98	3,4 (2,4)	98	-2,77 (0,323)	-0,16 (-0,88; 0,57); 0,672	-0,05 (-0,34; 0,25)
Adalimumab	124	79	6,3 (2,5)	79	3,7 (2,5)	79	-2,61 (0,359)		
mITT									
Baricitinib	487	279	6,2 (2,4)	277	3,5 (2,4)	277	-2,74 (0,158)	-0,27 (-0,70; 0,16); 0,222	-0,10 (-0,29; 0,08)
Adalimumab	330	190	6,3 (2,4)	190	3,8 (2,4)	190	-2,48 (0,183)		
Heterogenität^b								p = 0,018	

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der schwersten Müdigkeit									
Woche 52									
Subpopulation A1 Nicht zutreffend.									
Subpopulation A2									
Baricitinib	243	132	6,3 (2,4)	130	3,6 (2,4)	130	-3,05 (0,246)	-0,29 (-0,93; 0,36); 0,380	-0,10 (-0,38; 0,18)
Adalimumab	153	78	6,6 (2,3)	78	3,9 (2,4)	78	-2,76 (0,300)		
Subpopulation A3									
Baricitinib	170	98	5,9 (2,4)	98	2,9 (2,4)	98	-2,73 (0,349)	-0,77 (-1,50; - 0,04); 0,039	-0,22 (-0,52; 0,07)
Adalimumab	124	79	6,3 (2,5)	79	3,7 (2,6)	79	-1,96 (0,378)		
mITT									
Baricitinib	487	279	6,2 (2,4)	277	3,3 (2,4)	277	-2,89 (0,168)	-0,59 (-1,02; -0,16); 0,007	-0,22 (-0,40; -0,03)
Adalimumab	330	190	6,3 (2,4)	190	3,9 (2,5)	190	-2,29 (0,192)		
Heterogenität^b p = 0,354									
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.									
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.									
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.									
Quelle: [59].									

In **Subpopulation A2** verbesserte sich das Symptom schwerste Müdigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** verbesserte sich das Symptom schwerste Müdigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf die Reduktion der Müdigkeit und geht mit einer Verbesserung der Lebensqualität einher.

An Woche 24 liegt der p-Wert des Heterogenitätstest für den Endpunkt Veränderung der schwersten Müdigkeit unter 0,05, jedoch zeigen die Punktschätzer in die gleiche Richtung, mit sich weitestgehend überlappenden Konfidenzintervallen, und die Unterschiede zwischen den Subpopulationen sind medizinisch irrelevant. An Woche 52 ist der Heterogenitätstest unauffällig. Somit wird nicht von einer Modifikation des Effekts zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** verbesserte sich das Symptom schwerste Müdigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf die Reduktion der Müdigkeit und geht mit einer Verbesserung der Lebensqualität einher (Tabelle 4-70).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.3 *Lebensqualität - RCT*

4.3.1.3.3.1 *Short Form (36 item) Health Survey Version 2 akut (SF-36 v2 akut) – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-72: Operationalisierung des SF-36 v2 akut

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Baseline (Woche 0), Woche 24, Woche 52</p> <ul style="list-style-type: none"> • Veränderung des SF-36 v2 akut PCS von Baseline von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mLOCF) • Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5 von Baseline bis Woche 24 und Woche 52 (NRI) • Veränderung des SF-36 v2 akut MCS von Baseline von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mLOCF) • Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 von Baseline bis Woche 24 und Woche 52 (NRI) • Veränderung der SF-36 v2 akut Domänen von Baseline von Baseline bis Woche 24 und Woche 52 (ANCOVA mit mLOCF) <p><u>Definition:</u> Mittels SF-36 werden acht Dimensionen (körperliche Funktionsfähigkeit, körperliche Rollenfunktion, körperliche Schmerzen, allgemeiner Gesundheitszustand, Vitalität, soziale Funktionsfähigkeit, emotionale Rollenfunktion mentaler Gesundheitszustand) mit insgesamt 36 Fragen der allgemeinen Gesundheit, die wiederum in der physischen (PCS) und mentalen (MCS) Summenskala zusammengefasst werden, abgebildet. Ein hoher Wert zeigt einen besseren physischen oder mentalen Gesundheitsstatus bzw. eine Zunahme des Wertes spiegelt eine Verbesserung der physischen und mentalen Komponente des Gesundheitsstatus wider.</p> <p><u>Analysepopulation:</u> Subpopulation A2, Subpopulation A3, mITT</p>
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>	

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

Tabelle 4-73: Bewertung des Verzerrungspotenzials für SF-36 v2 akut in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	ja	niedrig
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].						

Begründen Sie für jede Studie die abschließende Einschätzung.

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt SF-36 v2 akut als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Stellen Sie die Ergebnisse für den Endpunkt für SF-36 v2 akut für jede einzelne Studie in tabellarischer Form dar. Fügen Sie für jede Studie eine neue Zeile ein. Beschreiben Sie die Ergebnisse zusammenfassend.

Veränderung des SF-36 v2 akut PCS

Tabelle 4-74: Ergebnisse für Veränderung des SF-36 v2 akut PCS aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des SF-36 v2 akut PCS								
Woche 24								
Subpopulation A1		Nicht zutreffend.						
Subpopulation A2								
Baricitinib	243	242	31,6 (8,1)	237	41,9 (9,1)	237	9,9 (0,60)	1,4 (-0,2; 3,0); 0,095
Adalimumab	153	152	31,0 (9,1)	148	40,1 (9,5)	148	8,6 (0,73)	
Subpopulation A3								
Baricitinib	170	170	33,5 (8,6)	169	43,7 (9,0)	169	10,3 (0,69)	1,8 (0,0; 3,6); 0,046
Adalimumab	124	124	32,0 (8,5)	121	41,0 (9,6)	121	8,4 (0,79)	
mITT								
Baricitinib	487	486	32,4 (8,2)	479	42,3 (8,9)	479	9,7 (0,38)	1,7 (0,7; 2,8); 0,001
Adalimumab	330	329	31,8 (9,2)	320	40,3 (9,6)	320	7,9 (0,45)	
Heterogenität^b								p = 0,404
Woche 52								
Subpopulation A1		Nicht zutreffend.						
Subpopulation A2								
Baricitinib	243	242	31,6 (8,1)	237	42,4 (9,6)	237	11,3 (0,70)	1,8 (0,1; 3,5); 0,033
Adalimumab	153	152	31,0 (9,1)	148	40,3 (9,8)	148	9,5 (0,82)	
Subpopulation A3								
Baricitinib	170	170	33,5 (8,6)	169	44,1 (9,7)	169	10,0 (0,88)	1,1 (-0,9; 3,2); 0,272
Adalimumab	124	124	32,0 (8,5)	121	42,1 (10,4)	121	8,8 (0,98)	

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des SF-36 v2 akut PCS								
mITT								
Baricitinib	487	486	32,4 (8,2)	479	42,9 (9,5)	479	10,4 (0,44)	1,7 (0,6; 2,9); 0,004
Adalimumab	330	329	31,8 (9,2)	320	40,9 (10,2)	320	8,6 (0,51)	
Heterogenität^b								p = 0,523
<p>a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.</p> <p>b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>								

In **Subpopulation A2** verbesserte sich die physische Komponente des allgemeinen Gesundheitszustandes unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf einen besseren physischen Gesundheitszustand und geht mit einer Verbesserung der Lebensqualität einher.

In **Subpopulation A3** verbesserte sich die physische Komponente des allgemeinen Gesundheitszustandes unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab. Der Unterschied war von Baseline bis Woche 24 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf einen besseren physischen Gesundheitszustand und geht mit einer Verbesserung der Lebensqualität einher.

Da der Heterogenitätstest für den Endpunkt SF-36 v2 akut PCS über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** verbesserte sich die physische Komponente des allgemeinen Gesundheitszustandes unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf einen besseren physischen Gesundheitszustand und geht mit einer Verbesserung der Lebensqualität einher (Tabelle 4-74).

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

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5 Tabelle 4-75: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel

			Baricitinib vs. Adalimumab		
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5					
Woche 24					
Subpopulation A1			Nicht zutreffend.		
Subpopulation A2					
Baricitinib	243	167 (68,7)	1,43 (0,93; 2,2); 0,101	1,14 (0,97; 1,32); 0,105	9,90 (0,16; 19,64); 0,052
Adalimumab	153	90 (58,8)			
Subpopulation A3					
Baricitinib	170	106 (62,4)	1,21 (0,75; 1,95); 0,441	1,08 (0,90; 1,30); 0,396	3,48 (-7,83; 14,80); 0,548
Adalimumab	124	73 (58,9)			
mITT					
Baricitinib	487	317 (65,1)	1,42 (1,06; 1,89); 0,018	1,15 (1,02; 1,28); 0,018	8,43 (1,61; 15,25); 0,016
Adalimumab	330	187 (56,7)			
Heterogenität^e p = 0,5519					
Woche 52					
Subpopulation A1			Nicht zutreffend.		
Subpopulation A2					
Baricitinib	243	156 (64,2)	1,56 (1,02; 2,39); 0,039	1,16 (0,98; 1,38); 0,086	10,60 (0,66; 20,54); 0,045
Adalimumab	153	82 (53,6)			
Subpopulation A3					
Baricitinib	170	95 (55,9)	1,10 (0,68; 1,77); 0,701	1,03 (0,83; 1,26); 0,811	1,04 (-10,46; 12,55); 0,906
Adalimumab	124	68 (54,8)			
mITT					
Baricitinib	487	292 (60,0)	1,41 (1,06; 1,88); 0,019	1,14 (1,00; 1,29); 0,047	8,14 (1,21; 15,07); 0,022
Adalimumab	330	171 (51,8)			
Heterogenität^e p = 0,3665					

			Baricitinib vs. Adalimumab		
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5 über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine klinisch relevante Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes (Tabelle 4-75).

Veränderung des SF-36 v2 akut MCS

Tabelle 4-76: Ergebnisse für Veränderung des SF-36 v2 akut MCS aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des SF-36 v2 akut MCS								
Woche 24								
Subpopulation A1		Nicht zutreffend.						
Subpopulation A2								
Baricitinib	243	242	45,9 (11,8)	237	50,4 (11,2)	237	3,8 (0,77)	1,4 (0,6; 3,5); 0,164
Adalimumab	153	152	44,8 (12,1)	148	48,6 (11,4)	148	2,3 (0,93)	
Subpopulation A3								
Baricitinib	170	170	48,1 (10,8)	169	50,9 (10,5)	169	3,4 (0,88)	-0,7 (-2,9; 1,6); 0,564
Adalimumab	124	124	46,7 (12,1)	121	50,7 (11,5)	121	4,1 (1,00)	
mITT								
Baricitinib	487	486	46,7 (11,5)	479	50,5 (10,8)	479	3,8 (0,48)	0,5 (0,9; 1,8); 0,503
Adalimumab	330	329	45,7 (12,1)	320	49,6 (11,3)	320	3,3 (0,57)	
Heterogenität^b								p = 0,438
Woche 52								
Subpopulation A1		Nicht zutreffend.						
Subpopulation A2								
Baricitinib	243	242	45,9 (11,8)	237	50,5 (10,7)	237	3,9 (0,81)	1,7 (0,2; 3,7); 0,081
Adalimumab	153	152	44,8 (12,1)	148	48,7 (10,9)	148	2,2 (0,95)	
Subpopulation A3								
Baricitinib	170	170	48,1 (10,8)	169	51,6 (9,9)	169	4,1 (0,91)	-0,0 (-2,1; 2,1); 0,996
Adalimumab	124	124	46,7 (12,1)	121	50,9 (11,6)	121	4,1 (1,02)	

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert
Veränderung des SF-36 v2 akut MCS								
mITT								
Baricitinib	487	486	46,7 (11,5)	479	50,8 (10,4)	479	4,3 (0,49)	0,9 (0,4; 2,2); 0,169
Adalimumab	330	329	45,7 (12,1)	320	49,4 (11,3)	320	3,4 (0,57)	
Heterogenität^b								p = 0,314
a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.								
b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.								
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.								
Quelle: [59].								

In **Subpopulation A2** verbesserte sich die mentale Komponente des allgemeinen Gesundheitszustandes von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** verbesserte sich die mentale Komponente des allgemeinen Gesundheitszustandes von Baseline bis Woche 24 zugunsten von Adalimumab. Der Unterschied war statistisch nicht signifikant. Zu Woche 52 gab es keine Unterschiede zwischen den Behandlungsarmen.

Da der Heterogenitätstest für den Endpunkt Veränderung des SF-36 v2 akut MCS über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In **mITT Population** verbesserte sich die mentale Komponente des allgemeinen Gesundheitszustandes von Baseline bis Woche 24 und Woche 52 zugunsten von Adalimumab. Der Unterschied war statistisch nicht signifikant (Tabelle 4-76).

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

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 Tabelle 4-77: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI)^d; p-Wert	RR^a (95% KI)^d; p-Wert	ARD^b (95% KI)^c; p-Wert^c
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	101 (41,6)	1,64 (1,05; 2,54); 0,028	1,38 (1,05; 1,82); 0,021	11,50 (1,95; 21,05); 0,025
Adalimumab	153	46 (30,1)			
Subpopulation A3					
Baricitinib	170	55 (32,4)	0,71 (0,43; 1,16); 0,171	0,81 (0,61; 1,09); 0,173	-8,78 (-19,93; 2,38); 0,140
Adalimumab	124	51 (41,1)			
mITT					
Baricitinib	487	184 (37,8)	1,21 (0,90; 1,63); 0,211	1,13 (0,94; 1,36); 0,202	-4,15 (2,53; 10,82); 0,236
Adalimumab	330	111 (33,6)			
Heterogenität^e p = 0,0239					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	90 (37,0)	1,46 (0,93; 2,28); 0,097	1,28 (0,96; 1,72); 0,098	-8,28 (1,12; 17,68); 0,102
Adalimumab	153	44 (28,8)			
Subpopulation A3					
Baricitinib	170	50 (29,4)	0,76 (0,45; 1,27); 0,290	0,87 (0,63; 1,19); 0,374	-6,88 (-17,77; 4,01); 0,256
Adalimumab	124	45 (36,3)			
mITT					
Baricitinib	487	166 (34,1)	1,24 (0,91; 1,69); 0,172	1,14 (0,93; 1,39); 0,219	-4,69 (1,78; 11,16); 0,170
Adalimumab	330	97 (29,4)			
Heterogenität^e p = 0,0704					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI) ^d ; p-Wert	RR ^a (95% KI) ^d ; p-Wert	ARD ^b (95% KI) ^c ; p-Wert ^c
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Adalimumab; ein Wert > 1 spricht für eine Behandlung mit Baricitinib.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Adalimumab; ein Wert > 0 spricht für eine Behandlung mit Baricitinib.</p> <p>c: 95%-KI mittels Newcombe-Wilson Methode (ohne Kontinuitätskorrektur) berechnet (NRI); p-Wert mittels Fisher's Test berechnet.</p> <p>d: 95%-KI mittels Logistisches Regressionsmodell (NRI) berechnet.</p> <p>e: Interaktions-p-Wert mittels eines <i>Mixed Models</i> berechnet (NRI); Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>					

In **Subpopulation A2** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Der Unterschied war von Baseline bis Woche 24 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf einen besseren mentalen Gesundheitszustand und geht mit einer Verbesserung der Lebensqualität einher.

In **Subpopulation A3** erreichte ein geringerer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Der Unterschied war nicht signifikant.

An Woche 24 liegt der p-Wert des Heterogenitätstest für den Endpunkt Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 unter 0,05. Die Ergebnisse deuten auf einen Zusatznutzen für den Baricitinib-Behandlungsarm in Subpopulation A2 hin. An Woche 52 ist der Heterogenitätstest unauffällig. Daher wird nicht von einer Modifikation des Effekts zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch nicht signifikant (Tabelle 4-77).

Veränderung der SF-36 v2 akut Domänen

Tabelle 4-78: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Körperliche Funktionsfähigkeit aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges' g (95% KI)
Veränderung der SF-36 v2 akut Domäne – Körperliche Funktionsfähigkeit									
Woche 24									
Subpopulation A1 Nicht zutreffend.									
Subpopulation A2									
Baricitinib	243	242	31,0 (9,9)	237	40,8 (11,3)	237	9,7 (0,71)	1,4 (0,5; 3,3); 0,143	0,13 (-0,07; 0,34)
Adalimumab	153	152	30,4 (11,2)	148	39,1 (11,6)	148	8,3 (0,84)		
Subpopulation A3									
Baricitinib	170	170	34,4 (10,4)	169	43,5 (10,3)	169	9,5 (0,77)	1,7 (-0,3; 3,7); 0,091	0,17 (-0,06; 0,41)
Adalimumab	124	124	32,5 (9,9)	121	40,3 (11,4)	121	7,8 (0,88)		
mITT									
Baricitinib	487	486	32,3 (10,2)	479	41,6 (10,7)	479	9,2 (0,44)	1,7 (0,5; 3,0); 0,006	0,18 (0,04; 0,32)
Adalimumab	330	329	31,6 (10,7)	320	39,4 (11,4)	320	7,5 (0,52)		
Heterogenität^b p = 0,644									
Woche 52									
Subpopulation A1 Nicht zutreffend.									
Subpopulation A2									
Baricitinib	243	242	31,0 (9,9)	237	40,8 (11,6)	237	10,4 (0,82)	1,3 (0,7; 3,3); 0,195	0,11 (-0,10; 0,31)
Adalimumab	153	152	30,4 (11,2)	148	39,4 (11,7)	148	9,1 (0,96)		
Subpopulation A3									
Baricitinib	170	170	34,4 (10,4)	169	43,9 (10,3)	169	9,5 (0,92)	1,4 (-0,8; 3,5); 0,217	0,12 (-0,12; 0,35)
Adalimumab	124	124	32,5 (9,9)	121	41,3 (11,7)	121	8,2 (1,04)		

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der SF-36 v2 akut Domäne – Körperliche Funktionsfähigkeit									
mITT									
Baricitinib	487	486	32,3 (10,2)	479	41,9 (10,9)	479	9,9 (0,49)	1,4 (0,1; 2,7); 0,029	0,14 (-0,00; 0,28)
Adalimumab	330	329	31,6 (10,7)	320	40,1 (11,7)	320	8,4 (0,57)		
Heterogenität^b								p = 1,000	
<p>a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.</p> <p>b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>									

In **Subpopulation A2** verbesserte sich die körperliche Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** verbesserte sich die körperliche Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 24 statistisch signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung der SF-36 v2 akut Domäne – Körperliche Funktionsfähigkeit über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** verbesserte sich die körperliche Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant (Tabelle 4-78).

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

Tabelle 4-79: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Körperliche Rollenfunktion aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges' g (95% KI)
Veränderung der SF-36 v2 akut Domäne – Körperliche Rollenfunktion									
Woche 24									
Subpopulation A1		Nicht zutreffend,							
Subpopulation A2									
Baricitinib	243	242	34,4 (10,0)	237	43,5 (9,6)	237	8,4 (0,65)	1,3 (0,4; 3,0); 0,131	0,13 (-0,07; 0,34)
Adalimumab	153	152	33,2 (10,3)	148	41,8 (10,2)	148	7,1 (0,78)		
Subpopulation A3									
Baricitinib	170	170	37,0 (10,8)	169	45,7 (9,3)	169	8,9 (0,75)	0,8 (-1,1; 2,8); 0,396	0,09 (-0,15; 0,32)
Adalimumab	124	124	34,3 (10,0)	121	43,3 (10,5)	121	8,0 (0,86)		
mITT									
Baricitinib	487	486	35,5 (10,3)	479	44,2 (9,4)	479	8,3 (0,42)	1,3 (0,2; 2,5); 0,025	0,15 (0,01; 0,29)
Adalimumab	330	329	34,5 (10,5)	320	42,4 (10,2)	320	7,0 (0,49)		
Heterogenität^b								p = 0,634	
Woche 52									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	242	34,4 (10,0)	237	44,0 (9,4)	237	10,0 (0,73)	2,3 (0,6; 4,1); 0,010	0,21 (0,01; 0,42)
Adalimumab	153	152	33,2 (10,3)	148	41,7 (10,9)	148	7,7 (0,86)		
Subpopulation A3									
Baricitinib	170	170	37,0 (10,8)	169	46,2 (9,6)	169	9,3 (0,88)	1,0 (-1,1; 3,0); 0,357	0,09 (-0,15; 0,32)
Adalimumab	124	124	34,3 (10,0)	121	43,7 (10,9)	121	8,3 (0,99)		

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges`g (95% KI)
mITT									
Baricitinib	487	486	35,5 (10,3)	479	44,8 (9,4)	479	9,4 (0,45)	1,9 (0,7; 3,1); 0,002	0,20 (0,06; 0,34)
Adalimumab	330	329	34,5 (10,5)	320	42,5 (10,8)	320	7,5 (0,52)		
Heterogenität^b								p = 0,382	
<p>a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.</p> <p>b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>									

In **Subpopulation A2** verbesserte sich die körperliche Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant.

In **Subpopulation A3** verbesserte sich die körperliche Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung der SF-36 v2 akut Domäne – Körperliche Rollenfunktion über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** verbesserte sich die körperliche Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant (Tabelle 4-79).

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

Tabelle 4-80: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Körperliche Schmerzen aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges' g (95% KI)
Veränderung der SF-36 v2 akut Domäne – Körperliche Schmerzen									
Woche 24									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	242	33,8 (7,7)	237	45,2 (9,1)	237	10,5 (0,67)	1,7 (0,1; 3,4); 0,063	0,17 (-0,04; 0,37)
Adalimumab	153	152	33,4 (8,4)	148	43,5 (9,3)	148	8,8 (0,80)		
Subpopulation A3									
Baricitinib	170	170	35,6 (7,3)	169	46,6 (8,9)	169	10,8 (0,76)	2,3 (0,3; 4,2); 0,023	0,23 (-0,00; 0,47)
Adalimumab	124	124	35,0 (8,2)	121	44,0 (9,3)	121	8,6 (0,86)		
mITT									
Baricitinib	487	486	34,6 (7,5)	479	45,5 (9,0)	479	10,3 (0,42)	1,8 (0,6; 3,0); 0,002	0,20 (0,06; 0,34)
Adalimumab	330	329	34,5 (8,5)	320	43,7 (9,2)	320	8,5 (0,50)		
Heterogenität^b								p = 0,506	
Woche 52									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	242	33,8 (7,7)	237	46,0 (9,4)	237	12,1 (0,80)	2,1 (0,2; 4,0); 0,032	0,18 (-0,03; 0,38)
Adalimumab	153	152	33,4 (8,4)	148	44,1 (10,1)	148	10,0 (0,94)		
Subpopulation A3									
Baricitinib	170	170	35,6 (7,3)	169	47,2 (9,7)	169	10,8 (0,98)	0,7 (-1,6; 3,0); 0,541	0,06 (-0,18; 0,29)
Adalimumab	124	124	35,0 (8,2)	121	46,3 (10,6)	121	10,1 (1,10)		

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges`g (95% KI)
mITT									
Baricitinib	487	486	34,6 (7,5)	479	46,2 (9,4)	479	11,2 (0,50)	1,5 (0,2; 2,8); 0,028	0,14 (-0,00; 0,28)
Adalimumab	330	329	34,5 (8,5)	320	44,7 (10,3)	320	9,7 (0,58)		
Heterogenität^b								p = 0,344	
<p>a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet. b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [59].</p>									

In **Subpopulation A2** verbesserten sich die körperlichen Schmerzen von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant.

In **Subpopulation A3** verbesserten sich die körperlichen Schmerzen von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 24 statistisch signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung der SF-36 v2 akut Domäne – Körperliche Schmerzen über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte werden auf die Subpopulationen übertragen.

In der **mITT Population** verbesserten sich die körperlichen Schmerzen von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant (Tabelle 4-80).

Tabelle 4-81: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Allgemeiner Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der SF-36 v2 akut Domäne – Allgemeiner Gesundheitszustand									
Woche 24									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	242	37,3 (8,4)	237	43,5 (8,7)	237	5,8 (0,60)	1,7 (0,1; 3,3); 0,037	0,19 (-0,02; 0,39)
Adalimumab	153	152	36,1 (7,7)	148	41,2 (9,2)	148	4,1 (0,72)		
Subpopulation A3									
Baricitinib	170	170	37,3 (7,7)	169	43,1 (9,3)	169	5,8 (0,72)	-0,7 (-2,5; 1,2); 0,473	-0,07 (-0,31; 0,16)
Adalimumab	124	124	36,7 (8,0)	121	43,6 (9,9)	121	6,5 (0,82)		
mITT									
Baricitinib	487	486	37,3 (8,1)	479	43,0 (8,9)	479	5,6 (0,39)	0,7 (0,4; 1,8); 0,209	0,08 (-0,06; 0,22)
Adalimumab	330	329	36,3 (8,7)	320	41,7 (9,9)	320	4,9 (0,46)		
Heterogenität^b								p = 0,042	
Woche 52									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	242	37,3 (8,4)	237	43,9 (9,2)	237	6,5 (0,69)	2,4 (0,7; 4,0); 0,006	0,23 (0,02; 0,43)
Adalimumab	153	152	36,1 (7,7)	148	41,1 (8,8)	148	4,2 (0,81)		
Subpopulation A3									
Baricitinib	170	170	37,3 (7,7)	169	43,5 (9,4)	169	5,6 (0,86)	0,3 (-1,7; 2,3); 0,769	0,03 (-0,21; 0,26)
Adalimumab	124	124	36,7 (8,0)	121	43,1 (9,7)	121	5,3 (0,96)		

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
mITT									
Baricitinib	487	486	37,3 (8,1)	479	43,4 (9,2)	479	6,1 (0,44)	1,3 (0,2; 2,5); 0,023	0,14 (0,00; 0,28)
Adalimumab	330	329	36,3 (8,7)	320	41,6 (9,7)	320	4,8 (0,51)		
Heterogenität^b								p = 0,114	
<p>a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet. b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [59].</p>									

In **Subpopulation A2** verbesserte sich der allgemeine Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant.

In **Subpopulation A3** verbesserte sich der allgemeine Gesundheitszustand von Baseline bis Woche 24 zugunsten von Adalimumab und von Baseline bis Woche 52 zugunsten von Baricitinib. Der Unterschied war statistisch nicht signifikant.

An Woche 24 liegt der p-Wert des Heterogenitätstest für den Endpunkt allgemeiner Gesundheitszustand unter 0,05. Die Ergebnisse zeigen einen statistisch signifikanten Vorteil für den Baricitinib-Behandlungsarm in Subpopulation A2. An Woche 52 ist der Heterogenitätstest unauffällig. Daher wird nicht von einer Modifikation des Effekts zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** verbesserte sich der allgemeine Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war zu Woche 52 statistisch signifikant (Tabelle 4-81).

Tabelle 4-82: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Vitalität aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges' g (95% KI)
Veränderung der SF-36 v2 akut Domäne – Vitalität									
Woche 24									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	242	43,5 (9,7)	237	51,6 (10,2)	237	7,6 (0,71)	1,0 (0,9; 2,9); 0,290	0,09 (-0,11; 0,30)
Adalimumab	153	152	42,5 (10,4)	148	50,2 (9,5)	148	6,5 (0,85)		
Subpopulation A3									
Baricitinib	170	170	44,3 (9,3)	169	52,1 (10,3)	169	7,6 (0,85)	1,1 (-1,1; 3,2); 0,335	0,10 (-0,13; 0,33)
Adalimumab	124	124	44,2 (10,4)	121	50,8 (10,8)	121	6,5 (0,97)		
mITT									
Baricitinib	487	486	43,8 (9,5)	479	51,6 (10,1)	479	7,5 (0,46)	0,8 (0,4; 2,1); 0,195	0,09 (-0,06; 0,23)
Adalimumab	330	329	43,2 (10,5)	320	50,4 (10,2)	320	6,6 (0,54)		
Heterogenität^b								p = 0,902	
Woche 52									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	242	43,5 (9,7)	237	52,0 (9,7)	237	8,4 (0,75)	1,3 (0,5; 3,1); 0,164	0,11 (-0,09; 0,32)
Adalimumab	153	152	42,5 (10,4)	148	50,5 (10,0)	148	7,1 (0,88)		
Subpopulation A3									
Baricitinib	170	170	44,3 (9,3)	169	52,6 (10,4)	169	7,4 (0,98)	1,0 (-1,3; 3,3); 0,380	0,08 (-0,15; 0,31)
Adalimumab	124	124	44,2 (10,4)	121	51,6 (11,3)	121	6,4 (1,10)		

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
mITT									
Baricitinib	487	486	43,8 (9,5)	479	52,0 (9,9)	479	7,9 (0,48)	1,3 (0,0; 2,6); 0,045	0,13 (-0,02; 0,27)
Adalimumab	330	329	43,2 (10,5)	320	50,4 (10,7)	320	6,6 (0,56)		
Heterogenität^b								p = 0,890	
<p>a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.</p> <p>b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>									

In **Subpopulation A2** verbesserte sich die Vitalität von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** verbesserte sich die Vitalität von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung der SF-36 v2 akut Domäne – Vitalität über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** verbesserte sich die Vitalität von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant (Tabelle 4-82).

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

Tabelle 4-83: Ergebnisse für Veränderung der SF-36 v2 akut Domäne– Soziale Funktionsfähigkeit aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges' g (95% KI)
Veränderung der SF-36 v2 akut Domäne – Soziale Funktionsfähigkeit									
Woche 24									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	242	39,8 (11,8)	237	46,5 (10,2)	237	6,1 (0,74)	2,3 (0,4; 4,2); 0,020	0,21 (0,00; 0,41)
Adalimumab	153	152	39,3 (12,5)	148	44,2 (10,9)	148	3,8 (0,88)		
Subpopulation A3									
Baricitinib	170	170	43,0 (11,0)	169	48,4 (9,0)	169	6,2 (0,84)	1,4 (-0,8; 3,5); 0,218	0,13 (-0,11; 0,36)
Adalimumab	124	124	40,4 (11,6)	121	46,0 (11,2)	121	4,8 (0,96)		
mITT									
Baricitinib	487	486	40,9 (11,6)	479	47,1 (9,6)	479	6,0 (0,47)	2,0 (0,7; 3,3); 0,002	0,20 (0,06; 0,34)
Adalimumab	330	329	40,0 (12,2)	320	44,8 (11,0)	320	4,0 (0,55)		
Heterogenität^b								p = 0,770	
Woche 52									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	242	39,8 (11,8)	237	46,9 (10,1)	237	6,8 (0,81)	3,1 (1,1; 5,0); 0,002	0,25 (0,05; 0,46)
Adalimumab	153	152	39,3 (12,5)	148	44,2 (10,7)	148	3,8 (0,95)		
Subpopulation A3									
Baricitinib	170	170	43,0 (11,0)	169	48,8 (9,0)	169	6,1 (0,87)	0,8 (-1,3; 2,8); 0,453	0,07 (-0,16; 0,30)
Adalimumab	124	124	40,4 (11,6)	121	47,0 (11,0)	121	5,3 (0,98)		

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
mITT									
Baricitinib	487	486	40,9 (11,6)	479	47,5 (9,6)	479	6,6 (0,49)	1,9 (0,7; 3,2); 0,003	0,19 (0,05; 0,33)
Adalimumab	330	329	40,0 (12,2)	320	45,2 (11,1)	320	4,2 (0,56)		
Heterogenität^b								p = 0,138	
<p>a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.</p> <p>b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>									

In **Subpopulation A2** verbesserte sich die soziale Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant.

In **Subpopulation A3** verbesserte sich die soziale Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 24 statistisch signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung der SF-36 v2 akut Domäne –Soziale Funktionsfähigkeit über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** verbesserte sich die soziale Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant (Tabelle 4-83).

Tabelle 4-84: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Emotionale Rollenfunktion aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der SF-36 v2 akut Domäne – Emotional Rollenfunktion									
Woche 24									
Subpopulation A1					Nicht zutreffend.				
Subpopulation A2									
Baricitinib	243	242	39,6 (13,0)	237	46,3 (10,7)	237	5,9 (0,80)	2,5 (0,4; 4,6); 0,021	0,20 (-0,00; 0,41)
Adalimumab	153	152	39,3 (13,1)	148	44,2 (12,3)	148	3,5 (0,96)		
Subpopulation A3									
Baricitinib	170	170	44,3 (11,0)	169	48,0 (9,4)	169	5,3 (0,88)	-1,5 (-3,8; 0,8); 0,189	-0,13 (0,37; 0,10)
Adalimumab	124	124	40,6 (13,0)	121	47,8 (10,7)	121	6,8 (1,01)		
mITT									
Baricitinib	487	486	41,4 (12,5)	479	46,8 (10,2)	479	5,6 (0,50)	0,7 (0,7; 2,0); 0,355	0,06 (-0,08; 0,20)
Adalimumab	330	329	40,3 (12,9)	320	45,8 (11,5)	320	4,9 (0,59)		
Heterogenität^b								p = 0,047	
Woche 52									
Subpopulation A1					Nicht zutreffend.				
Subpopulation A2									
Baricitinib	243	242	39,6 (13,0)	237	46,4 (10,2)	237	6,7 (0,86)	2,6 (0,6; 4,7); 0,012	0,21 (0,00; 0,41)
Adalimumab	153	152	39,3 (13,1)	148	44,3 (11,9)	148	4,0 (1,01)		
Subpopulation A3									
Baricitinib	170	170	44,3 (11,0)	169	49,0 (8,6)	169	6,3 (0,89)	0,4 (-1,8; 2,5); 0,741	0,03 (0,20; 0,26)
Adalimumab	124	124	40,6 (13,0)	121	47,3 (11,0)	121	5,9 (1,01)		

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
mITT									
Baricitinib	487	486	41,4 (12,5)	479	47,3 (9,7)	479	6,6 (0,51)	1,4 (0,0; 2,7); 0,048	0,12 (-0,02; 0,27)
Adalimumab	330	329	40,3 (12,9)	320	45,7 (11,4)	320	5,3 (0,59)		
Heterogenität^b								p = 0,178	
<p>a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.</p> <p>b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [59].</p>									

In **Subpopulation A2** verbesserte sich die emotionale Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 24 und Woche 52 statistisch signifikant.

In **Subpopulation A3** verbesserte sich die emotionale Rollenfunktion von Baseline bis Woche 24 zugunsten von Adalimumab und Woche 52 zugunsten von Baricitinib. Der Unterschied war statistisch nicht signifikant.

An Woche 24 liegt der p-Wert des Heterogenitätstest für den Endpunkt Emotionale Rollenfunktion unter 0,05. Die Ergebnisse zeigen einen statistisch signifikanten Vorteil für den Baricitinib-Behandlungsarm in Subpopulation A2. An Woche 52 ist der Heterogenitätstest unauffällig. Daher wird nicht von einer Modifikation des Effekts zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** verbesserte sich die emotionale Rollenfunktion von Baseline bis Woche 52 zugunsten von Baricitinib. Der Unterschied war von Baseline bis Woche 52 statistisch signifikant. Von Baseline bis Woche 24 war die Verbesserung der emotionalen Rollenfunktion in beiden Behandlungsarmen gleich. (Tabelle 4-84).

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

Tabelle 4-85: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Mentaler Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges' g (95% KI)
Veränderung der SF-36 v2 akut Domäne – Mentaler Gesundheitszustand									
Woche 24									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	242	43,1 (11,5)	237	48,1 (11,1)	237	4,4 (0,75)	0,4 (1,5; 2,4); 0,663	0,04 (-0,17; 0,24)
Adalimumab	153	152	41,0 (11,4)	148	46,6 (11,2)	148	3,9 (0,90)		
Subpopulation A3									
Baricitinib	170	170	43,8 (10,8)	169	48,3 (11,0)	169	4,6 (0,88)	0,7 (-1,5; 3,0); 0,534	0,06 (-0,17; 0,30)
Adalimumab	124	124	44,1 (11,3)	121	47,6 (11,8)	121	3,9 (1,00)		
mITT									
Baricitinib	487	486	43,3 (11,1)	479	48,2 (10,9)	479	4,7 (0,48)	0,5 (0,8; 1,9); 0,441	0,05 (-0,09; 0,19)
Adalimumab	330	329	42,5 (11,5)	320	47,2 (11,4)	320	4,1 (0,57)		
Heterogenität^b								p = 0,916	
Woche 52									
Subpopulation A1		Nicht zutreffend.							
Subpopulation A2									
Baricitinib	243	242	43,1 (11,5)	237	48,0 (10,7)	237	4,5 (0,81)	0,7 (1,3; 2,6); 0,502	0,05 (-0,15; 0,26)
Adalimumab	153	152	41,0 (11,4)	148	46,8 (11,1)	148	3,8 (0,94)		
Subpopulation A3									
Baricitinib	170	170	43,8 (10,8)	169	48,9 (10,6)	169	5,2 (0,99)	0,4 (-1,9; 2,7); 0,714	0,03 (-0,20; 0,27)
Adalimumab	124	124	44,1 (11,3)	121	48,5 (11,8)	121	4,7 (1,10)		

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
mITT									
Baricitinib	487	486	43,3 (11,1)	479	48,3 (10,6)	479	5,1 (0,51)	0,7 (0,7; 2,0); 0,327	0,06 (-0,08; 0,20)
Adalimumab	330	329	42,5 (11,5)	320	47,2 (11,6)	320	4,4 (0,59)		
Heterogenität^b								p = 0,988	
<p>a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet. b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOCF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [59].</p>									

In **Subpopulation A2** verbesserte sich der mentale Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Adalimumab. Der Unterschied war statistisch nicht signifikant.

In **Subpopulation A3** verbesserte sich der mentale Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war statistisch nicht signifikant.

Da der Heterogenitätstest für den Endpunkt Veränderung der SF-36 v2 akut Domäne – Mentaler Gesundheitszustand über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In **mITT Population** verbesserte sich der mentale Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib. Der Unterschied war statistisch nicht signifikant. (Tabelle 4-85).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.4 Sicherheit – 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 sowie Angabe des relativen Risikos, des Odds Ratios und der absoluten Risikoreduktion – entsprechende Maße bei weiteren Messniveaus
- Effektschätzer mit zugehörigem Standardfehler
- Angabe der verwendeten statistischen Methodik inklusive der Angabe der Faktoren, nach denen ggf. adjustiert wurde

Bei Überlebenszeitanalysen soll die Kaplan-Meier-Kurve einschließlich Angaben zu den Patienten unter Risiko im Zeitverlauf (zu mehreren Zeitpunkten) abgebildet werden.

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

Sofern mehrere Studien vorliegen, sollen diese in einer Meta-Analyse zusammengefasst werden, wenn die Studien aus medizinischen (z. B. Patientengruppen) und methodischen (z. B. Studiendesign) Gründen ausreichend vergleichbar sind. Es ist jeweils zu begründen, warum eine Meta-Analyse durchgeführt wurde oder warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Sofern die vorliegenden Studien für eine Meta-Analyse geeignet sind, sollen die Meta-Analysen als Forest-Plot dargestellt werden. Die Darstellung soll ausreichende Informationen zur Einschätzung der Heterogenität der Ergebnisse zwischen den Studien in Form von geeigneten statistischen Maßzahlen enthalten (siehe Abschnitt 4.2.5.3). Eine Gesamtanalyse aller Patienten aus mehreren Studien ohne Berücksichtigung der Studienzugehörigkeit (z. B. Gesamt-Vierfeldertafel per Addition der Einzel-Vierfeldertafeln) soll vermieden werden, da so die Heterogenität nicht eingeschätzt werden kann.

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

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

Tabelle 4-86: Operationalisierung der Sicherheit

Studie	Operationalisierung
RA-BEAM (JADV)	<p><u>Erhebungszeitpunkt:</u> Woche 24, Woche 52</p> <p>Dokumentiert wurde die Gesamtrate der</p> <ul style="list-style-type: none"> • Todesfälle • Schwerwiegende Unerwünschte Ereignisse (SUE) • Unerwünschte Ereignisse (UE) • Schwere UE (ohne ausschließliche Laborereignisse) • Studienabbruch aufgrund von UE • Therapieabbruch aufgrund von UE • UE von besonderem Interesse <ul style="list-style-type: none"> ○ Infektionen ○ Schwere Infektionen ○ Kardiovaskuläre UE ○ Malignität ○ Depressionen: QIDS-SR16 ○ Allergische Reaktionen/Hypersensitivität ○ Gastrointestinale Perforation <p><u>Definition:</u> Unerwünschte Ereignisse wurden nach SOC und PT unter Verwendung des MedDRA Version 18.0 klassifiziert.</p> <p><u>Analysepopulation:</u> Safety-Population (Subpopulation A2, Subpopulation A3, mITT)</p>
<p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].</p>	

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

Tabelle 4-87: Bewertung des Verzerrungspotenzials für Sicherheit in RCT mit dem zu bewertenden Arzneimittel

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
RA-BEAM (JADV)	niedrig	ja	ja	ja	ja	niedrig
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [7-10].						

Begründen Sie für jede Studie die abschließende Einschätzung.

Für die Studie RA-BEAM (JADV) wurde das Verzerrungspotenzial für den Endpunkt Sicherheit als niedrig eingeschätzt, da das Verzerrungspotenzial auf Studienebene niedrig war sowie die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz adäquat umgesetzt wurden. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung und es lagen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

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

Todesfälle

Tabelle 4-88: Ergebnisse für Todesfälle aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI); p-Wert^c	Peto OR^a (95% KI); p-Wert^c	ARD^b (95% KI); p-Wert^d
Todesfälle					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	2 (0,82)	3,18 (0,181; Inf.); 0,525	5,12 (0,296; 88,559); 0,525	0,8 (-9,3; 10,9); 0,691
Adalimumab	153	0 (0,00)			
Subpopulation A3					
Baricitinib	170	0 (0,00)	NE	NE	NE
Adalimumab	124	0 (0,00)			
mITT					
Baricitinib	487	2 (0,41)	3,40 (0,195; Inf.); 0,518	5,36 (0,318; 90,546); 0,518	0,4 (-6,6; 7,4); 0,657
Adalimumab	330	0 (0,00)			
Heterogenität^e -					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	2 (0,82)	1,26 (0,065; 74,903); 1,000	1,25 (0,122; 12,868); 1,000	0,2 (-9,9; 10,3); 1,000
Adalimumab	153	1 (0,65)			
Subpopulation A3					
Baricitinib	170	0 (0,00)	NE	NE	NE
Adalimumab	124	0 (0,00)			
mITT					
Baricitinib	487	2 (0,41)	1,36 (0,070; 80,281); 1,000	1,34 (0,133; 13,502); 1,000	0,1 (-6,9; 7,1); 1,000
Adalimumab	330	1 (0,30)			
Heterogenität^e -					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Baricitinib; ein Wert > 1 spricht für eine Behandlung mit Adalimumab.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Baricitinib; ein Wert > 0 spricht für eine Behandlung mit Adalimumab.</p> <p>c: Fisher´s Test.</p> <p>d: normal Z-Test (mit Kontinuität Korrektur).</p> <p>e: kein Heterogenitätstest durchgeführt aufgrund zu geringer oder keiner Ereignisse.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [60].</p>					

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

Für **Subpopulation A2** traten vergleichbar wenige Todesfälle in den Behandlungsarmen Baricitinib bzw. Adalimumab auf. Es gab keine statistisch signifikanten Unterschiede zwischen den Behandlungsarmen.

Für **Subpopulation A3** traten bis zum Studienende keine Todesfälle in den Behandlungsarmen auf.

Aufgrund zu geringer oder keiner Ereignisse konnte zu Woche 24 und Woche 52 kein Heterogenitätstest durchgeführt werden.

Für die **mITT Population** traten vergleichbar wenige Todesfälle in den Behandlungsarmen Baricitinib bzw. Adalimumab auf. Es gab keine statistisch signifikanten Unterschiede zwischen den Behandlungsarmen (Tabelle 4-88).

Schwerwiegende Unerwünschte Ereignisse (SUE)

Tabelle 4-89: Ergebnisse für SUE aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI); p-Wert ^c	RR ^a (95% KI); p-Wert ^c	ARD ^b (95% KI); p-Wert ^d
SUE					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	14 (5,76)	4,62 (1,034; 42,299); 0,034	4,41 (1,016; 19,125); 0,034	4,5 (-5,7; 14,5); 0,054
Adalimumab	153	2 (1,31)			
Subpopulation A3					
Baricitinib	170	8 (4,71)	6,07 (0,793; 271,671); 0,084	3,70 ^f (0,969; 14,167); 0,084	3,9 (-7,7; 15,4); 0,115
Adalimumab	124	1 (0,81)			
mITT					
Baricitinib	487	23 (4,72)	2,68 (1,043; 8,120); 0,033	2,60 (1,069; 6,310); 0,033	2,9 (-4,1; 9,9); 0,045
Adalimumab	330	6 (1,82)			
Heterogenität^e p = 0,828					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	21 (8,64)	3,52 (1,153; 14,366); 0,019	3,31 (1,157; 9,446); 0,019	6,0 (-4,1; 16,1); 0,029
Adalimumab	153	4 (2,61)			
Subpopulation A3					
Baricitinib	170	12 (7,06)	1,81 (0,572; 6,719); 0,320	1,75 (0,633; 4,841); 0,320	3,0 (-8,5; 14,5); 0,398
Adalimumab	124	5 (4,03)			

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI); p-Wert ^c	RR ^a (95% KI); p-Wert ^c	ARD ^b (95% KI); p-Wert ^d
SUE					
mITT					
Baricitinib	487	38 (7,80)	2,06 (1,054; 4,290); 0,027	1,98 (1,072; 3,660); 0,027	3,9 (-3,1; 10,8); 0,036
Adalimumab	330	13 (3,94)			
Heterogenität^e					p = 0,394
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Baricitinib; ein Wert > 1 spricht für eine Behandlung mit Adalimumab.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Baricitinib; ein Wert > 0 spricht für eine Behandlung mit Adalimumab.</p> <p>c: Fisher´s Test.</p> <p>d: normal Z-Test (mit Kontinuität Korrektur).</p> <p>e: Interaktions-p-Wert mittels Logistischem Regressionsmodell berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>f: Peto´s Odds Ratio.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [60].</p>					

In **Subpopulation A2** traten unter Baricitinib an Woche 24 und Woche 52 statistisch signifikant mehr schwerwiegende unerwünschte Ereignisse auf als unter Adalimumab.

In **Subpopulation A3** unterschieden sich die schwerwiegenden unerwünschten Ereignisse in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

Da der Heterogenitätstest für den Endpunkt SUE über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** traten unter Baricitinib an Woche 24 und Woche 52 statistisch signifikant mehr schwerwiegende unerwünschte Ereignisse auf als unter Adalimumab (Tabelle 4-89).

Unerwünschte Ereignisse (UE)

Tabelle 4-90: Ergebnisse für UE aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI); p-Wert^c	RR^a 95% KI); p-Wert^c	ARD^b (95% KI); p-Wert^c
UE					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	151 (62,14)	1,09 (0,719; 1,647); 0,689	1,0 (0,879; 1,215); 0,689	2,0 (-7,9; 11,9); 0,769
Adalimumab	153	92 (60,13)			
Subpopulation A3					
Baricitinib	170	138 (81,18)	1,38 (0,784; 2,416); 0,265	1,07 (0,947; 1,211); 0,265	5,4 (-4,2; 14,9); 0,332
Adalimumab	124	94 (75,81)			
mITT					
Baricitinib	487	347 (71,25)	1,17 (0,866; 1,588); 0,302	1,05 (0,956; 1,152); 0,302	3,4 (-3,1; 9,8); 0,340
Adalimumab	330	224 (67,88)			
Heterogenität^e p = 0,732					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	173 (71,19)	0,97 (0,617; 1,513); 0,880	0,99 (0,872; 1,125); 0,880	-0,7 (-9,8; 8,4); 0,971
Adalimumab	153	110 (71,90)			
Subpopulation A3					
Baricitinib	170	145 (85,29)	1,32 (0,710; 2,457); 0,379	1,05 (0,943; 1,163); 0,379	3,8 (-4,8; 12,5); 0,471
Adalimumab	124	101 (81,45)			
mITT					
Baricitinib	487	384 (78,85)	1,13 (0,811; 1,587); 0,460	1,03 (0,954; 1,109); 0,460	2,2 (-3,6; 8,0); 0,514
Adalimumab	330	253 (76,67)			
Heterogenität^e p = 0,506					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI); p-Wert ^c	RR ^a 95% KI); p-Wert ^c	ARD ^b (95% KI); p-Wert ^c
UE					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Baricitinib; ein Wert > 1 spricht für eine Behandlung mit Adalimumab.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Baricitinib; ein Wert > 0 spricht für eine Behandlung mit Adalimumab.</p> <p>c: Chi-Quadrat Test.</p> <p>d: normal Z-Test (mit Kontinuität Korrektur).</p> <p>e: Interaktions-p-Wert mittels Logistischem Regressionsmodell berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [60].</p>					

In **Subpopulation A2** unterschieden sich die unerwünschten Ereignisse in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

In **Subpopulation A3** unterschieden sich die unerwünschten Ereignisse in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

Da der Heterogenitätstest für den Endpunkt UE über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** unterschieden sich die unerwünschten Ereignisse in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander (Tabelle 4-90).

Schwere UE (ohne ausschließliche Laborereignisse)

Tabelle 4-91: Ergebnisse für schwere UE (ohne ausschließliche Laborereignisse) aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI); p-Wert^c	RR^a (95% KI); p-Wert^c	ARD (95% KI); p-Wert^d
Schwere UE (ohne ausschließliche Laborereignisse)					
Woche 24					
Subpopulation A1			Nicht zutreffend.		
Subpopulation A2					
Baricitinib	243	10 (4,12)	2,15 (0,540; 12,308); 0,386	2,10 (0,587; 7,505); 0,386	2, (-8,0; 12,2); 0,378
Adalimumab	153	3 (1,96)			
Subpopulation A3					
Baricitinib	170	4 (2,35)	2,96 (0,288; 147,128); 0,401	2,51 ^f (0,421; 15,015); 0,401	1,5 (-10,0; 13,1); 0,578
Adalimumab	124	1 (0,81)			
mITT					
Baricitinib	487	18 (3,70)	2,49 (0,880; 8,675); 0,084	2,44 (0,915; 6,506); 0,084	2,2 (-4,8; 9,2); 0,102
Adalimumab	330	5 (1,52)			
Heterogenität^e					
p = 0,798					
Woche 52					
Subpopulation A1			Nicht zutreffend.		
Subpopulation A2					
Baricitinib	243	16 (6,58)	1,73 (0,623; 5,508); 0,368	1,68 (0,672; 4,197); 0,368	2, (-7,5; 12,7); 0,368
Adalimumab	153	6 (3,92)			
Subpopulation A3					
Baricitinib	170	6 (3,53)	1,48 (0,307; 9,286); 0,738	1,46 (0,372; 5,721); 0,738	1,1 (-10,4; 12,7); 0,839
Adalimumab	124	3 (2,42)			
mITT					
Baricitinib	487	28 (5,75)	1,77 (0,838; 3,995); 0,133	1,72 (0,871; 3,416); 0,133	2,4 (-4,6; 9,4); 0,155
Adalimumab	330	11 (3,33)			
Heterogenität^e					
p = 0,867					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI); p-Wert ^c	RR ^a (95% KI); p-Wert ^c	ARD (95% KI); p-Wert ^d
Schwere UE (ohne ausschließliche Laborereignisse)					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Baricitinib; ein Wert > 1 spricht für eine Behandlung mit Adalimumab.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Baricitinib; ein Wert > 0 spricht für eine Behandlung mit Adalimumab.</p> <p>c: Fisher´s Test.</p> <p>d: normal Z-Test (mit Kontinuität Korrektur).</p> <p>e: Interaktions-p-Wert mittels Logistischem Regressionsmodell berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>f: Peto´s Odds Ratio.</p> <p>Alle in der vorliegenden Nutzenbewertung aufgeführten schweren unerwünschten Ereignisse wurden ohne Laborereignisse ausgewertet, in deren Zusammenhang keine Symptome beschrieben wurden oder dies angenommen werden musste und die somit asymptomatisch waren (Abschnitt 4.2.5.2). Diese schweren UE waren ausschließlich Laborereignisse ohne assoziierte Symptomatik und daher für den Patienten nicht zu spüren, also nicht patientenrelevant. Sollte eine mit diesen Laborereignissen in Zusammenhang stehende Symptomatik aufgetreten sein, dann wurde diese Symptomatik innerhalb der unerwünschten Ereignisse erfasst.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [60].</p>					

In **Subpopulation A2** unterschieden sich die schweren unerwünschten Ereignisse (ohne ausschließliche Laborereignisse) in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

In **Subpopulation A3** unterschieden sich die schweren unerwünschten Ereignisse (ohne ausschließliche Laborereignisse) in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

Da der Heterogenitätstest für den Endpunkt schwere UE (ohne ausschließliche Laborereignisse) über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** unterschieden sich die schweren unerwünschten Ereignisse (ohne ausschließliche Laborereignisse) in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander (Tabelle 4-91).

Studienabbrüche aufgrund von UE

Tabelle 4-92: Ergebnisse für Studienabbrüche aufgrund von UE aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI); p-Wert^c	RR^a (95% KI); p-Wert^c	ARD^b (95% KI); p-Wert^d
Studienabbruch aufgrund von UE					
Woche 24					
Subpopulation A1			Nicht zutreffend.		
Subpopulation A2					
Baricitinib	243	9 (3,70)	1,92 (0,469; 11,202); 0,383	1,89 (0,519; 6,868); 0,383	1,7 (-8,4; 11,8); 0,494
Adalimumab	153	3 (1,96)			
Subpopulation A3					
Baricitinib	170	7 (4,12)	1,73 (0,385; 10,572); 0,527	1,70 (0,449; 6,452); 0,527	1,7 (-9,9; 13,2); 0,640
Adalimumab	124	3 (2,42)			
mITT					
Baricitinib	487	22 (4,52)	2,18 (0,887; 6,119); 0,083	2,13 (0,920; 4,928); 0,083	2,4 (-4,6; 9,4); 0,104
Adalimumab	330	7 (2,12)			
Heterogenität^e					
p = 0,912					
Woche 52					
Subpopulation A1			Nicht zutreffend.		
Subpopulation A2					
Baricitinib	243	14 (5,76)	1,50 (0,526; 4,861); 0,486	1,47 (0,577; 3,741); 0,486	1,8 (-8,3; 11,9); 0,563
Adalimumab	153	6 (3,92)			
Subpopulation A3					
Baricitinib	170	14 (8,24)	2,69 (0,816; 11,486); 0,088	2,55 (0,861; 7,569); 0,088	5,0 (-6,5; 16,5); 0,128
Adalimumab	124	4 (3,23)			
mITT					
Baricitinib	487	34 (6,98)	1,99 (0,985; 4,283); 0,045	1,92 (1,009; 3,652); 0,045	3,3 (-3,6; 10,3); 0,060
Adalimumab	330	12 (3,64)			
Heterogenität^e					
p = 0,450					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI); p-Wert ^c	RR ^a (95% KI); p-Wert ^c	ARD ^b (95% KI); p-Wert ^d
Studienabbruch aufgrund von UE					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Baricitinib; ein Wert > 1 spricht für eine Behandlung mit Adalimumab.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Baricitinib; ein Wert > 0 spricht für eine Behandlung mit Adalimumab.</p> <p>c: Fisher´s Test.</p> <p>d: normal Z-Test (mit Kontinuität Korrektur).</p> <p>e: Interaktions-p-Wert mittels Logistischem Regressionsmodell berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [60].</p>					

In **Subpopulation A2** unterschieden sich die Studienabbrüche aufgrund von unerwünschten Ereignissen in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

In **Subpopulation A3** unterschieden sich die Studienabbrüche aufgrund von unerwünschten Ereignissen in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

Da der Heterogenitätstest für den Endpunkt Studienabbrüche über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** unterschieden sich die Studienabbrüche aufgrund von unerwünschten Ereignissen in den beiden Behandlungsarmen an Woche 24 nicht statistisch signifikant voneinander. An Woche 52 traten unter Baricitinib signifikant mehr Studienabbrüche aufgrund von unerwünschten Ereignissen auf als unter Adalimumab (Tabelle 4-92).

Therapieabbrüche aufgrund von UE

Tabelle 4-93: Ergebnisse für Therapieabbrüche aufgrund von UE aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI); p-Wert^c	RR^a (95% KI); p-Wert^c	ARD^b (95% KI); p-Wert^d
Therapieabbruch aufgrund von UE					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	9 (3,70)	1,92 (0,469; 11,202); 0,383	1,89 (0,519; 6,868); 0,383	1,7 (-8,4; 11,8); 0,494
Adalimumab	153	3 (1,96)			
Subpopulation A3					
Baricitinib	170	8 (4,71)	1,99 (0,465; 11,871); 0,367	1,95 (0,527; 7,184); 0,367	2,3 (-9,3; 13,8); 0,478
Adalimumab	124	3 (2,42)			
mITT					
Baricitinib	487	23 (4,72)	2,29 (0,935; 6,383); 0,059	2,23 (0,967; 5,129); 0,059	2,6 (-4,4; 9,6); 0,080
Adalimumab	330	7 (2,12)			
Heterogenität^e p = 0,975					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	14 (5,76)	1,28 (0,468; 3,823); 0,654	1,26 (0,520; 3,049); 0,654	1,2 (-8,9; 11,3); 0,777
Adalimumab	153	7 (4,58)			
Subpopulation A3					
Baricitinib	170	14 (8,24)	2,69 (0,816; 11,486); 0,088	2,55 (0,861; 7,569); 0,088	5,0 (-6,5; 16,5); 0,128
Adalimumab	124	4 (3,23)			
mITT					
Baricitinib	487	34 (6,98)	1,69 (0,868; 3,474); 0,129	1,65 (0,897; 3,018); 0,129	2,7 (-4,3; 9,7); 0,138
Adalimumab	330	14 (4,24)			
Heterogenität^e p = 0,323					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI); p-Wert ^c	RR ^a (95% KI); p-Wert ^c	ARD ^b (95% KI); p-Wert ^d
Therapieabbruch aufgrund von UE					
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Baricitinib; ein Wert > 1 spricht für eine Behandlung mit Adalimumab.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Baricitinib; ein Wert > 0 spricht für eine Behandlung mit Adalimumab.</p> <p>c: Fisher´s Test.</p> <p>d: normal Z-Test (mit Kontinuität Korrektur).</p> <p>e: Interaktions-p-Wert mittels Logistischem Regressionsmodell berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [60].</p>					

In **Subpopulation A2** unterschieden sich die Therapieabbrüche aufgrund von unerwünschten Ereignissen in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

In **Subpopulation A3** unterschieden sich die Therapieabbrüche aufgrund von unerwünschten Ereignissen in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

Da der Heterogenitätstest für den Endpunkt Therapieabbrüche über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** unterschieden sich die Therapieabbrüche aufgrund von unerwünschten Ereignissen in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander (Tabelle 4-93).

UE von besonderem Interesse

Tabelle 4-94: Ergebnisse für UE von besonderem Interesse aus RCT mit dem zu bewertenden Arzneimittel

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI); p-Wert^c	RR^a (95% KI); p-Wert^c	ARD^b (95% KI); p-Wert^d
Infektionen					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	72 (29,63)	1,08 (0,689; 1,685); 0,745	1,05 (0,766; 1,450); 0,745	1,5 (-7,6; 10,7); 0,832
Adalimumab	153	43 (28,10)			
Subpopulation A3					
Baricitinib	170	79 (46,47)	1,42 (0,887; 2,279); 0,143	1,23 (0,929; 1,617); 0,143	8,6 (-2,8; 19,9); 0,178
Adalimumab	124	47 (37,90)			
mITT					
Baricitinib	487	176 (36,14)	1,13 (0,843; 1,519); 0,409	1,08 (0,894; 1,315); 0,409	2,8 (-3,8; 9,4); 0,453
Adalimumab	330	110 (33,33)			
Heterogenität^e p = 0,484					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	100 (41,15)	1,18 (0,777; 1,785); 0,440	1,10 (0,856; 1,425); 0,440	3,9 (-6,0; 13,7); 0,505
Adalimumab	153	57 (37,25)			
Subpopulation A3					
Baricitinib	170	98 (57,65)	1,32 (0,828; 2,099); 0,245	1,13 (0,914; 1,408); 0,245	6,8 (-4,7; 18,4); 0,296
Adalimumab	124	63 (50,81)			
mITT					
Baricitinib	487	233 (47,84)	1,17 (0,884; 1,550); 0,272	1,09 (0,934; 1,269); 0,272	3,9 (-3,0; 10,9); 0,305
Adalimumab	330	145 (43,94)			
Heterogenität^e p = 0,875					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI); p-Wert^c	RR^a (95% KI); p-Wert^c	ARD^b (95% KI); p-Wert^d
schwere Infektionen					
Woche 24					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	1 (0,41)	0,63 (0,008; 49,625); 1,000	0,62 ^g (0,036; 10,691); 1,000	-0,2 (-10,3; 9,9); 1,000
Adalimumab	153	1 (0,65)			
Subpopulation A3					
Baricitinib	170	1 (0,59)	2,20 (0,038; Inf); 1,000	5,64 ^g (0,107; 298,336); 1,000	0,6 (-11,0; 12,1); 1,000
Adalimumab	124	0 (0,00)			
mITT					
Baricitinib	487	3 (0,62)	1,02 (0,116; 12,232); 1,000	1,02 ^g (0,170; 6,093); 1,000	0,0 (-7,0; 7,0); 1,000
Adalimumab	330	2 (0,61)			
Heterogenität^f					
-					
Woche 52					
Subpopulation A1		Nicht zutreffend.			
Subpopulation A2					
Baricitinib	243	4 (1,65)	1,26 (0,179; 14,121); 1,000	1,26 (0,233; 6,792); 1,000	0,3 (-9,8; 10,4); 1,000
Adalimumab	153	2 (1,31)			
Subpopulation A3					
Baricitinib	170	2 (1,18)	1,46 (0,075; 87,081); 1,000	1,44 ^g (0,145; 14,359); 1,000	0,4 (-11,2; 11,9); 1,000
Adalimumab	124	1 (0,81)			
mITT					
Baricitinib	487	8 (1,64)	1,36 (0,361; 6,226); 0,771	1,36 (0,411; 4,464); 0,771	0,4 (-6,6; 7,4); 0,837
Adalimumab	330	4 (1,21)			
Heterogenität^e					
p = 0,921					

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

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI); p-Wert^c	RR^a (95% KI); p-Wert^c	ARD^b (95% KI); p-Wert^d
Kardiovaskuläre Ereignisse					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	1 (0,41)	1,90 (0,033; Inf.); 1,000	5,10 [§] (0,091; 285,686); 1,000	0,4 (-9,7; 10,5); 1,000
Adalimumab	153	0 (0,00)			
Subpopulation A3					
Baricitinib	170	0 (0,00)	NE	NE	NE
Adalimumab	124	0 (0,00)			
mITT					
Baricitinib	487	1 (0,21)	2,04 (0,036; Inf.); 1,000	5,35 [§] (0,099; 290,613); 1,000	0,2 (-6,8; 7,2); 1,000
Adalimumab	330	0 (0,00)			
Heterogenität^f -					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	1 (0,41)	0,63 (0,008; 49,625); 1,000	0,62 [§] (0,036; 10,691); 1,000	-0,2 (-10,3; 9,9); 1,000
Adalimumab	153	1 (0,65)			
Subpopulation A3					
Baricitinib	170	0 (0,00)	NE	NE	NE
Adalimumab	124	0 (0,00)			
mITT					
Baricitinib	487	2 (0,41)	1,36 (0,070; 80,281); 1,000	1,34 [§] (0,133; 13,502); 1,000	0,1 (-6,9; 7,1); 1,000
Adalimumab	330	1 (0,30)			
Heterogenität^f -					

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

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR^a (95% KI); p-Wert^c	RR^a (95% KI); p-Wert^c	ARD^b (95% KI); p-Wert^d
Malignität					
Woche 24					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	1 (0,41)	1,90 (0,033; Inf.); 1,000	5,10 ^g (0,091; 285,686); 1,000	0,4 (-9,7; 10,5); 1,000
Adalimumab	153	0 (0,00)			
Subpopulation A3					
Baricitinib	170	0 (0,00)	NE	NE	NE
Adalimumab	124	0 (0,00)			
mITT					
Baricitinib	487	2 (0,41)	3,40 (0,195; Inf.); 0,518	5,36 ^g (0,318; 90,546); 0,518	0,4 (-6,6; 7,4); 0,657
Adalimumab	330	0 (0,0)			
Heterogenität^f -					
Woche 52					
Subpopulation A1 Nicht zutreffend.					
Subpopulation A2					
Baricitinib	243	2 (0,82)	3,18 (0,181; Inf.); 0,525	5,12 ^g (0,296; 88,559); 0,525	0,8 (-9,3; 10,9); 0,691
Adalimumab	153	0 (0,00)			
Subpopulation A3					
Baricitinib	170	0 (0,00)	NE	NE	NE
Adalimumab	124	0 (0,00)			
mITT					
Baricitinib	487	3 (0,62)	4,78 (0,396; Inf.); 0,277	5,37 ^g (0,534; 54,094); 0,277	0,6 (-6,4; 7,6); 0,401
Adalimumab	330	0 (0,0)			
Heterogenität^f -					

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	N	Patienten mit Ereignis n (%)	OR ^a (95% KI); p-Wert ^c	RR ^a (95% KI); p-Wert ^c	ARD ^b (95% KI); p-Wert ^d
<p>a: Ein Wert < 1 spricht für eine Behandlung mit Baricitinib; ein Wert > 1 spricht für eine Behandlung mit Adalimumab.</p> <p>b: Ein Wert < 0 spricht für eine Behandlung mit Baricitinib; ein Wert > 0 spricht für eine Behandlung mit Adalimumab.</p> <p>c: Fisher´s Test.</p> <p>d: normal Z-Test (mit Kontinuität Korrektur).</p> <p>e: Interaktions-p-Wert mittels Logistischem Regressionsmodell berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>f: kein Heterogenitätstest durchgeführt aufgrund zu geringer oder keiner Ereignisse.</p> <p>g: Peto´s Odds Ratio.</p> <p>Für die UE von besonderem Interesse Gastrointestinale Perforation und Allergische Reaktionen/Hypersensitivität traten keine Ereignisse, respektive ein Ereignis, auf.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [60].</p>					

In **Subpopulation A2** unterschieden sich die UE von besonderem Interesse (Infektionen, schwere Infektionen, kardiovaskuläre Ereignisse und Malignität) in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

In **Subpopulation A3** unterschieden sich die UE von besonderem Interesse (Infektionen, schwere Infektionen, kardiovaskuläre Ereignisse und Malignität) in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

Da der Heterogenitätstest zu Woche 24 und Woche 52 für das UE von besonderem Interesse Infektionen sowie zu Woche 52 für das UE von besonderem Interesse schwere Infektionen über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

Aufgrund zu geringer oder keiner Ereignisse konnten zu Woche 24 und Woche 52 keine Heterogenitätstests für die UE von besonderem Interesse Kardiovaskuläre Ereignisse und Malignität durchgeführt werden.

In der **mITT Population** unterschieden sich die UE von besonderem Interesse (Infektionen, schwere Infektionen, kardiovaskuläre Ereignisse und Malignität) in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander (Tabelle 4-94).

Tabelle 4-95: Ergebnisse für Depressionen QIDS-SR16 aus RCT mit dem zu bewertenden Arzneimittel

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges' g (95% KI)
Depressionen QIDS-SR16									
Woche 24									
Subpopulation A1 Nicht zutreffend.									
Subpopulation A2									
Baricitinib	243	242	7,8 (5,1)	211	5,0 (3,8)	211	-3,1 (0,2)	-0,4 (-1,2; 0,4); 0,321	-0,11 (-0,33; 0,11)
Adalimumab	153	152	9,1 (5,6)	126	5,9 (4,6)	126	-2,7 (0,3)		
Subpopulation A3									
Baricitinib	170	170	7,1 (4,4)	152	4,7 (3,5)	152	-2,4 (0,3)	0,0 (-0,8; 0,9); 0,907	0,01 (-0,23; 0,26)
Adalimumab	124	124	7,6 (4,8)	104	4,8 (4,0)	104	-2,5 (0,2)		
mITT									
Baricitinib	487	486	7,6 (4,8)	427	5,0 (3,7)	427	-2,7 (0,2)	-0,0 (-0,5; 0,5); 0,937	-0,01 (-0,16; 0,15)
Adalimumab	330	329	8,3 (5,4)	271	5,3 (4,2)	271	-2,7 (0,2)		
Heterogenität^b p = 0,469									
Woche 52									
Subpopulation A1 Nicht zutreffend.									
Subpopulation A2									
Baricitinib	243	242	7,8 (5,1)	198	5,0 (4,0)	198	-3,2 (0,2)	0,2 (-0,7; 1,0); 0,706	0,04 (-0,19; 0,28)
Adalimumab	153	152	9,1 (5,6)	110	5,2 (3,8)	110	-3,4 (0,3)		
Subpopulation A3									
Baricitinib	170	170	7,1 (4,4)	141	4,9 (4,1)	141	-2,6 (0,2)	0,2 (-0,4; 0,8); 0,449	0,06 (-0,10; 0,22)
Adalimumab	124	124	7,6 (4,8)	96	4,5 (4,2)	96	-2,9 (0,2)		

RA-BEAM (JADV)	N	Baseline		Woche 24/52		Veränderung von Baseline bis Woche 24/52		Baricitinib vs. Adalimumab	
		n	MW (SD)	n	MW (SD)	n	LSM (SE) ^a	LSMD ^a (95% KI); p-Wert	Hedges'g (95% KI)
Depressionen QIDS-SR16									
mITT									
Baricitinib	487	486	7,6 (4,8)	397	5,1 (4,1)	397	-2,6 (0,2)	0,2 (-0,4; 0,8); 0,449	0,06 (-0,10; 0,22)
Adalimumab	330	329	8,3 (5,4)	241	5,0 (4,0)	241	-2,9 (0,2)		
Heterogenität^b								p = 0,390	
<p>a: p-Werte, LSMD, LSM, SE und 95% KI wurden mittels eines ANCOVA Modells (mLOCF) berechnet.</p> <p>b: Interaktions-p-Wert mittels eines ANCOVA Modells (mLOF) berechnet; Heterogenitätstest über alle Subpopulationen hinweg.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [60].</p>									

In **Subpopulation A2** unterschieden sich die UE von besonderem Interesse (Depressionen QIDS-SR16) in den beiden Behandlungsarmen von Baseline bis Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

In **Subpopulation A3** unterschieden sich die UE von besonderem Interesse (Depressionen QIDS-SR16) in den beiden Behandlungsarmen von Baseline bis Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

Da der Heterogenitätstest für den Endpunkt UE von besonderem Interesse Depressionen QIDS-SR16 über alle Subpopulationen hinweg unauffällig war, wird nicht von einer Heterogenität des Ergebnisses zwischen den Subpopulationen ausgegangen. Im Folgenden werden daher die Ergebnisse der mITT Population betrachtet und die dort beobachteten Effekte auf die Subpopulationen übertragen.

In der **mITT Population** unterschieden sich die UE von besonderem Interesse (Depressionen QIDS-SR16) in den beiden Behandlungsarmen von Baseline bis Woche 24 und Woche 52 nicht statistisch signifikant voneinander (Tabelle 4-95).

Sofern die vorliegenden Studien bzw. Daten für eine Meta-Analyse medizinisch und methodisch 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, warum eine Meta-Analyse durchgeführt wurde bzw. warum eine Meta-Analyse nicht durchgeführt wurde bzw. warum einzelne Studien ggf. nicht in die Meta-Analyse einbezogen wurden. Machen Sie auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Eine quantitative Zusammenfassung der Ergebnisse im Rahmen einer Meta-Analyse wurde nicht durchgeführt, da die Bewertung auf der Grundlage einer klinischen Studie erfolgt.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus RCT mit dem zu bewertenden Arzneimittel fortlaufend in einem eigenen Abschnitt dar.

4.3.1.3.5 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. 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.

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

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

Für die Interpretation der Ergebnisse der Subgruppenanalysen sollte Folgendes berücksichtigt werden:

Zur Durchführung einer Metaanalyse über verschiedene Studien hinweg gibt der G-BA Subgruppenanalysen für die einzelnen Studien vor. Grenzwerte für den Interaktions-p-Wert sind laut IQWiG-Methodenpapier 4.2 nur in Zusammenhang mit Metaanalysen erwähnt. Hier wird der Grenzwert von $< 0,2$ als Hinweis auf eine Effektmodifikation verstanden. Das Methodenpapier des IQWiG beschreibt in Bezug auf einzelne Studien, dass

„... im Allgemeinen post-hoc durchgeführte Subgruppenanalysen auf Studienebene kritisch zu interpretieren sind...“ [6].

Bei der Beurteilung der Ergebnisse von Subgruppenanalysen sollte ferner die mit multiplem Testen in Zusammenhang stehende α -Fehler-Kumulierung, die die Zunahme der Wahrscheinlichkeit für Fehler 1. Art (falsch positive Testergebnisse) beschreibt, berücksichtigt werden. Auch sollten bei der Interpretation der Resultate kleine Fallzahlen innerhalb der Subgruppen berücksichtigt werden.

Im vorliegenden Fall handelt es sich nur um eine Studie, sodass keine Metaanalyse durchgeführt werden kann. In diesem Kontext gibt es daher verschiedene Limitierungen für die Interpretation der Ergebnisse der Subgruppenanalysen auf Ebene dieser einzelnen Studie.

In Anlehnung an den Entwurf des IQWiG-Methodenpapier 5.0 [151], der ausschließlich einen Grenzwert von 0,05 als Beleg für eine Effektmodifikation berücksichtigt, werden im vorliegenden Dossier lediglich die Details von Subgruppenanalysen mit einem Interaktions-p-Wert von $< 0,05$ dargestellt.

Nach den Anforderungen der Verfo wurden Interaktionstests für:

- Geschlecht
(weiblich, männlich)
- Alter
(< 65 , ≥ 65 Jahre)
- Krankheitsschwere bzw. –stadium
- Zentrums- und Ländereffekte

sowie alle *a priori* geplanten Subgruppen durchgeführt, um zu überprüfen, ob der Therapieeffekt über die Subgruppen hinweg konstant ist oder ob es Hinweise für eine Effektmodifikation gibt (Abschnitt 4.2.5.5).

Im Folgenden werden die p-Werte der Interaktionstests der präspezifizierten Subgruppenanalysen gemäß Studienbericht, sowie gem. Verfo für alle patientenrelevanten Endpunkte von Baseline bis Woche 52 angegeben.

Wurden in einer Subgruppe durch einen Interaktionstest basierend auf dem vorgegebenen p-Wert von $< 0,05$ signifikant unterschiedliche Effekte ermittelt, so werden die Ergebnisse auf Endpunktebene diesbezüglich präzisiert aufgeführt, verglichen und bewertet. Da es sich um eine einzelne Studie und keine Metaanalyse handelt, sind die Ergebnisse, wie oben dargestellt, kritisch zu interpretieren.

Es wurden für 43 (Unter-) Endpunkte bis zu 18 Subgruppenanalysen durchgeführt. Unter der Annahme, dass keine Subgruppeneffekte vorliegen, wären daher zufallsbedingt ca. 38 – 39 Belege ($p < 0,05$) pro Population für eine Effektmodifikation zu erwarten.

In der mITT Population zeigten sich 37 Belege, in der Subpopulation A2 18 Belege sowie in der Subpopulation A3 9 Belege für eine Effektmodifikation.

Die Anzahl der beobachteten Effektmodifikationen liegt somit im oder deutlich unterhalb des Bereiches der zu erwarteten Ergebnisse. Auch sind keine auffälligen Häufungen über die Endpunkte oder Subgruppen hinweg zu beobachten. Auf Grund dessen wird insgesamt nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Die Daten der Ergebnisse der Subgruppenanalysen sind dem Studienprotokoll, dem statistischen Analyseplan, dem Studienbericht und dem *post-hoc* Nachberechnungsdokument zur Studie RA-BEAM (JADV) zu entnehmen [7-9,11].

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

Interaktionstests in der Subpopulation A2

Tabelle 4-96: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A2 von Baseline bis Woche 52 (1)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a				
	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)	Ethnie (Kaukasisch; Andere)
Nutzendimension					
Endpunkt					
Mortalität^b					
Morbidität					
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,491	0,308	0,548	0,791	0,128
Veränderung der Anzahl der geschwollenen Gelenke	0,232	0,574	0,991	0,803	0,339
Veränderung des DAS28-hsCRP	0,306	0,444	0,949	0,691	0,731
Anteil Patienten mit DAS28-hsCRP $\leq 3,2$	0,7156	0,6302	0,7446	0,8315	0,9151
Anteil Patienten mit DAS28-hsCRP $< 2,6$	0,1930	0,5659	0,8277	0,8735	0,4406
Veränderung des SDAI	0,201	0,184	0,640	0,896	0,330
Anteil Patienten mit SDAI $\leq 3,3$	0,2104	0,6190	0,9805	0,3657	0,5518
Veränderung des CDAI	0,148	0,281	0,706	0,899	0,266
Anteil Patienten mit CDAI ≤ 10	0,1642	0,9458	0,5279	0,9978	0,8517
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,6466	0,5802	0,6783	0,7944	0,9876

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a				
	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)	Ethnie (Kaukasisch; Andere)
Veränderung des PAP (VAS)	0,004 [†]	0,204	0,357	0,883	0,783
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,0395 [†]	0,1411	0,2791	0,9266	0,6954
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,0254 [†]	0,0056 [†]	0,2785	0,9447	0,5399
Veränderung des PtGADA (VAS)	0,021 [†]	0,404	0,720	0,849	0,982
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,3968	0,1081	0,8622	0,9651	0,8915
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,0969	0,0175 [†]	0,8200	0,7076	0,1353
Veränderung des HAQ-DI	0,600	0,380	0,434	0,905	0,898
Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$	0,7132	0,2112	0,7846	0,6340	0,9100
Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$	0,4442	0,2658	0,9360	0,7001	0,7547
Veränderung der Dauer der Morgensteifigkeit (min)	0,440	0,261	0,607	0,441	0,843
Veränderung der EQ-5D VAS	0,554	0,029 [†]	0,267	0,104	0,310
Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm	0,4594	0,1592	0,1968	0,3836	0,9602
Veränderung der schwersten Müdigkeit	0,068	0,577	0,365	0,553	0,675

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a				
Merkmal	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)	Ethnie (Kaukasisch; Andere)
Lebensqualität					
Veränderung des SF-36 v2 akut PCS	0,185	0,547	0,488	0,479	0,962
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,7711	0,9662	0,9455	0,3120	0,9334
Veränderung des SF-36 v2 akut MCS	0,510	0,384	0,703	0,607	0,771
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	0,6386	0,3035	0,2270	0,0569	0,7849
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Funktionsfähigkeit</i>	0,393	0,940	0,194	0,533	0,716
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Rollenfunktion</i>	0,318	0,646	0,149	0,502	0,795
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Schmerzen</i>	0,173	0,784	0,758	0,595	0,791
Veränderung der SF-36 v2 akut Domäne <i>Allgemeiner Gesundheitszustand</i>	0,109	0,234	0,058	0,985	0,061
Veränderung der SF-36 v2 akut Domäne <i>Vitalität</i>	0,216	0,853	0,862	0,879	0,757
Veränderung der SF-36 v2 akut Domäne <i>Soziale Funktionsfähigkeit</i>	0,679	0,220	0,085	0,691	0,893
Veränderung der SF-36 v2 akut Domäne <i>Emotionale Rollenfunktion</i>	0,677	0,874	0,489	0,516	0,905
Veränderung der SF-36 v2 akut Domäne <i>Mentaler Gesundheitszustand</i>	0,261	0,420	0,853	0,678	0,728

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a				
Merkmal	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)	Ethnie (Kaukasisch; Andere)
Sicherheit					
SUE	0,9687	0,7216	NE	NE	NE
UE	0,5432	0,5150	0,2963	0,1153	0,9844
Schwere UE (ohne ausschließliche Laborereignisse)	NE	NE	NE	NE	NE
Studienabbruch aufgrund von UE	NE	NE	NE	NE	NE
Therapieabbruch aufgrund von UE	NE	NE	NE	NE	NE
UE von besonderem Interesse					
Infektionen	0,3052	0,2172	0,8281	0,3109	0,4470
Depressionen: QIDS-SR16	0,178	0,775	0,507	0,385	0,439
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechrates und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>†: $p < 0,05$</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>					

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

Tabelle 4-97: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A2 von Baseline bis Woche 52 (2)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Region (US, Kanada und restliche Welt; Zentral- und Südamerika und Mexiko; Europa, Asien)	spezifische Region I (Europa, Andere)	spezifische Region II (Westeuropa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Nutzendimension				
Endpunkt				
<i>Mortalität^b</i>				
<i>Morbidität</i>				
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,779	0,332	0,853	0,149
Veränderung der Anzahl der geschwollenen Gelenke	0,936	0,772	0,466	0,361
Veränderung des DAS28-hsCRP	0,313	0,087	0,989	0,581
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	0,3689	0,1450	0,8700	0,9618
Anteil Patienten mit DAS28-hsCRP < 2,6	0,3915	0,1202	0,9677	0,9513
Veränderung des SDAI	0,629	0,179	0,674	0,332
Anteil Patienten mit SDAI ≤ 3,3	0,1651	0,0241 [†]	0,7442	0,9615
Veränderung des CDAI	0,635	0,192	0,666	0,240
Anteil Patienten mit CDAI ≤ 10	0,2704	0,0803	0,5667	0,8245
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,4075	0,1703	0,9140	0,9092

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Region (US, Kanada und restliche Welt; Zentral- und Südamerika und Mexiko; Europa, Asien)	spezifische Region I (Europa, Andere)	spezifische Region II (Westeuropa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Veränderung des PAP (VAS)	0,189	0,059	0,543	0,825
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,4826	0,4511	0,9122	0,1889
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,1015	0,4420	0,9285	0,9108
Veränderung des PtGADA (VAS)	0,214	0,039 [†]	0,858	0,930
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,8337	0,5900	0,7659	0,9211
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,9682	0,6475	0,9466	0,9690
Veränderung des HAQ-DI	0,217	0,074	0,666	0,868
Anteil Patienten mit einer Verbesserung HAQ-DI ≥ 0,22	0,7666	0,9013	0,6928	0,7301
Anteil Patienten mit einer Verbesserung HAQ-DI ≥ 0,3	0,7489	0,9298	0,9822	0,6693
Veränderung der Dauer der Morgensteifigkeit (min)	0,691	0,637	0,915	0,442
Veränderung der EQ-5D VAS	0,507	0,244	0,586	0,805
Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm	0,9726	0,9340	0,5050	0,9762
Veränderung der schwersten Müdigkeit	0,996	0,990	0,185	0,358

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Region (US, Kanada und restliche Welt; Zentral- und Südamerika und Mexiko; Europa, Asien)	spezifische Region I (Europa, Andere)	spezifische Region II (Westeuropa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Lebensqualität				
Veränderung des SF-36 v2 akut PCS	0,386	0,180	0,374	0,993
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,5205	0,3556	0,1223	0,9228
Veränderung des SF-36 v2 akut MCS	0,381	0,721	0,860	0,536
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	0,7902	0,9362	NE	0,0260 [†]
Veränderung der SF-36 v2 akut Domäne Körperliche Funktionsfähigkeit	0,457	0,161	0,982	0,883
Veränderung der SF-36 v2 akut Domäne Körperliche Rollenfunktion	0,897	0,491	0,205	0,817
Veränderung der SF-36 v2 akut Domäne Körperliche Schmerzen	0,289	0,168	0,269	0,923
Veränderung der SF-36 v2 akut Domäne Allgemeiner Gesundheitszustand	0,272	0,982	0,482	0,465
Veränderung der SF-36 v2 akut Domäne Vitalität	0,320	0,459	0,751	0,824
Veränderung der SF-36 v2 akut Domäne Soziale Funktionsfähigkeit	0,174	0,099	0,987	0,737
Veränderung der SF-36 v2 akut Domäne Emotionale Rollenfunktion	0,484	0,654	0,525	0,569
Veränderung der SF-36 v2 akut Domäne Mentaler Gesundheitszustand	0,635	0,685	0,866	0,631

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Region (US, Kanada und restliche Welt; Zentral- und Südamerika und Mexiko; Europa, Asien)	spezifische Region I (Europa, Andere)	spezifische Region II (Westeuropa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Sicherheit				
SUE	NE	0,4212	NE	0,2045 ^c
UE	0,8356	0,9562	0,8501	0,3489
Schwere UE (ohne ausschließliche Laborereignisse)	NE	NE	NE	NE
Studienabbruch aufgrund von UE	NE	NE	NE	NE
Therapieabbruch aufgrund von UE	NE	NE	NE	NE
UE von besonderem Interesse				
Infektionen	0,1034	0,0591	0,2767	0,5961
Depressionen: QIDS-SR16	0,943	0,800	0,552	0,994
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechrare und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>c: Auf Grund zu geringer Patienten- bzw. Ereigniszahlen wurden die beiden untersten und die beiden obersten Subgruppenkategorien jeweils zusammengefasst: < 5 Jahre; ≥ 5 Jahre.</p> <p>†: $p < 0,05$</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>				

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

Tabelle 4-98: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A2 von Baseline bis Woche 52 (3)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP >5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF oder ACPA positiv; RF und ACPA negativ)
Nutzendimension				
Endpunkt				
Mortalität^b				
Morbidität				
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,381	0,695	0,968	0,915
Veränderung der Anzahl der geschwollenen Gelenke	0,211	0,776	0,864	0,813
Veränderung des DAS28-hsCRP	0,708	0,659	0,561	0,611
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	0,6879	0,8037	0,9554	0,6116
Anteil Patienten mit DAS28-hsCRP < 2,6	0,4934	0,7148	0,8913	NE
Veränderung des SDAI	0,297	0,519	0,616	0,912
Anteil Patienten mit SDAI ≤ 3,3	0,8501	0,1413	0,1038	NE
Veränderung des CDAI	0,357	0,447	0,651	0,887
Anteil Patienten mit CDAI ≤ 10	0,7000	0,9163	0,8142	0,8625
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,5946	0,8690	0,8821	0,5920

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP >5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF oder ACPA positiv; RF und ACPA negativ)
Veränderung des PAP (VAS)	0,886	0,065	0,115	0,853
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,1815	0,0513	0,2543	0,3904
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,1015	0,0159 [†]	0,0481 [†]	0,7411
Veränderung des PtGADA (VAS)	0,481	0,021 [†]	0,050	0,814
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,1299	0,0542	0,3309	0,2589
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,4973	0,0056 [†]	0,0135 [†]	0,9824
Veränderung des HAQ-DI	0,401	0,802	0,572	0,573
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,22	0,5794	0,6230	0,7523	0,2216
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,3	0,4905	0,6231	0,9869	0,4054
Veränderung der Dauer der Morgen- steifigkeit (min)	0,176	0,523	0,701	0,595
Veränderung der EQ-5D VAS	0,323	0,122	0,299	0,453
Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm	0,1681	0,5700	0,9871	0,4359
Veränderung der schwersten Müdigkeit	0,449	0,648	0,941	0,228

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP >5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF oder ACPA positiv; RF und ACPA negativ)
Lebensqualität				
Veränderung des SF-36 v2 akut PCS	0,358	0,397	0,255	0,607
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,2794	0,3625	0,4083	0,6672
Veränderung des SF-36 v2 akut MCS	0,886	0,659	0,296	0,195
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	0,2253	0,0482 [†]	0,0487 [†]	NE
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Funktionsfähigkeit</i>	0,363	0,578	0,679	0,994
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Rollenfunktion</i>	0,904	0,323	0,412	0,511
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Schmerzen</i>	0,468	0,987	0,821	0,691
Veränderung der SF-36 v2 akut Domäne <i>Allgemeiner Gesundheitszustand</i>	0,429	0,635	0,460	0,828
Veränderung der SF-36 v2 akut Domäne <i>Vitalität</i>	0,389	0,905	0,961	0,802
Veränderung der SF-36 v2 akut Domäne <i>Soziale Funktionsfähigkeit</i>	0,810	0,868	0,934	0,637
Veränderung der SF-36 v2 akut Domäne <i>Emotionale Rollenfunktion</i>	0,494	0,848	0,882	0,371
Veränderung der SF-36 v2 akut Domäne <i>Mentaler Gesundheitszustand</i>	0,999	0,548	0,137	0,087

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP >5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF oder ACPA positiv; RF und ACPA negativ)
Sicherheit				
SUE	NE	NE	NE	NE
UE	0,1882	0,1019	0,0592	0,6342
Schwere UE (ohne ausschließliche Laborereignisse)	NE	NE	NE	NE
Studienabbruch aufgrund von UE	NE	NE	NE	NE
Therapieabbruch aufgrund von UE	NE	NE	NE	NE
UE von besonderem Interesse				
Infektionen	0,1115	0,3892	0,4097	0,8160
Depressionen: QIDS-SR16	0,102	0,948	0,810	0,862
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechrates und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>†: p < 0,05</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>				

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

Tabelle 4-99: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A2 von Baseline bis Woche 52 (4)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Seropositivität (RF positiv; RF negativ)	Seropositivität (ACPA positiv; ACPA negativ)	mTSS zu Baseline (≤ Median; > Median)	Therapie mit Kortikosteroiden zu Baseline (Ja; Nein)
Nutzendimension Endpunkt				
Mortalität^b				
Morbidität				
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,574	1,000	0,621	0,788
Veränderung der Anzahl der geschwollenen Gelenke	0,482	0,714	0,894	0,673
Veränderung des DAS28-hsCRP	0,939	0,438	0,359	0,764
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	0,7590	0,7250	0,9188	0,2866
Anteil Patienten mit DAS28-hsCRP < 2,6	0,8304	0,5688	0,3607	0,5949
Veränderung des SDAI	0,751	0,805	0,455	0,963
Anteil Patienten mit SDAI ≤ 3,3	NE	NE	0,5056	0,1600
Veränderung des CDAI	0,531	0,700	0,536	0,971
Anteil Patienten mit CDAI ≤ 10	0,6171	0,8887	0,8271	0,5680
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,9343	0,6875	0,9407	0,2869

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Seropositivität (RF positiv; RF negativ)	Seropositivität (ACPA positiv; ACPA negativ)	mTSS zu Baseline (≤ Median; > Median)	Therapie mit Kortikosteroiden zu Baseline (Ja; Nein)
Veränderung des PAP (VAS)	0,926	0,964	0,164	0,813
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,6844	0,8506	0,5933	0,8353
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,9380	0,9682	0,1736	0,6490
Veränderung des PtGADA (VAS)	0,580	0,557	0,232	0,679
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,2114	0,1726	0,1225	0,9055
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,8523	0,4479	0,0611	0,9423
Veränderung des HAQ-DI	0,161	0,383	0,961	0,620
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,22	0,0733	0,2992	0,6554	0,3051
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,3	0,1602	0,4594	0,8162	0,2160
Veränderung der Dauer der Morgensteifigkeit (min)	0,513	0,138	0,911	0,068
Veränderung der EQ-5D VAS	0,877	0,703	0,975	0,243
Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm	0,2589	0,3982	0,5442	0,5711
Veränderung der schwersten Müdigkeit	0,589	0,650	0,364	0,153

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Seropositivität (RF positiv; RF negativ)	Seropositivität (ACPA positiv; ACPA negativ)	mTSS zu Baseline (≤ Median; > Median)	Therapie mit Kortikosteroiden zu Baseline (Ja; Nein)
Lebensqualität				
Veränderung des SF-36 v2 akut PCS	0,490	0,785	0,311	0,743
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,5782	0,6749	0,7579	0,2459
Veränderung des SF-36 v2 akut MCS	0,096	0,426	0,263	0,182
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	0,2032	0,6913	0,0244 [†]	0,4027
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Funktionsfähigkeit</i>	0,680	0,551	0,095	0,589
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Rollenfunktion</i>	0,910	0,299	0,596	0,812
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Schmerzen</i>	0,909	0,962	0,659	0,726
Veränderung der SF-36 v2 akut Domäne <i>Allgemeiner Gesundheitszustand</i>	0,428	0,945	0,574	0,437
Veränderung der SF-36 v2 akut Domäne <i>Vitalität</i>	0,563	0,651	0,460	0,128
Veränderung der SF-36 v2 akut Domäne <i>Soziale Funktionsfähigkeit</i>	0,465	0,437	0,937	0,263
Veränderung der SF-36 v2 akut Domäne <i>Emotionale Rollenfunktion</i>	0,247	0,342	0,808	0,313
Veränderung der SF-36 v2 akut Domäne <i>Mentaler Gesundheitszustand</i>	0,079	0,304	0,234	0,608

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Seropositivität (RF positiv; RF negativ)	Seropositivität (ACPA positiv; ACPA negativ)	mTSS zu Baseline (≤ Median; > Median)	Therapie mit Kortikosteroide zu Baseline (Ja; Nein)
Sicherheit				
SUE	NE	NE	NE	NE
UE	0,1304	0,5735	0,5093	0,4797
Schwere UE (ohne ausschließliche Laborereignisse)	NE	NE	NE	NE
Studienabbruch aufgrund von UE	NE	NE	NE	NE
Therapieabbruch aufgrund von UE	NE	NE	NE	0,3796
UE von besonderem Interesse				
Infektionen	0,8970	0,5109	0,3557	0,4396
Depressionen: QIDS-SR16	0,211	0,839	0,610	0,068
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechrates und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>†: p < 0,05</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>				

Interaktionstests in der Subpopulation A3

Tabelle 4-100: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A3 von Baseline bis Woche 52 (1)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)
Nutzendimension				
Endpunkt				
Mortalität^b				
Morbidität				
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,520	0,489	0,581	0,118
Veränderung der Anzahl der geschwollenen Gelenke	0,105	0,317	0,328	0,013 [†]
Veränderung des DAS28-hsCRP	0,231	0,751	0,419	0,192
Anteil Patienten mit DAS28-hsCRP $\leq 3,2$	0,8094	0,7231	0,7043	0,1517
Anteil Patienten mit DAS28-hsCRP $< 2,6$	0,6178	0,9276	0,5129	0,1145
Veränderung des SDAI	0,260	0,733	0,768	0,110
Anteil Patienten mit SDAI $\leq 3,3$	NE	0,5563	0,9006	0,2548
Veränderung des CDAI	0,191	0,755	0,513	0,107
Anteil Patienten mit CDAI ≤ 10	0,8693	0,4668	0,9150	0,1325
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,9407	0,6638	0,9143	0,1716

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)
Veränderung des PAP (VAS)	0,284	0,101	0,412	0,617
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,2028	0,2055	0,3208	0,8187
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,1189	0,1151	0,7178	0,8526
Veränderung des PtGADA (VAS)	0,243	0,150	0,466	0,691
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,1297	0,3699	0,9237	0,9341
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,6438	0,1496	0,5318	0,3274
Veränderung des HAQ-DI	0,627	0,382	0,568	0,379
Anteil Patienten mit einer Verbesserung HAQ-DI ≥ 0,22	0,6065	0,3030	0,7803	0,9009
Anteil Patienten mit einer Verbesserung HAQ-DI ≥ 0,3	0,3774	0,1581	0,6222	0,5407
Veränderung der Dauer der Morgensteifigkeit (min)	0,201	0,096	0,367	0,982
Veränderung der EQ-5D VAS	0,496	1,000	0,195	0,576
Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm	0,2578	0,4258	0,1554	0,0694
Veränderung der schwersten Müdigkeit	0,119	0,184	0,732	0,781

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)
Lebensqualität				
Veränderung des SF-36 v2 akut PCS	0,870	0,231	0,702	0,819
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,5949	0,5470	0,0973	0,1633
Veränderung des SF-36 v2 akut MCS	0,368	0,321	0,161	0,251
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	0,3162	NE	0,5135	0,3242
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Funktionsfähigkeit</i>	0,231	0,615	0,965	0,424
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Rollenfunktion</i>	0,745	0,515	0,178	0,191
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Schmerzen</i>	0,418	0,293	0,684	0,681
Veränderung der SF-36 v2 akut Domäne <i>Allgemeiner Gesundheitszustand</i>	0,870	0,590	0,672	0,731
Veränderung der SF-36 v2 akut Domäne <i>Vitalität</i>	0,858	0,923	0,433	0,929
Veränderung der SF-36 v2 akut Domäne <i>Soziale Funktionsfähigkeit</i>	0,864	0,753	0,076	0,577
Veränderung der SF-36 v2 akut Domäne <i>Emotionale Rollenfunktion</i>	0,714	0,333	0,331	0,291
Veränderung der SF-36 v2 akut Domäne <i>Mentaler Gesundheitszustand</i>	0,055	0,685	0,373	0,121

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)
Sicherheit				
SUE	NE	NE	NE	NE
UE	0,8802	0,3897	0,7023	0,9304
Studienabbruch aufgrund von UE	NE	NE	NE	NE
Therapieabbruch aufgrund von UE	NE	NE	NE	NE
UE von besonderem Interesse				
Infektionen	0,8240	0,5309	0,5539	0,7128
Depressionen: QIDS-SR16	0,658	0,136	0,993	0,181
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechrare und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>†: p < 0,05</p> <p>Auf Grund zu geringer Fallzahlen (n=9) wurden für den Endpunkt Schwere UE (ohne ausschließliche Laborereignisse) keine Subgruppenanalysen durchgeführt. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>				

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

Tabelle 4-101: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A3 von Baseline bis Woche 52 (2)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Ethnie (Kaukasisch; Andere)	Region (US, Kanada und restliche Welt; Zentral- und Südamerika und Mexiko; Europa, Asien)	spezifische Region I (Europa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Nutzendimension Endpunkt				
Mortalität^b				
Morbidität				
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,482	0,103	0,510	0,560
Veränderung der Anzahl der geschwollenen Gelenke	0,591	0,146	0,195	0,507
Veränderung des DAS28-hsCRP	0,438	0,153	0,514	0,423
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	0,0625	0,4423	0,3558	0,2045 ^c
Anteil Patienten mit DAS28-hsCRP < 2,6	0,2217	0,9290	0,8445	0,2306 ^c
Veränderung des SDAI	0,741	0,130	0,346	0,568
Anteil Patienten mit SDAI ≤ 3,3	0,7705	0,6680 ^d	0,3676	0,8585 ^c
Veränderung des CDAI	0,726	0,138	0,443	0,645
Anteil Patienten mit CDAI ≤ 10	0,0287 [†]	0,2657	0,6040	0,1339
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,0366 [†]	0,3154	0,2986	0,1188 ^c

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Ethnie (Kaukasisch; Andere)	Region (US, Kanada und restliche Welt; Zentral- und Südamerika und Mexiko; Europa, Asien)	spezifische Region I (Europa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Veränderung des PAP (VAS)	0,417	0,388	0,764	0,581
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,7101	0,7141	0,7540	0,3091
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,8662	0,9673	0,7289	0,7833
Veränderung des PtGADA (VAS)	0,464	0,425	0,907	0,664
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,1912	0,6100	0,3861	0,7905
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,4525	0,9151	0,6635	0,9396
Veränderung des HAQ-DI	0,326	0,945	0,706	0,531
Anteil Patienten mit einer Verbesserung HAQ-DI ≥ 0,22	0,4363	0,3122	0,3275	0,6007
Anteil Patienten mit einer Verbesserung HAQ-DI ≥ 0,3	0,4001	0,4270	0,2281	0,3869
Veränderung der Dauer der Morgensteifigkeit (min)	0,827	0,217	0,233	0,733
Veränderung der EQ-5D VAS	0,696	0,907	0,400	0,775
Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm	0,9250	0,5170	0,8348	0,9044
Veränderung der schwersten Müdigkeit	0,223	0,490	0,741	0,571

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Ethnie (Kaukasisch; Andere)	Region (US, Kanada und restliche Welt; Zentral- und Südamerika und Mexiko; Europa, Asien)	spezifische Region I (Europa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Lebensqualität				
Veränderung des SF-36 v2 akut PCS	0,392	0,792	0,935	0,483
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,9658	0,8191	0,3873	0,3695
Veränderung des SF-36 v2 akut MCS	0,789	0,501	0,260	0,540
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	0,9266	0,9631	0,7492	0,9837 ^c
Veränderung der SF-36 v2 akut Domäne Körperliche Funktionsfähigkeit	0,351	0,671	0,594	0,557
Veränderung der SF-36 v2 akut Domäne Körperliche Rollenfunktion	0,732	0,926	0,960	0,366
Veränderung der SF-36 v2 akut Domäne Körperliche Schmerzen	0,314	0,344	0,529	0,973
Veränderung der SF-36 v2 akut Domäne Allgemeiner Gesundheitszustand	0,641	0,507	0,375	0,436
Veränderung der SF-36 v2 akut Domäne Vitalität	0,912	0,778	0,509	0,862
Veränderung der SF-36 v2 akut Domäne Soziale Funktionsfähigkeit	0,941	0,548	0,489	0,591
Veränderung der SF-36 v2 akut Domäne Emotionale Rollenfunktion	0,771	0,629	0,870	0,511
Veränderung der SF-36 v2 akut Domäne Mentaler Gesundheitszustand	0,927	0,509	0,174	0,487

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Ethnie (Kaukasisch; Andere)	Region (US, Kanada und restliche Welt; Zentral- und Südamerika und Mexiko; Europa, Asien)	spezifische Region I (Europa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Sicherheit				
SUE	NE	NE	NE	NE
UE	0,9853	0,4982	0,3248	0,4233 ^c
Studienabbruch aufgrund von UE	NE	NE	NE	NE
Therapieabbruch aufgrund von UE	NE	NE	NE	NE
UE von besonderem Interesse				
Infektionen	0,2307	0,5178	0,2988	0,3527 ^c
Depressionen: QIDS-SR16 (mLOCF)	0,813	0,607	0,469	0,955
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechrates und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>c: Auf Grund zu geringer Patienten- bzw. Ereigniszahlen wurden die beiden untersten Subgruppenkategorien zusammengefasst: < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre.</p> <p>d: Auf Grund zu geringer Patienten- bzw. Ereigniszahlen wurden die folgenden Subgruppenkategorien gebildet: Europa; Asien; Rest der Welt.</p> <p>†: p < 0,05</p> <p>Auf Grund zu geringer Fallzahlen (n=9) wurden für den Endpunkt Schwere UE (ohne ausschließliche Laborereignisse) keine Subgruppenanalysen durchgeführt. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>				

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

Tabelle 4-102: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A3 von Baseline bis Woche 52 (3)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP >5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF positiv; RF negativ)
Nutzendimension Endpunkt				
Mortalität^b				
Morbidität				
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,896	0,693	0,749	0,424
Veränderung der Anzahl der geschwollenen Gelenke	0,707	0,450	0,448	0,783
Veränderung des DAS28-hsCRP	0,433	0,456	0,314	0,987
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	0,3601	0,4749	0,3426	0,6643
Anteil Patienten mit DAS28-hsCRP < 2,6	0,5136	0,0937	0,0471 [†]	0,4069
Veränderung des SDAI	0,496	0,857	0,751	0,725
Anteil Patienten mit SDAI ≤ 3,3	0,2778	0,2481	0,2888	NE
Veränderung des CDAI	0,683	0,772	0,697	0,829
Anteil Patienten mit CDAI ≤ 10	0,8184	0,4881	0,3331	0,6539
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,5503	0,4198	0,2955	0,6305

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP >5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF positiv; RF negativ)
Veränderung des PAP (VAS)	0,361	0,948	0,384	0,349
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,3891	0,0721	0,0124 [†]	0,0852
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,5244	0,5308	0,3874	0,4629
Veränderung des PtGADA (VAS)	0,179	0,954	0,465	0,314
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,3477	0,1178	0,0250 [†]	0,2028
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,6057	0,9628	0,5276	0,1793
Veränderung des HAQ-DI	0,449	0,536	0,507	0,770
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,22	0,9874	0,4898	0,9796	NE
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,3	0,3872	0,5286	0,7700	0,0735
Veränderung der Dauer der Morgen- steifigkeit (min)	0,449	0,636	0,509	0,339
Veränderung der EQ-5D VAS	0,517	0,971	0,753	0,501
Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm	0,6825	0,3614	0,3155	0,8149
Veränderung der schwersten Müdigkeit	0,590	0,570	0,698	0,399

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP >5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF positiv; RF negativ)
Lebensqualität				
Veränderung des SF-36 v2 akut PCS	0,449	0,534	0,626	0,643
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,3521	0,6915	0,9080	0,5385
Veränderung des SF-36 v2 akut MCS	0,824	0,408	0,729	0,633
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	0,4084	0,5972	0,8527	NE
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Funktionsfähigkeit</i>	0,228	0,596	0,711	0,716
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Rollenfunktion</i>	0,295	0,249	0,370	0,777
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Schmerzen</i>	0,627	0,721	0,811	0,400
Veränderung der SF-36 v2 akut Domäne <i>Allgemeiner Gesundheitszustand</i>	0,605	0,348	0,460	0,897
Veränderung der SF-36 v2 akut Domäne <i>Vitalität</i>	0,918	0,759	0,918	0,266
Veränderung der SF-36 v2 akut Domäne <i>Soziale Funktionsfähigkeit</i>	0,540	0,289	0,652	0,682
Veränderung der SF-36 v2 akut Domäne <i>Emotionale Rollenfunktion</i>	0,862	0,506	0,655	0,892
Veränderung der SF-36 v2 akut Domäne <i>Mentaler Gesundheitszustand</i>	0,916	0,238	0,369	0,832

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP >5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF positiv; RF negativ)
Sicherheit				
SUE	NE	NE	NE	NE
UE	0,3068	0,1848	0,2977	0,8192
Studienabbruch aufgrund von UE	NE	NE	NE	NE
Therapieabbruch aufgrund von UE	NE	NE	NE	NE
UE von besonderem Interesse				
Infektionen	0,5478	0,4213	0,5739	0,6497
Depressionen: QIDS-SR16 (mLOCF)	0,558	0,827	0,607	0,734
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechrates und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>†: p < 0,05</p> <p>Auf Grund zu geringer Fallzahlen (n=9) wurden für den Endpunkt Schwere UE (ohne ausschließliche Laborereignisse) keine Subgruppenanalysen durchgeführt.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>				

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

Tabelle 4-103: Übersicht über alle durchgeführten Interaktionstests in der Subpopulation A3 von Baseline bis Woche 52 (4)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a		
	Seropositivität (ACPA positiv; ACPA negativ)	mTSS zu Baseline (≤ Median; > Median)	Therapie mit Kortikosteroide zu Baseline (Ja; Nein)
Nutzendimension			
Endpunkt			
Mortalität^b			
Morbidität			
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,718	0,788	0,632
Veränderung der Anzahl der geschwollenen Gelenke	0,117	0,988	0,887
Veränderung des DAS28-hsCRP	0,288	0,663	0,501
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	0,3255	0,6331	0,3094
Anteil Patienten mit DAS28-hsCRP < 2,6	0,5966	0,7944	0,6399
Veränderung des SDAI	0,182	0,691	0,488
Anteil Patienten mit SDAI ≤ 3,3	NE	0,8011	0,7434
Veränderung des CDAI	0,230	0,843	0,452
Anteil Patienten mit CDAI ≤ 10	0,8643	0,8661	0,9417
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,2922	0,7513	0,2807

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a		
	Seropositivität (ACPA positiv; ACPA negativ)	mTSS zu Baseline (≤ Median; > Median)	Therapie mit Kortikosteroiden zu Baseline (Ja; Nein)
Veränderung des PAP (VAS)	0,047 [†]	0,809	0,377
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,0648	0,2339	0,8937
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,1678	0,5140	0,4287
Veränderung des PtGADA	0,100	0,870	0,174
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,1314	0,2408	0,3795
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,1009	0,5086	0,4198
Veränderung des HAQ-DI	0,963	0,261	0,316
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,22	0,0075 [†]	0,1481	0,8176
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,3	0,6073	0,4898	0,2615
Veränderung der Dauer der Morgen- steifigkeit (min)	0,408	0,950	0,924
Veränderung der EQ-5D VAS	0,233	0,812	0,117
Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm	0,2410	0,7351	0,0875
Veränderung der schwersten Müdigkeit	0,452	0,705	0,056

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a		
	Seropositivität (ACPA positiv; ACPA negativ)	mTSS zu Baseline (≤ Median; > Median)	Therapie mit Kortikosteroiden zu Baseline (Ja; Nein)
Lebensqualität			
Veränderung des SF-36 v2 akut PCS	0,806	0,657	0,236
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,1769	0,4347	0,7262
Veränderung des SF-36 v2 akut MCS	0,122	0,810	0,132
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	NE	0,3188	0,8386
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Funktionsfähigkeit</i>	0,854	0,557	0,331
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Rollenfunktion</i>	0,559	0,326	0,115
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Schmerzen</i>	0,810	0,946	0,127
Veränderung der SF-36 v2 akut Domäne <i>Allgemeiner Gesundheitszustand</i>	0,286	0,913	0,218
Veränderung der SF-36 v2 akut Domäne <i>Vitalität</i>	0,458	0,670	0,204
Veränderung der SF-36 v2 akut Domäne <i>Soziale Funktionsfähigkeit</i>	0,258	0,265	0,194
Veränderung der SF-36 v2 akut Domäne <i>Emotionale Rollenfunktion</i>	0,274	0,654	0,038 [†]
Veränderung der SF-36 v2 akut Domäne <i>Mentaler Gesundheitszustand</i>	0,331	0,628	0,177

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a		
	Merkmals	Seropositivität (ACPA positiv; ACPA negativ)	mTSS zu Baseline (≤ Median; > Median)
Sicherheit			
SUE	NE	NE	NE
UE	0,1308	0,4789	0,8910
Studienabbruch aufgrund von UE	NE	NE	NE
Therapieabbruch aufgrund von UE	NE	NE	NE
UE von besonderem Interesse			
Infektionen	0,2032	0,6366	0,1916
Depressionen: QIDS-SR16	0,327	0,159	0,478
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechrate und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>†: p < 0,05</p> <p>Auf Grund zu geringer Fallzahlen (n=9) wurden für den Endpunkt Schwere UE (ohne ausschließliche Laborereignisse) keine Subgruppenanalysen durchgeführt.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>			

Interaktionstests in der mITT

Tabelle 4-104: Übersicht über alle durchgeführten Interaktionstests in der mITT Population von Baseline bis Woche 52 (1)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a				
Merkmal	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg - < 100 kg; ≥ 100 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)	Ethnie (Kaukasisch; Andere)
Nutzendimension Endpunkt					
<i>Mortalität^b</i>					
<i>Morbidität</i>					
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,920	0,340	0,740	0,532	0,746
Veränderung der Anzahl der geschwollenen Gelenke	0,530	0,348	0,718	0,019 [†]	0,778
Veränderung des DAS28-hsCRP	0,893	0,965	0,730	0,162	0,434
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	0,8222	0,9642	0,8988	0,1145	0,1048
Anteil Patienten mit DAS28-hsCRP < 2,6	0,2774	0,8937	0,9528	0,3521	0,0901
Veränderung des SDAI	0,646	0,475	0,690	0,335	0,887
Anteil Patienten mit SDAI ≤ 3,3	0,8948	0,5942	0,9295 ^c	0,7658	0,7035
Veränderung des CDAI	0,610	0,573	0,704	0,281	0,999
Anteil Patienten mit CDAI ≤ 10	0,2157	0,5972	0,6766	0,2213	0,0712
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,8075	0,9244	0,8544	0,1077	0,0639

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a				
	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg - < 100 kg; ≥ 100 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)	Ethnie (Kaukasisch; Andere)
Veränderung des PAP (VAS)	0,143	0,229	0,691	0,525	0,196
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,5262	0,2201	0,8756	0,8352	0,7402
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,5343	0,0404 [†]	0,7741	0,7570	0,2315
Veränderung des PtGADA (VAS)	0,377	0,259	0,798	0,505	0,500
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,7140	0,2101	0,5289	0,6680	0,4726
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,3475	0,0338 [†]	0,5936	0,5121	0,9445
Veränderung des HAQ-DI	0,870	0,697	0,721	0,592	0,505
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,22	0,4231	0,2965	0,8457	0,7722	0,5258
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,3	0,1931	0,3334	0,7772	0,9018	0,7187
Veränderung der Dauer der Morgensteifigkeit (min)	0,059	0,119	0,534 ^c	0,553	0,443
Veränderung der EQ-5D (VAS)	0,614	0,070	0,841	0,252	0,282
Anteil Patienten mit einer Verbesserung der EQ-5D (VAS) ≥ 10 mm	0,8293	0,2120	0,7765	0,9395	0,7058
Veränderung der schwersten Müdigkeit	0,017 [†]	0,226	0,600 ^c	0,626	0,522

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a				
Merkmal	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg - < 100 kg; ≥ 100 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)	Ethnie (Kaukasisch; Andere)
Lebensqualität					
Veränderung des SF-36 v2 akut PCS	0,250	0,993	0,994	0,534	0,448
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,4962	0,6737	0,5054	0,5273	0,5641
Veränderung des SF-36 v2 akut MCS	0,969	0,948	0,311	0,725	0,951
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	0,5150	0,8579	0,2475	0,0745	0,5647
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Funktionsfähigkeit</i>	0,832	0,840	0,759	0,216	0,205
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Rollenfunktion</i>	0,288	0,864	0,664	0,875	0,434
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Schmerzen</i>	0,491	0,655	0,936	0,815	0,262
Veränderung der SF-36 v2 akut Domäne <i>Allgemeiner Gesundheitszustand</i>	0,057	0,398	0,046 [†]	0,743	0,240
Veränderung der SF-36 v2 akut Domäne <i>Vitalität</i>	0,534	0,869	0,864	0,909	0,645
Veränderung der SF-36 v2 akut Domäne <i>Soziale Funktionsfähigkeit</i>	0,751	0,412	0,927	0,975	0,953
Veränderung der SF-36 v2 akut Domäne <i>Emotionale Rollenfunktion</i>	0,770	0,501	0,305	0,917	0,653
Veränderung der SF-36 v2 akut Domäne <i>Mentaler Gesundheitszustand</i>	0,850	0,854	0,399	0,356	0,674

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a				
Merkmal	Geschlecht (Weiblich, Männlich)	Alter (< 65 Jahre, ≥ 65 Jahre)	Gewicht zu Baseline (< 60 kg; ≥ 60 kg - < 100 kg; ≥ 100 kg)	BMI zu Baseline (< 25 kg/m ² ; ≥ 25 kg/m ² - < 30 kg/m ² ; ≥ 30 kg/m ²)	Ethnie (Kaukasisch; Andere)
Sicherheit					
SUE	0,6273	0,3893	0,1554 ^c	0,0191 [†]	0,5570
UE	0,5507	0,6200	0,3074	0,1763	0,5459
Schwere UE (ohne ausschließliche Laborereignisse)	0,5607	0,1483	0,8439 ^c	0,7240	0,9784
Studienabbruch aufgrund von UE	0,2075	0,3022	0,1989 ^c	0,1913 ^d	0,5893
Therapieabbruch aufgrund von UE	0,3037	0,2055	0,1123 ^c	0,0810 ^d	0,9724
UE von besonderem Interesse					
Infektionen	0,9409	0,5354	0,7109	0,1026	0,9207
Depressionen: QIDS-SR16	0,221	0,181	0,031 [†]	0,141	0,708
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechrates und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>c: Auf Grund zu geringer Patienten- bzw. Ereigniszahlen wurden die beiden obersten Subgruppenkategorien zusammengefasst: < 60 kg; ≥ 60 kg.</p> <p>d: Auf Grund zu geringer Patienten- bzw. Ereigniszahlen wurden die beiden obersten Subgruppenkategorien zusammengefasst: < 25 kg/m²; ≥ 25 kg/m².</p> <p>†: p < 0,05</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>					

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

Tabelle 4-105: Übersicht über alle durchgeführten Interaktionstests in der mITT Population von Baseline bis Woche 52 (2)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Region (US und Kanada; Zentral- und Südamerika und Mexiko; Osteuropa, Westeuropa; Asien ohne Japan; Japan; Rest der Welt)	spezifische Region I (Europa, Andere)	spezifische Region II (Westeuropa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Nutzendimension Endpunkt				
<i>Mortalität^b</i>				
<i>Morbidität</i>				
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,608	0,602	0,718	0,790
Veränderung der Anzahl der geschwollenen Gelenke	0,322	0,641	0,466	0,616
Veränderung des DAS28-hsCRP	0,535	0,512	0,568	0,996
Anteil Patienten mit DAS28-hsCRP $\leq 3,2$	0,1617	0,0722	0,8964	0,8045
Anteil Patienten mit DAS28-hsCRP $< 2,6$	0,5395	0,2137	0,9430	0,8900
Veränderung des SDAI	0,312	0,777	0,339	0,896
Anteil Patienten mit SDAI $\leq 3,3$	0,8644 ^c	0,3374	0,8663	0,6963
Veränderung des CDAI	0,376	0,658	0,423	0,805
Anteil Patienten mit CDAI ≤ 10	0,1833	0,0826	0,8243	0,6322
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,1093	0,0630	0,9230	0,7130

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Region (US und Kanada; Zentral- und Südamerika und Mexiko; Osteuropa, Westeuropa; Asien ohne Japan; Japan; Rest der Welt)	spezifische Region I (Europa, Andere)	spezifische Region II (Westeuropa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Veränderung des PAP (VAS)	0,094	0,423	0,047 [†]	0,720
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,5830	0,9030	0,1532	0,2431
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,5076	0,8836	0,2771	0,7521
Veränderung des PtGADA (VAS)	0,442	0,378	0,315	0,812
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,8106	0,8345	0,6187	0,9511
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,9170	0,7365	0,6307	0,5068
Veränderung des HAQ-DI	0,858	0,380	0,730	0,365
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,22	0,4135	0,4706	0,1026	0,9282
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,3	0,7163	0,3452	0,2405	0,9486
Veränderung der Dauer der Morgensteifigkeit (min)	0,277	0,094	0,055	0,966
Veränderung der EQ-5D (VAS)	0,245	0,946	0,070	0,528
Anteil Patienten mit einer Verbesserung der EQ-5D (VAS) ≥ 10 mm	0,5924	0,7546	0,1335	0,7552
Veränderung der schwersten Müdigkeit	0,409	0,998	0,075	0,351

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Region (US und Kanada; Zentral- und Südamerika und Mexiko; Osteuropa, Westeuropa; Asien ohne Japan; Japan; Rest der Welt)	spezifische Region I (Europa, Andere)	spezifische Region II (Westeuropa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Lebensqualität				
Veränderung des SF-36 v2 akut PCS	0,681	0,743	0,318	0,597
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,1211	0,0991	0,0128 [†]	0,8337
Veränderung des SF-36 v2 akut MCS	0,246	0,839	0,615	0,618
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	0,5844	0,7884	0,7000	0,2870
Veränderung der SF-36 v2 akut Domäne Körperliche Funktionsfähigkeit	0,578	0,387	0,962	0,845
Veränderung der SF-36 v2 akut Domäne Körperliche Rollenfunktion	0,618	0,674	0,547	0,331
Veränderung der SF-36 v2 akut Domäne Körperliche Schmerzen	0,259	0,828	0,067	0,744
Veränderung der SF-36 v2 akut Domäne Allgemeiner Gesundheitszustand	0,360	0,308	0,051	0,463
Veränderung der SF-36 v2 akut Domäne Vitalität	0,777	0,822	0,732	0,973
Veränderung der SF-36 v2 akut Domäne Soziale Funktionsfähigkeit	0,327	0,356	0,583	0,584
Veränderung der SF-36 v2 akut Domäne Emotionale Rollenfunktion	0,067	0,589	0,429	0,584
Veränderung der SF-36 v2 akut Domäne Mentaler Gesundheitszustand	0,766	0,890	0,333	0,507

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Region (US und Kanada; Zentral- und Südamerika und Mexiko; Osteuropa, Westeuropa; Asien ohne Japan; Japan; Rest der Welt)	spezifische Region I (Europa, Andere)	spezifische Region II (Westeuropa, Andere)	Erkrankungsdauer (< 1 Jahre; ≥ 1 Jahre - < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre)
Sicherheit				
SUE	0,4856 ^d	0,7964	NE	0,0475 ^{†e}
UE	0,4580	0,3220	0,4555	0,4614
Schwere UE (ohne ausschließliche Laboreignisse)	NE	0,9784	NE	0,9784 ^e
Studienabbruch aufgrund von UE	NE	NE	NE	0,7656 ^e
Therapieabbruch aufgrund von UE	NE	NE	NE	0,6097 ^e
UE von besonderem Interesse				
Infektionen	0,7973	0,3112	0,6742	0,5119
Depressionen: QIDS-SR16 (mLOCF)	0,333	0,892	0,265	0,945
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechrare und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>c: Auf Grund zu geringer Patienten- bzw. Ereigniszahlen wurden die Subgruppenkategorien Asien ohne Japan, und Japan in eine Kategorie zusammengefasst: US und Kanada; Zentral- und Südamerika und Mexiko; Osteuropa, Westeuropa; Asien; Rest der Welt.</p> <p>d: Auf Grund zu geringer Patienten- bzw. Ereigniszahlen wurden die folgenden drei Subgruppenkategorien gebildet: Europa; Asien; Rest der Welt.</p> <p>e: Auf Grund zu geringer Patienten- bzw. Ereigniszahlen wurden die beiden untersten Subgruppenkategorien zusammengefasst: < 5 Jahre; ≥ 5 Jahre - < 10 Jahre; ≥ 10 Jahre.</p> <p>†: p < 0,05</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>				

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

Tabelle 4-106: Übersicht über alle durchgeführten Interaktionstests in der mITT Population von Baseline bis Woche 52 (3)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP > 5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF oder ACPA positiv; RF und ACPA negativ)
Nutzendimension				
Endpunkt				
Mortalität^b				
Morbidität				
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,434	0,405	0,608	0,961
Veränderung der Anzahl der geschwollenen Gelenke	0,170	0,355	0,346	0,882
Veränderung des DAS28-hsCRP	0,294	0,639	0,474	0,559
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	0,2275	0,7966	0,7122	0,7748
Anteil Patienten mit DAS28-hsCRP < 2,6	0,9757	0,7324	0,5731	0,5992
Veränderung des SDAI	0,153	0,341	0,352	0,785
Anteil Patienten mit SDAI ≤ 3,3	0,1848	0,1163	0,1145	NE
Veränderung des CDAI	0,252	0,302	0,380	0,955
Anteil Patienten mit CDAI ≤ 10	0,8213	0,7419	0,7283	0,9750
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,3548	0,8446	0,7614	0,7895

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP > 5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF oder ACPA positiv; RF und ACPA negativ)
Veränderung des PAP (VAS)	0,672	0,102	0,066	0,577
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,2940	0,0029 [†]	0,0173 [†]	0,5596
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,3142	0,0116 [†]	0,0207 [†]	0,9151
Veränderung des PtGADA (VAS)	0,756	0,043 [†]	0,036 [†]	0,694
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,6110	0,0046 [†]	0,0133 [†]	0,8933
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,8368	0,0074 [†]	0,0066 [†]	0,6810
Veränderung des HAQ-DI	0,996	0,844	0,930	0,495
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,22	0,9655	0,9056	0,7099	0,5580
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,3	0,5955	0,9921	0,7623	0,4578
Veränderung der Dauer der Morgen- steifigkeit (min)	0,502	0,983	0,887	0,607
Veränderung der EQ-5D (VAS)	0,778	0,291	0,642	0,305
Anteil Patienten mit einer Verbesserung der EQ-5D (VAS) ≥ 10 mm	0,7457	0,5536	0,8726	0,5836
Veränderung der schwersten Müdigkeit	0,443	0,864	0,605	0,343

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP > 5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF oder ACPA positiv; RF und ACPA negativ)
Lebensqualität				
Veränderung des SF-36 v2 akut PCS	0,983	0,525	0,448	0,982
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,7993	0,9886	0,9616	0,6364
Veränderung des SF-36 v2 akut MCS	0,812	0,174	0,739	0,050
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	0,1102	0,1240	0,0763	0,5162
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Funktionsfähigkeit</i>	0,680	0,314	0,414	0,748
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Rollenfunktion</i>	0,518	0,184	0,349	0,270
Veränderung der SF-36 v2 akut Domäne <i>Körperliche Schmerzen</i>	0,356	0,888	0,982	0,663
Veränderung der SF-36 v2 akut Domäne <i>Allgemeiner Gesundheitszustand</i>	0,949	0,355	0,373	0,742
Veränderung der SF-36 v2 akut Domäne <i>Vitalität</i>	0,353	0,457	0,774	0,381
Veränderung der SF-36 v2 akut Domäne <i>Soziale Funktionsfähigkeit</i>	0,293	0,131	0,501	0,108
Veränderung der SF-36 v2 akut Domäne <i>Emotionale Rollenfunktion</i>	0,652	0,155	0,433	0,156
Veränderung der SF-36 v2 akut Domäne <i>Mentaler Gesundheitszustand</i>	0,938	0,136	0,689	0,050

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a			
Merkmal	Krankheitsaktivität zu Baseline (DAS28-hsCRP ≤ 5,1; DAS28-hsCRP > 5,1)	Gelenkerosionsstatus (1-2 Gelenkerosionen + Seropositivität; mindestens 3 Gelenkerosionen)	Seropositivität und Gelenkerosionsstatus (Seropositivität und mind. 3 Gelenkerosionen; Andere)	Seropositivität (RF oder ACPA positiv; RF und ACPA negativ)
Sicherheit				
SUE	NE	NE	0,0292 [†]	NE
UE	0,8442	0,0068 [†]	0,0061 [†]	0,8839
Schwere UE (ohne ausschließliche Laborereignisse)	NE	NE	NE	NE
Studienabbruch aufgrund von UE	0,9384	NE	NE	NE
Therapieabbruch aufgrund von UE	0,8020	NE	0,3917	NE
UE von besonderem Interesse				
Infektionen	0,3947	0,3682	0,3804	0,7929
Depressionen: QIDS-SR16	0,043 [†]	0,688	0,684	0,826
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechrare und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>†: p < 0,05</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>				

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

Tabelle 4-107: Übersicht über alle durchgeführten Interaktionstests in der mITT Population von Baseline bis Woche 52 (4)

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest ^a				
Merkmal	Seropositivität (RF positiv; RF negativ)	Seropositivität (ACPA positiv; ACPA negativ)	Art der Begleitmedikation (nur MTX; MTX + andere DMARDs)	mTSS zu Baseline (≤ Median; > Median)	Therapie mit Kortikosteroiden zu Baseline (Ja; Nein)
Nutzendimension Endpunkt					
Mortalität^b					
Morbidität					
Veränderung der Anzahl der druckschmerzhaften Gelenke	0,586	0,488	0,480	0,270	0,670
Veränderung der Anzahl der geschwollenen Gelenke	0,887	0,362	0,310	0,365	0,944
Veränderung des DAS28-hsCRP	0,859	0,086	0,126	0,495	0,875
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	0,7325	0,5697	0,2173	0,6564	0,2566
Anteil Patienten mit DAS28-hsCRP < 2,6	0,9696	0,4487	0,4128	0,4636	0,7219
Veränderung des SDAI	0,778	0,335	0,211	0,206	0,729
Anteil Patienten mit SDAI ≤ 3,3	0,6962	0,7942	0,9809	0,8419	0,3457
Veränderung des CDAI	0,644	0,455	0,294	0,232	0,698
Anteil Patienten mit CDAI ≤ 10	0,4962	0,6900	0,4104	0,8942	0,6662
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	0,8861	0,5446	0,1545	0,8158	0,2131

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a				
Merkmal	Seropositivität (RF positiv; RF negativ)	Seropositivität (ACPA positiv; ACPA negativ)	Art der Begleitmedikation (nur MTX; MTX + andere DMARDs)	mTSS zu Baseline (≤ Median; > Median)	Therapie mit Kortikosteroiden zu Baseline (Ja; Nein)
Veränderung des PAP (VAS)	0,858	0,114	0,163	0,203	0,780
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm	0,8825	0,5182	0,3884	0,2342	0,7974
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm	0,5919	0,2947	0,1962	0,0983	0,9729
Veränderung des PtGADA (VAS)	0,721	0,295	0,357	0,148	0,545
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm	0,7663	0,7563	0,7035	0,0425 [†]	0,4344
Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm	0,7919	0,8484	0,3968	0,0244 [†]	0,6842
Veränderung des HAQ-DI	0,422	0,791	0,250	0,449	0,627
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,22	0,6109	0,4464	0,1068	0,3954	0,6743
Anteil Patienten mit einer Verbesserung des HAQ-DI ≥ 0,3	0,7487	0,9478	0,3593	0,6928	0,9499
Veränderung der Dauer der Morgensteifigkeit (min)	0,839	0,963	0,276	0,527	0,171
Veränderung der EQ-5D (VAS)	0,477	0,125	0,270	0,986	0,048 [†]
Anteil Patienten mit einer Verbesserung der EQ-5D (VAS) ≥ 10 mm	0,2071	0,0449 [†]	0,5153	0,4038	0,3232
Veränderung der schwersten Müdigkeit	0,304	0,952	0,188	0,667	0,029 [†]

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a				
Merkmal	Seropositivität (RF positiv; RF negativ)	Seropositivität (ACPA positiv; ACPA negativ)	Art der Begleitmedikation (nur MTX; MTX + andere DMARDs)	mTSS zu Baseline (≤ Median; > Median)	Therapie mit Kortikosteroiden zu Baseline (Ja; Nein)
Lebensqualität					
Veränderung des SF-36 v2 akut PCS	0,689	0,629	0,330	0,326	0,612
Anteil Patienten mit SF-36 v2 akut PCS ≥ 5	0,7407	0,5682	0,2962	0,4320	0,3579
Veränderung des SF-36 v2 akut MCS	0,052	0,071	0,477	0,753	0,029 [†]
Anteil Patienten mit SF-36 v2 akut MCS ≥ 5	0,3234	0,6102	0,0370 [†]	0,0439 [†]	0,5788
Veränderung der SF-36 v2 akut Domäne Körperliche Funktionsfähigkeit	0,939	0,980	0,446	0,031 [†]	0,212
Veränderung der SF-36 v2 akut Domäne Körperliche Rollenfunktion	0,556	0,345	0,138	0,170	0,402
Veränderung der SF-36 v2 akut Domäne Körperliche Schmerzen	0,368	0,748	0,603	0,743	0,263
Veränderung der SF-36 v2 akut Domäne Allgemeiner Gesundheitszustand	0,318	0,659	0,824	0,625	0,591
Veränderung der SF-36 v2 akut Domäne Vitalität	0,370	0,431	0,350	0,702	0,070
Veränderung der SF-36 v2 akut Domäne Soziale Funktionsfähigkeit	0,262	0,362	0,569	0,226	0,071
Veränderung der SF-36 v2 akut Domäne Emotionale Rollenfunktion	0,104	0,071	0,873	0,252	0,023 [†]
Veränderung der SF-36 v2 akut Domäne Mentaler Gesundheitszustand	0,096	0,130	0,187	0,505	0,105

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

Studie RA-BEAM (JADV)	p-Wert aus Interaktionstest^a				
Merkmal	Seropositivität (RF positiv; RF negativ)	Seropositivität (ACPA positiv; ACPA negativ)	Art der Begleitmedikation (nur MTX; MTX + andere DMARDs)	mTSS zu Baseline (≤ Median; > Median)	Therapie mit Kortikosteroiden zu Baseline (Ja; Nein)
Sicherheit					
SUE	NE	NE	NE	0,1054	0,9690
UE	0,1853	0,1626	0,2423	0,2474	0,4336
Schwere UE (ohne ausschließliche Laborereignisse)	NE	NE	NE	0,0148 [†]	0,6806
Studienabbruch aufgrund von UE	NE	NE	NE	0,1160	0,3570
Therapieabbruch aufgrund von UE	NE	NE	NE	0,0449 [†]	0,4585
UE von besonderem Interesse					
Infektionen	0,9713	0,2464	0,6365	0,7544	0,8870
Depressionen: QIDS-SR16 (mLOCF)	0,352	0,563	0,198	0,244	0,007 [†]
<p>a: p-Wert für Interaktion aus einem Cox-Modell (bzw. logistischen Modell für objektive Ansprechraten und Krankheitskontrollrate) mit den Einflussfaktoren Behandlung, Subgruppe und Behandlung*Subgruppe.</p> <p>b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegendsten unerwünschten Ereignisse ausgewertet.</p> <p>†: p < 0,05</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [11].</p>					

4.3.1.3.5.1 Druckschmerzhaftes Gelenke und geschwollene Gelenke – RCT**Veränderung der Anzahl der geschwollenen Gelenke nach BMI zu Baseline**

Tabelle 4-108: Ergebnisse für Veränderung der Anzahl der geschwollenen Gelenke aus RCT mit dem zu bewertenden Arzneimittel nach BMI zu Baseline

Baricitinib vs. Adalimumab				
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der Anzahl der geschwollenen Gelenke nach BMI zu Baseline				
Subpopulation A3				
Interaktionstest^a: p = 0,013				
N	170	124		
< 25 kg/m²				
N'	91	53		
Baseline, MW (SD)	10,92 (4,80)	9,89 (4,48)	-1,04 (-2,30; 0,21); 0,103	-0,28 (-0,62; 0,06)
Woche 52, MW (SD)	2,53 (3,33)	3,29 (4,64)		
LSM (SE)	-8,12 (0,384)	-7,08 (0,503)		
≥ 25 kg/m² und < 30 kg/m²				
N'	54	43		
Baseline, MW (SD)	11,48 (5,21)	11,09 (6,04)	1,48 (-0,64; 3,59); 0,169	0,28 (-0,12; 0,68)
Woche 52, MW (SD)	4,17 (5,99)	2,59 (4,52)		
LSM (SE)	-7,19 (0,710)	-8,66 (0,795)		
≥ 30 kg/m²				
N'	25	27		
Baseline, MW (SD)	10,22 (3,95)	10,74 (4,14)	-2,59 (-4,60; -0,57); 0,013	-0,71 (-1,26; -0,15)
Woche 52, MW (SD)	1,59 (2,64)	4,29 (4,41)		
LSM (SE)	-8,84 (0,722)	-6,25 (0,695)		
mITT				
Interaktionstest^a: p = 0,019				
N	487	330		
< 25 kg/m²				
N'	210	148		
Baseline, MW (SD)	11,05 (5,00)	11,02 (5,55)	-1,43 (-2,31; -0,56); 0,001	-0,35 (-0,56; -0,14)
Woche 52, MW (SD)	2,41 (3,74)	3,84 (5,31)		
LSM (SE)	-8,63 (0,285)	-7,20 (0,339)		
≥ 25 kg/m² und < 30 kg/m²				
N'	159	107		
Baseline, MW (SD)	11,23 (5,15)	11,24 (5,77)	0,44 (-0,76; 1,65); 0,471	0,09 (-0,15; 0,33)
Woche 52, MW (SD)	3,70 (5,21)	3,26 (5,11)		
LSM (SE)	-7,53 (0,388)	-7,98 (0,473)		

Baricitinib vs. Adalimumab				
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der Anzahl der geschwollenen Gelenke nach BMI zu Baseline				
≥ 30 kg/m²				
N'	113	73	-1,60 (-3,00; -0,20); 0,025	-0,34 (-0,63; -0,04)
Baseline, MW (SD)	10,91 (4,99)	11,15 (5,39)		
Woche 52, MW (SD)	2,70 (4,76)	4,39 (5,50)		
LSM (SE)	-8,27 (0,444)	-6,67 (0,553)		
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.				

Subpopulation A3

Für den Endpunkt Veränderung der Anzahl der geschwollenen Gelenke waren die Effektschätzer nach BMI zu Baseline nicht gleichgerichtet. In den Gruppen der Patienten mit einem BMI von $< 25 \text{ kg/m}^2$ und von $\geq 25 \text{ kg/m}^2$ und $< 30 \text{ kg/m}^2$ war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. In der Gruppe der Patienten mit einem BMI von $\geq 30 \text{ kg/m}^2$ war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffektes kann nicht ausgeschlossen werden.

mITT

Für den Endpunkt Veränderung der Anzahl der geschwollenen Gelenke waren die Effektschätzer nach BMI zu Baseline nicht gleichgerichtet. In der Gruppe der Patienten mit einem BMI von ≥ 25 und $< 30 \text{ kg/m}^2$ war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. In der Gruppe der Patienten mit einem BMI von $< 25 \text{ kg/m}^2$ sowie $\geq 30 \text{ kg/m}^2$ war Baricitinib jeweils statistisch signifikant überlegen zu Adalimumab. Eine Modifikation des Behandlungseffektes kann nicht ausgeschlossen werden.

4.3.1.3.5.2 Disease Activity Score 28 High Sensitivity C-Reactive Protein (DAS28-hsCRP) – RCT

Anteil Patienten mit DAS28-hsCRP < 2,6 nach Seropositivität und Gelenkerosionsstatus

Tabelle 4-109: Ergebnisse für Anteil Patienten mit DAS28-hsCRP < 2,6 aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit DAS28-hsCRP < 2,6 nach Seropositivität und Gelenkerosionsstatus					
Subpopulation A3			Interaktionstest^a: p = 0,0471		
N	170	124			
Seropositiv und mind. 3 Gelenkerosionen					
N´	124	92	1,36 (0,78; 2,37); 0,276	1,20 (0,86; 1,68); 0,282	7,40 (-5,78; 20,58); 0,328
Anzahl Patienten mit Ereignis n (%)	55 (44,4)	34 (37,0)			
Andere					
N´	46	32	0,45 (0,17; 1,17); 0,099	0,60 (0,33; 1,10); 0,097	-18,61 (-40,25; 3,02); 0,101
Anzahl Patienten mit Ereignis n (%)	13 (28,3)	15 (46,9)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A3

Die Hauptanalyse des Anteils der Patienten mit DAS28-hsCRP < 2,6 an Woche 52 konnte keinen statistisch signifikanten Behandlungseffekt für Baricitinib gegenüber Adalimumab nachweisen. Die Subgruppenanalyse nach Seropositivität und Gelenkerosionsstatus ist dazu konsistent, d. h. in keiner der beiden Subgruppen gab es einen statistisch signifikanten Unterschied gegenüber Adalimumab. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

4.3.1.3.5.3 Simplified Disease Activity Index (SDAI) – RCT**Anteil Patienten mit SDAI \leq 3,3 nach spezifischer Region I**

Tabelle 4-110: Ergebnisse für Anteil Patienten mit SDAI \leq 3,3 aus RCT mit dem zu bewertenden Arzneimittel nach spezifischer Region I

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit SDAI \leq 3,3 nach spezifischer Region I					
Subpopulation A2			Interaktionstest^a: p = 0,0241		
N	243	153			
Europa					
N ^c	62	33	0,40 (0,13; 1,16); 0,091	0,47 (0,20; 1,12); 0,089	-14,37 (-31,71; 2,97); 0,097
Anzahl Patienten mit Ereignis n (%)	8 (12,6)	9 (27,3)			
Andere					
N ^c	181	120	1,61 (0,90; 2,87); 0,109	1,45 (0,91; 2,31); 0,114	7,91 (-1,38; 17,21); 0,120
Anzahl Patienten mit Ereignis n (%)	46 (25,4)	21 (17,5)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert.					
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Die Hauptanalyse des Anteils der Patienten mit SDAI \leq 3,3 an Woche 52 konnte keinen statistisch signifikanten Behandlungseffekt für Baricitinib gegenüber Adalimumab nachweisen. Die Subgruppenanalyse nach spezifischer Region I ist dazu konsistent, d. h. in keiner der beiden Subgruppen gab es einen statistisch signifikanten Unterschied gegenüber Adalimumab. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

4.3.1.3.5.4 Clinical Disease Activity Index (CDAI) – RCT**Anteil Patienten mit CDAI ≤ 10 nach Ethnie**

Tabelle 4-111: Ergebnisse für Anteil Patienten mit CDAI ≤ 10 aus RCT mit dem zu bewertenden Arzneimittel nach Ethnie

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit CDAI ≤ 10 nach Ethnie					
Subpopulation A3			Interaktionstest^a: p = 0,0287		
N	170	124			
Kaukasisch					
N´	101	78	0,75	0,88	-7,08
Anzahl Patienten mit Ereignis n (%)	55 (54,5)	48 (61,5)	(0,41; 1,37); 0,344	(0,69; 1,14); 0,339	(-21,61; 7,44); 0,364
Andere					
N´	69	46	2,29	1,47	20,29
Anzahl Patienten mit Ereignis n (%)	44 (63,8)	20 (43,5)	(1,06; 4,94); 0,035	(1,00; 2,14); 0,047	(2,02; 38,56); 0,037
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A3

Für den Endpunkt Anteil Patienten mit CDAI ≤ 10 waren die Effektschätzer nach Ethnie nicht gleichgerichtet. Für kaukasische Patienten war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten anderer Ethnie war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

4.3.1.3.5.5 EULAR Responsekriterien („good Response“, DAS28-basiert) – RCT**Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert) nach Ethnie**

Tabelle 4-112: Ergebnisse für Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert) aus RCT mit dem zu bewertenden Arzneimittel nach Ethnie

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert) nach Ethnie					
Subpopulation A3			Interaktionstest^a: p = 0,0366		
N	170	124			
Kaukasisch					
N´	101	78	0,79	0,90	-5,92
Anzahl Patienten mit Ereignis n (%)	51 (50,5)	44 (56,4)	(0,43; 1,43); 0,433	(0,68; 1,18); 0,430	(-20,62; 8,79); 0,454
Andere					
N´	69	46	2,28	1,52	20,29
Anzahl Patienten mit Ereignis n (%)	41 (59,4)	18 (39,1)	(1,05; 4,92); 0,037	(1,00; 2,30); 0,048	(2,04; 38,54); 0,038
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A3

Für den Endpunkt Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien waren die Effektschätzer nach Ethnie nicht gleichgerichtet. Für kaukasische Patienten war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten anderer Ethnie war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

4.3.1.3.5.6 Patient Assessment of Pain (PAP) (VAS) – RCT**Veränderung des PAP (VAS) nach Geschlecht**

Tabelle 4-113: Ergebnisse für Veränderung des PAP (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach Geschlecht

Baricitinib vs. Adalimumab			
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert
Veränderung des PAP (VAS) nach Geschlecht			
Subpopulation A2		Interaktionstest^a: p = 0,004	
N	243	153	
Männlich			
N´	62	48	-15,75 (-24,26; -7,25); < 0,001
Baseline, MW (SD)	64,60 (24,64)	61,06 (24,03)	
Woche 52, MW (SD)	21,27 (21,51)	35,88 (25,92)	
LSM (SE)	-42,28 (2,830)	-26,53 (3,217)	
Weiblich			
N´	176	104	0,30 (-5,72; 6,31); 0,923
Baseline, MW (SD)	63,90 (21,79)	63,74 (22,89)	
Woche 52, MW (SD)	30,09 (25,42)	29,77 (25,67)	
LSM (SE)	-33,90 (1,861)	-34,20 (2,421)	
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.			

Subpopulation A2

Für den Endpunkt Veränderung des PAP (VAS) waren die Effektschätzer nach Geschlecht nicht gleichgerichtet. In der Gruppe der männlichen Patienten war Baricitinib statistisch signifikant überlegen zu Adalimumab. In der Gruppe der weiblichen Patienten war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Veränderung des PAP (VAS) nach spezifischer Region II

Tabelle 4-114: Ergebnisse für Veränderung des PAP (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach spezifischer Region II

Baricitinib vs. Adalimumab			
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert
Veränderung des PAP (VAS) nach spezifischer Region II			
mITT		Interaktionstest^a: p = 0,047	
N	487	330	
Westeuropa			
N'	27	20	-19,49 (-30,58; -8,41); < 0,001
Baseline, MW (SD)	52,89 (18,08)	55,40 (29,04)	
Woche 52, MW (SD)	15,67 (15,05)	35,60 (23,07)	
LSM (SE)	-38,10 (3,585)	-18,61 (4,166)	
Andere			
N'	455	307	-5,05 (-8,46; -1,63); 0,004
Baseline, MW (SD)	62,39 (21,91)	61,38 (22,25)	
Woche 52, MW (SD)	25,67 (23,72)	30,38 (25,81)	
LSM (SE)	-36,51 (1,104)	-31,46 (1,344)	
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.			

mITT

Für den Endpunkt Veränderung des PAP (VAS) waren die Effektschätzer nach spezifischer Region II gleichgerichtet und statistisch signifikant zugunsten von Baricitinib. Dies entspricht auch dem Ergebnis der Hauptanalyse. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Veränderung des PAP (VAS) nach Seropositivität (ACPA)

Tabelle 4-115: Ergebnisse für Veränderung des PAP (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität (ACPA)

Baricitinib vs. Adalimumab			
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert
Veränderung des PAP (VAS) nach Seropositivität (ACPA)			
Subpopulation A3			Interaktionstest^a: p = 0,047
N	170	124	
ACPA positiv			
N'	153	111	-7,62 (-12,92; -2,32); 0,005
Baseline, MW (SD)	58,44 (21,85)	58,70 (22,22)	
Woche 52, MW (SD)	20,46 (19,96)	28,14 (24,81)	
LSM (SE)	-38,06 (1,745)	-30,44 (2,048)	
ACPA negativ			
N'	17	12	9,79 (-8,95; 28,52); 0,293
Baseline, MW (SD)	59,24 (17,70)	61,33 (20,86)	
Woche 52, MW (SD)	25,06 (27,05)	15,92 (19,92)	
LSM (SE)	-34,78 (5,859)	-44,56 (6,975)	
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.			

Subpopulation A3

Für den Endpunkt Veränderung des PAP (VAS) waren die Effektschätzer nach Seropositivität (ACPA) nicht gleichgerichtet. In der Gruppe der ACPA positiven Patienten war Baricitinib signifikant überlegen zu Adalimumab. In der Gruppe der ACPA negativen Patienten war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 10 mm nach Geschlecht

Tabelle 4-116: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Geschlecht

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 10 mm nach Geschlecht					
Subpopulation A2			Interaktionstest^a: p = 0,0395		
N	243	153			
Männlich					
N´	62	48	4,67 (1,64; 13,28); 0,004	1,35 (1,09; 1,69); 0,007	23,66 (8,42; 38,89); 0,003
Anzahl Patienten mit Ereignis n (%)	56 (90,3)	32 (66,7)			
Weiblich					
N´	177	105	1,18 (0,66; 2,11); 0,570	1,04 (0,91; 1,18); 0,576	2,91 (-7,21; 13,02); 0,656
Anzahl Patienten mit Ereignis n (%)	140 (79,1)	80 (76,2)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Anteil Patienten mit einer Veränderung des PAP (VAS) \geq 10 waren die Effektschätzer nach Geschlecht gleichgerichtet. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm nach Gelenkerosionsstatus

Tabelle 4-117: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm nach Gelenkerosionsstatus					
mITT Interaktionstest^a: p = 0,0029					
N	487	330			
1-2 Gelenkerosionen + Seropositivität					
N'	115	80	0,70 (0,33; 1,47); 0,344	0,93 (0,81; 1,07); 0,331	-5,49 (-16,54; 5,56); 0,365
Anzahl Patienten mit Ereignis n (%)	90 (78,3)	67 (83,8)			
Mindestens 3 Gelenkerosionen					
N'	367	245	2,31 (1,55; 3,44); < 0,001	1,20 (1,09; 1,32); < 0,001	14,13 (7,34; 20,92); < 0,001
Anzahl Patienten mit Ereignis n (%)	311 (84,7)	173 (70,6)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 10 waren die Effektschätzer nach Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten mit 1-2 Gelenkerosionen + Seropositivität war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten mit mindestens 3 Gelenkerosionen war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm nach Seropositivität und Gelenkerosionsstatus

Tabelle 4-118: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm nach Seropositivität und Gelenkerosionsstatus					
Subpopulation A3			Interaktionstest^a: p = 0,0124		
N	170	124			
Seropositiv und mind. 3 Gelenkerosionen					
N'	124	92			
Anzahl Patienten mit Ereignis n (%)	108 (87,1)	66 (71,7)	2,66 (1,32; 5,34); 0,006	1,21 (1,05; 1,40); 0,009	15,36 (4,43; 26,29); 0,006
Andere					
N'	46	31			
Anzahl Patienten mit Ereignis n (%)	35 (76,1)	27 (87,1)	0,47 (0,13; 1,68); 0,242	0,87 (0,70; 1,08); 0,214	-11,01 (-28,07; 6,05); 0,260
mITT			Interaktionstest^a: p = 0,0173		
N	487	330			
Seropositiv und mind. 3 Gelenkerosionen					
N'	340	233			
Anzahl Patienten mit Ereignis n (%)	291 (85,6)	166 (71,2)	2,40 (1,58; 3,63); < 0,001	1,20 (1,09; 1,32); < 0,001	14,34 (7,44; 21,25); < 0,001
Andere					
N'	143	95			
Anzahl Patienten mit Ereignis n (%)	111 (77,6)	75 (78,9)	0,92 (0,49; 1,74); 0,809	0,98 (0,86; 1,13); 0,808	-1,32 (-12,00; 9,35); 0,873
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A3

Für den Endpunkt Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 10 waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten, die seropositiv waren und mind. 3 Gelenkerosionen hatten, war Baricitinib statistisch signifikant überlegen zu Adalimumab. Für Patienten mit einem anderen Status war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

mITT

Für den Endpunkt Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 10 waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten, die seropositiv waren und mind. 3 Gelenkerosionen hatten, war Baricitinib statistisch signifikant überlegen zu Adalimumab. Für Patienten mit einem anderen Status war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm nach Geschlecht

Tabelle 4-119: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Geschlecht

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm nach Geschlecht					
Subpopulation A2			Interaktionstest^a: p = 0,0254		
N	243	153			
Männlich					
N´	62	48	3,19 (1,37; 7,42); 0,008	1,46; (1,09; 1,96); 0,012	24,87 (7,51; 42,23); 0,007
Anzahl Patienten mit Ereignis n (%)	49 (79,0)	26 (54,2)			
Weiblich					
N´	177	105	0,98 (0,59; 1,65); 0,947	0,99 (0,84; 1,18); 0,946	-0,39 (-11,70; 10,92); 1,000
Anzahl Patienten mit Ereignis n (%)	119 (67,2)	71 (67,6)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 20 mm waren die Effektschätzer nach Geschlecht nicht gleichgerichtet. Für die Gruppe der weiblichen Patienten waren die Behandlungseffekte ausgeglichen und statistisch nicht signifikant verschieden zu Adalimumab. Für die Gruppe der männlichen Patienten war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 20 mm nach Alter

Tabelle 4-120: Ergebnisse für Anteil Patienten mit einer Veränderung des PAP (VAS) \geq 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Alter

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 20 mm nach Alter					
Subpopulation A2			Interaktionstest^a: p = 0,0056		
N	243	153			
< 65 Jahre					
N´	184	127	0,95 (0,58; 1,55); 0,837	0,98 (0,84; 1,15); 0,836	-1,11 (-11,66; 9,43); 0,902
Anzahl Patienten mit Ereignis n (%)	124 (67,4)	87 (68,5)			
\geq 65 Jahre					
N´	55	26	6,40 (2,25; 18,22); < 0,001	2,08 (1,25; 3,47); 0,006	41,54 (20,06; 63,02); < 0,001
Anzahl Patienten mit Ereignis n (%)	44 (80,0)	10 (38,5)			
mITT			Interaktionstest^a: p = 0,0404		
N	487	330			
< 65 Jahre					
N´	381	273	1,39 (1,00; 1,95); 0,052	1,11 (1,00; 1,24); 0,056	7,18 (-0,08; 14,44); 0,060
Anzahl Patienten mit Ereignis n (%)	273 (71,7)	176 (64,5)			
\geq 65 Jahre					
N´	102	55	3,46 (1,68; 7,11); < 0,001	1,51 (1,15; 1,98); 0,003	26,68 (11,33; 42,04); < 0,001
Anzahl Patienten mit Ereignis n (%)	81 (79,4)	29 (52,7)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Anteil Patienten mit einer Veränderung des PAP (VAS) \geq 20 mm waren die Effektschätzer nach Alter nicht gleichgerichtet. Für die Gruppe der Patienten < 65 Jahre waren die Behandlungseffekte ausgeglichen und statistisch nicht signifikant verschieden zu Adalimumab. Für die Gruppe der Patienten \geq 65 Jahre war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

mITT

Für den Endpunkt Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 20 waren die Effektschätzer nach Alter gleichgerichtet. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm nach Gelenkerosionsstatus

Tabelle 4-121: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm nach Gelenkerosionsstatus					
Subpopulation A2			Interaktionstest^a: p = 0,0159		
N	243	153			
1-2 Gelenkerosionen + Seropositivität					
N´	59	40	0,52	0,81	-13,98
Anzahl Patienten mit Ereignis n (%)	36 (61,0)	30 (75,0)	(0,21; 1,28); 0,154	(0,62; 1,07); 0,139	(-32,28; 4,32); 0,193
Mindestens 3 Gelenkerosionen					
N´	179	111	1,79	1,21	12,82
Anzahl Patienten mit Ereignis n (%)	131 (73,2)	67 (60,4)	(1,08; 2,97); 0,024	(1,02; 1,45); 0,032	(1,65; 24,00); 0,027
mITT			Interaktionstest^a: p = 0,0116		
N	487	330			
1-2 Gelenkerosionen + Seropositivität					
N´	115	80	0,82	0,94	-4,29
Anzahl Patienten mit Ereignis n (%)	77 (67,0)	57 (71,3)	(0,44; 1,53); 0,526	(0,78; 1,14); 0,521	(-17,42; 8,83); 0,638
Mindestens 3 Gelenkerosionen					
N´	367	245	2,02	1,25	15,20
Anzahl Patienten mit Ereignis n (%)	276 (75,2)	147 (60,0)	(1,43; 2,87); < 0,001	(1,11; 1,41); < 0,001	(7,64; 22,76); < 0,001
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert.					
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 20 mm waren die Effektschätzer nach Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten mit 1-2 Gelenkerosionen + Seropositivität war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten mit mindestens 3 Gelenkerosionen war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

mITT

Für den Endpunkt Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 20 mm waren die Effektschätzer nach Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten mit 1-2 Gelenkerosionen + Seropositivität war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten mit mindestens 3 Gelenkerosionen war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 20 mm nach Seropositivität und Gelenkerosionsstatus

Tabelle 4-122: Ergebnisse für Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 20 mm nach Seropositivität und Gelenkerosionsstatus					
Subpopulation A2			Interaktionstest^a: p = 0,0481		
N	243	153			
Seropositiv und mind. 3 Gelenkerosionen					
N'	164	103	1,86	1,23	13,59
Anzahl Patienten mit Ereignis n (%)	121 (73,8)	62 (60,2)	(1,10; 3,16); 0,021	(1,02; 1,47); 0,029	(1,98; 25,19); 0,022
Andere					
N'	75	50	0,72	0,90	-7,33
Anzahl Patienten mit Ereignis n (%)	47 (62,7)	35 (70,0)	(0,33; 1,56); 0,400	(0,69; 1,15); 0,391	(-24,10; 9,43); 0,446
mITT			Interaktionstest^a: p = 0,0207		
N	487	330			
Seropositiv und mind. 3 Gelenkerosionen					
N'	340	233	2,12	1,26	15,96
Anzahl Patienten mit Ereignis n (%)	260 (76,5)	141 (60,5)	(1,47; 3,05); < 0,001	(1,12; 1,42); < 0,001	(8,23; 23,68); < 0,001
Andere					
N'	143	95	0,93	0,98	-1,63
Anzahl Patienten mit Ereignis n (%)	94 (65,7)	64 (67,4)	(0,53; 1,62); 0,794	(0,81; 1,17), 0,793	(-13,86; 10,59); 0,889
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert.					
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Anteil Patienten mit einer Veränderung des PAP (VAS) \geq 20 mm waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten, die seropositiv waren und mind. 3 Gelenkerosionen hatten, war Baricitinib statistisch signifikant überlegen zu Adalimumab. Für Patienten mit einem anderen Status war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

mITT

Für den Endpunkt Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 20 mm waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten, die seropositiv waren und mind. 3 Gelenkerosionen hatten, war Baricitinib statistisch signifikant überlegen zu Adalimumab. Für Patienten mit einem anderen Status war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

4.3.1.3.5.7 Patient's Global Assessment of Disease Activity (PtGADA) (VAS) – RCT**Veränderung des PtGADA (VAS) nach Geschlecht**

Tabelle 4-123: Ergebnisse für Veränderung des PtGADA (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach Geschlecht

Baricitinib vs. Adalimumab			
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert
Veränderung des PtGADA (VAS) nach Geschlecht			
Subpopulation A2			Interaktionstest^a: p = 0,021
N	243	153	
Männlich			
N'	62	48	-14,69 (-23,39; -5,99); 0,001
Baseline, MW (SD)	65,42 (22,82)	67,73 (18,28)	
Woche 52, MW (SD)	22,79 (20,92)	38,13 (26,32)	
LSM (SE)	-43,36 (2,896)	-28,67 (3,292)	
Weiblich			
N'	176	104	-2,30 (-8,07; 3,46); 0,432
Baseline, MW (SD)	64,76 (21,63)	62,74 (23,00)	
Woche 52, MW (SD)	30,33 (23,91)	31,97 (25,42)	
LSM (SE)	-34,02 (1,783)	-31,72 (2,320)	
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.			

Subpopulation A2

Für den Endpunkt Veränderung des PtGADA waren die Effektschätzer nach Geschlecht gleichgerichtet. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Veränderung des PtGADA (VAS) nach spezifischer Region I

Tabelle 4-124: Ergebnisse für Veränderung des PtGADA (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach spezifischer Region I

Baricitinib vs. Adalimumab			
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert
Veränderung des PtGADA (VAS) nach spezifischer Region I			
Subpopulation A2			Interaktionstest^a: p = 0,039
N	243	153	
Europa			
N'	60	32	3,59 (-5,55; 12,72); 0,438
Baseline, MW (SD)	62,37 (19,08)	61,53 (21,23)	
Woche 52, MW (SD)	31,48 (21,03)	28,06 (21,58)	
LSM (SE)	-30,65 (2,712)	-34,24 (3,714)	
Andere			
N'	178	120	-8,55 (-14,16; -2,94); 0,003
Baseline, MW (SD)	65,79 (22,76)	65,06 (21,83)	
Woche 52, MW (SD)	27,33 (24,06)	35,51 (26,67)	
LSM (SE)	-38,39 (1,809)	-29,85 (2,204)	
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.			

Subpopulation A2

Für den Endpunkt Veränderung des PtGADA waren die Effektschätzer nach spezifischer Region I nicht gleichgerichtet. Für Patienten aus Europa war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten aus anderen Regionen war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Veränderung des PtGADA (VAS) nach Gelenkerosionsstatus

Tabelle 4-125: Ergebnisse für Veränderung des PtGADA (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus

Baricitinib vs. Adalimumab			
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert
Veränderung des PtGADA (VAS) nach Gelenkerosionsstatus			
Subpopulation A2			Interaktionstest^a: p = 0,021
N	243	153	
1-2 Gelenkerosionen + Seropositivität			
N´	59	40	3,34 (-6,17; 12,86); 0,487
Baseline, MW (SD)	62,05 (23,44)	66,33 (21,36)	
Woche 52, MW (SD)	29,59 (23,17)	26,95 (23,91)	
LSM (SE)	-33,90 (3,039)	-37,24 (3,694)	
Mindestens 3 Gelenkerosionen			
N´	178	110	-9,04 (-14,62; -3,46); 0,002
Baseline, MW (SD)	65,79 (21,38)	64,26 (21,41)	
Woche 52, MW (SD)	28,12 (23,46)	36,50 (26,14)	
LSM (SE)	-37,43 (1,750)	-28,39 (2,227)	
mITT			Interaktionstest^a: p = 0,043
N	487	330	
1-2 Gelenkerosionen + Seropositivität			
N´	115	80	-0,24 (-6,54; 6,07); 0,941
Baseline, MW (SD)	62,46 (24,05)	64,08 (21,84)	
Woche 52, MW (SD)	27,19 (22,49)	27,71 (21,91)	
LSM (SE)	-35,82 (2,046)	-35,58 (2,454)	
Mindestens 3 Gelenkerosionen			
N´	366	244	-7,82 (-11,61; -4,03); < 0,001
Baseline, MW (SD)	63,29 (20,21)	64,06 (20,83)	
Woche 52, MW (SD)	26,27 (22,67)	34,32 (26,96)	
LSM (SE)	-37,29 (1,220)	-29,47 (1,495)	
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.			

Subpopulation A2

Für den Endpunkt Veränderung des PtGADA waren die Effektschätzer nach Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten mit 1-2 Gelenkerosionen + Seropositivität war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten mit mindestens 3 Gelenkerosionen war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

mITT

Für den Endpunkt Veränderung des PtGADA waren die Effektschätzer nach Gelenkerosionsstatus gleichgerichtet. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Veränderung des PtGADA (VAS) nach Seropositivität und Gelenkerosionsstatus

Tabelle 4-126: Ergebnisse für Veränderung des PtGADA (VAS) aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus

Baricitinib vs. Adalimumab			
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert
mITT			Interaktionstest^a: p = 0,036
N	487	330	
Seropositiv und mindestens 3 Gelenkerosionen			
N´	339	232	-8,22 (-12,13; -4,31); < 0,001
Baseline, MW (SD)	63,88 (20,33)	63,85 (20,83)	
Woche 52, MW (SD)	25,93 (22,49)	34,09 (27,00)	
LSM (SE)	-38,01(1,269)	-29,79 (1,534)	
Andere			
N´	143	95	-0,77 (-6,59; 5,05); 0,794
Baseline, MW (SD)	61,35 (22,99)	63,49 (22,30)	
Woche 52, MW (SD)	27,64 (22,93)	28,89 (22,73)	
LSM (SE)	-34,38 (1,866)	-33,60 (2,290)	
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.			

mITT

Für den Endpunkt Veränderung des PtGADA waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus gleichgerichtet. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 10 mm nach Gelenkerosionsstatus

Tabelle 4-127: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 10 mm nach Gelenkerosionsstatus					
mITT Interaktionstest^a: p = 0,0046					
N	487	330			
1-2 Gelenkerosionen + Seropositivität					
N´	115	80	0,55	0,90	-8,86
Anzahl Patienten mit Ereignis n (%)	89 (77,4)	69 (86,3)	(0,25; 1,19); 0,126	(0,79; 1,02); 0,109	(-19,60; 1,88); 0,140
Mindestens 3 Gelenkerosionen					
N´	367	245	1,77	1,13	9,50
Anzahl Patienten mit Ereignis n (%)	306 (83,4)	181 (73,9)	(1,19; 2,64); 0,005	(1,03; 1,23); 0,007	(2,81; 16,19); 0,006
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt Anteil Patienten mit einer Verbesserung der PtGADA (VAS) \geq 10 mm waren die Effektschätzer nach Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten mit 1-2 Gelenkerosionen + Seropositivität war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten mit mindestens 3 Gelenkerosionen war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 10 mm nach Seropositivität und Gelenkerosionsstatus

Tabelle 4-128: Ergebnisse für Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 10 mm nach Seropositivität und Gelenkerosionsstatus					
Subpopulation A3			Interaktionstest^a: p = 0,0250		
N	170	124			
Seropositiv und mind. 3 Gelenkerosionen					
N'	124	92			
Anzahl Patienten mit Ereignis n (%)	110 (88,7)	72 (78,3)	2,18 (1,03; 4,62); 0,041	1,13 (1,00; 1,28); 0,050	10,45 (0,35; 20,55); 0,058
Andere					
N'	46	31			
Anzahl Patienten mit Ereignis n (%)	34 (73,9)	27 (87,1)	0,42 (0,12; 1,48); 0,174	0,85 (0,68; 1,06); 0,146	-13,18 (-30,51; 4,14); 0,252
mITT			Interaktionstest^a: p = 0,0133		
N	487	330			
Seropositiv und mind. 3 Gelenkerosionen					
N'	340	233			
Anzahl Patienten mit Ereignis n (%)	285 (83,8)	172 (73,8)	1,84 (1,22; 2,77); 0,004	1,14 (1,04; 1,24); 0,006	10,00 (3,13; 16,87); 0,004
Andere					
N'	143	95			
Anzahl Patienten mit Ereignis n (%)	111 (77,6)	79 (83,2)	0,70 (0,36; 1,37); 0,300	0,93 (0,82; 1,06); 0,286	-5,54 (-15,70; 4,63); 0,326
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A3

Für den Endpunkt Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus nicht gleichgerichtet. Diese Ergebnisse waren jedoch statistisch nicht signifikant. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

mITT

Für den Endpunkt Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten, die seropositiv waren und mind. 3 Gelenkerosionen hatten, war Baricitinib statistisch signifikant überlegen zu Adalimumab. Für Patienten mit einem anderen Status war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 10 mm nach mTSS zu Baseline

Tabelle 4-129: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 10 mm nach mTSS zu Baseline					
mITT			Interaktionstest^a: p = 0,0425		
N	487	330			
\leq Median					
N´	246	153	0,98	1,00	-0,31
Anzahl Patienten mit Ereignis n (%)	197 (80,1)	123 (80,4)	(0,59; 1,63); 0,940	(0,90; 1,10); 0,940	(-8,34; 7,72); 1,000
$>$ Median					
N´	227	160	2,01	1,16	11,83
Anzahl Patienten mit Ereignis n (%)	190 (83,7)	115 (71,9)	(1,23; 3,29); 0,006	(1,04; 1,30); 0,008	(3,36; 20,29); 0,006
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm waren die Effektschätzer nach mTSS zu Baseline gleichgerichtet. Es wird nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 20 mm nach Alter

Tabelle 4-130: Ergebnisse für Anteil Patienten mit einer Verbesserung der PtGADA (VAS) \geq 20 aus RCT mit dem zu bewertenden Arzneimittel nach Alter

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 20 mm nach Alter					
Subpopulation A2			Interaktionstest^a: p = 0,0175		
N	243	153			
< 65 Jahre					
N´	184	127	1,04 (0,64; 1,69); 0,887	1,01 (0,87; 1,18); 0,888	0,76 (-9,78; 11,31); 0,902
Anzahl Patienten mit Ereignis n (%)	126 (68,5)	86 (67,7)			
\geq 65 Jahre					
N´	55	26	4,89 (1,76; 13,59); 0,003	1,85 (1,15; 2,98); 0,012	35,87 (13,97; 57,78); 0,002
Anzahl Patienten mit Ereignis n (%)	43 (78,2)	11 (42,3)			
mITT			Interaktionstest^a: p = 0,0338		
N	487	330			
< 65 Jahre					
N´	381	273	1,21 (0,86; 1,69); 0,274	1,06 (0,95; 1,18); 0,279	3,99 (-3,19; 11,17); 0,301
Anzahl Patienten mit Ereignis n (%)	272 (71,4)	184 (67,4)			
\geq 65 Jahre					
N´	102	55	3,03 (1,48; 6,20); 0,003	1,44 (1,10; 1,87); 0,007	23,89 (8,50; 39,28); 0,003
Anzahl Patienten mit Ereignis n (%)	80 (78,4)	30 (54,5)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Anteil Patienten mit einer Verbesserung der PtGADA (VAS) \geq 20 mm waren die Effektschätzer nach Alter gleichgerichtet. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

mITT

Für den Endpunkt Anteil Patienten mit einer Verbesserung der PtGADA (VAS) \geq 20 mm waren die Effektschätzer nach Alter gleichgerichtet. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 20 mm nach Gelenkerosionsstatus

Tabelle 4-131: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 20 mm nach Gelenkerosionsstatus					
Subpopulation A2 Interaktionstest^a: p = 0,0056					
N	243	153			
1-2 Gelenkerosionen + Seropositivität					
N'	59	40	0,45	0,81	-15,59
Anzahl Patienten mit Ereignis n (%)	38 (64,4)	32 (80,0)	(0,17; 1,17); 0,101	(0,63; 1,03); 0,086	(-33,00; 1,81); 0,117
Mindestens 3 Gelenkerosionen					
N'	179	111	1,88	1,24	14,07
Anzahl Patienten mit Ereignis n (%)	130 (72,6)	65 (58,6)	(1,14; 3,10); 0,014	(1,03; 1,49); 0,020	(2,81; 25,32); 0,015
mITT Interaktionstest^a: p = 0,0074					
N	487	330			
1-2 Gelenkerosionen + Seropositivität					
N'	115	80	0,68	0,89	-8,04
Anzahl Patienten mit Ereignis n (%)	77 (67,0)	60 (75,0)	(0,36; 1,28); 0,230	(0,74; 1,07); 0,219	(-20,85; 4,76); 0,266
Mindestens 3 Gelenkerosionen					
N'	367	245	1,77	1,20	12,21
Anzahl Patienten mit Ereignis n (%)	274 (74,7)	153 (62,4)	(1,25; 2,51); 0,001	(1,07; 1,34); 0,002	(4,69; 19,73); 0,002
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert.					
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Anteil Patienten mit einer Verbesserung der PtGADA (VAS) \geq 20 mm waren die Effektschätzer nach Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten mit 1-2 Gelenkerosionen + Seropositivität war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten mit mindestens 3 Gelenkerosionen war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

mITT

Für den Endpunkt Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm waren die Effektschätzer nach Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten mit 1-2 Gelenkerosionen + Seropositivität war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten mit mindestens 3 Gelenkerosionen war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm nach Seropositivität und Gelenkerosionsstatus

Tabelle 4-132: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm nach Seropositivität und Gelenkerosionsstatus					
Subpopulation A2			Interaktionstest^a: p = 0,0135		
N	243	153			
Seropositiv und mind. 3 Gelenkerosionen					
N´	164	103			
Anzahl Patienten mit Ereignis n (%)	121 (73,8)	60 (58,3)	2,02 (1,19; 3,41); 0,009	1,27 (1,05; 1,53); 0,014	15,53 (3,87; 27,19); 0,010
Andere					
N´	75	50			
Anzahl Patienten mit Ereignis n (%)	48 (64,0)	37 (74,0)	0,62 (0,28; 1,38); 0,244	0,86 (0,68; 1,10); 0,231	-10,00 (-26,30; 6,30); 0,328
mITT			Interaktionstest^a: p = 0,0066		
N	487	330			
Seropositiv und mind. 3 Gelenkerosionen					
N´	340	233			
Anzahl Patienten mit Ereignis n (%)	259 (76,2)	146 (62,7)	1,91 (1,32; 2,74); < 0,001	1,22 (1,08; 1,36); < 0,001	13,52 (5,83; 21,20); < 0,001
Andere					
N´	143	95			
Anzahl Patienten mit Ereignis n (%)	93 (65,0)	68 (71,6)	0,74 (0,42; 1,30); 0,292	0,91 (0,76; 1,08); 0,283	-6,54 (-18,52; 5,43); 0,324

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 20 mm nach Seropositivität und Gelenkerosionsstatus					
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Anteil Patienten mit einer Verbesserung der PtGADA (VAS) \geq 20 mm waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten, die seropositiv waren und mind. 3 Gelenkerosionen hatten, war Baricitinib statistisch signifikant überlegen zu Adalimumab. Für Patienten mit einem anderen Status war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

mITT

Für den Endpunkt Anteil Patienten mit einer Verbesserung der PtGADA (VAS) \geq 20 mm waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten, die seropositiv waren und mind. 3 Gelenkerosionen hatten, war Baricitinib statistisch signifikant überlegen zu Adalimumab. Für Patienten mit einem anderen Status war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm nach mTSS zu Baseline

Tabelle 4-133: Ergebnisse für Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm nach mTSS zu Baseline					
mITT Interaktionstest^a: p = 0,0244					
N	487	330			
\leq Median					
N´	246	153	0,98 (0,63; 1,52); 0,929	0,99 (0,87; 1,14); 0,929	-0,42 (-9,69; 8,84); 1,000
Anzahl Patienten mit Ereignis n (%)	171 (69,5)	107 (69,9)			
$>$ Median					
N´	227	160	2,03 (1,30; 3,15); 0,002	1,24 (1,08; 1,44); 0,003	14,96 (5,60; 24,32); 0,002
Anzahl Patienten mit Ereignis n (%)	73 (76,2)	98 (61,3)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm waren die Effektschätzer nach mTSS zu Baseline nicht gleichgerichtet. Für Patienten mit mTSS \leq Median war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten mit mTSS $>$ Median war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

4.3.1.3.5.8 Health Assessment Questionnaire-Disability Index (HAQ-DI) – RCT**Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$ nach Seropositivität (ACPA)**

Tabelle 4-134: Ergebnisse für Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$ aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität (ACPA)

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$ nach Seropositivität (ACPA)					
Subpopulation A3			Interaktionstest^a: p = 0,0075		
N	170	124			
ACPA positiv					
N´	153	112	1,56 (0,93; 2,61); 0,089	1,17 (0,97; 1,41); 0,097	10,11 (-1,52; 21,74); 0,091
Anzahl Patienten mit Ereignis n (%)	107 (69,9)	67 (59,8)			
ACPA negativ					
N´	17	12	0,10 (0,01; 1,09); 0,058	0,58 (0,35; 0,95); 0,033	-38,73 (-67,14; -10,31); 0,043
Anzahl Patienten mit Ereignis n (%)	9 (52,9)	11 (91,7)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A3

Für den Endpunkt Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$ waren die Effektschätzer nach Seropositivität (ACPA) nicht gleichgerichtet. Für ACPA positive Patienten war Baricitinib statistisch nicht signifikant verschieden zu Adalimumab. Für ACPA negative Patienten war Baricitinib statistisch signifikant verschieden zuungunsten von Baricitinib. Hierbei ist zu beachten, dass die Anzahl der Patienten in diesen Subgruppen nur 17, bzw. 12 Patienten betrug. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

4.3.1.3.5.9 European Quality of Life-5 Dimensions (EQ-5D) VAS – RCT**Veränderung der EQ-5D VAS nach Alter**

Tabelle 4-135: Ergebnisse für Veränderung der EQ-5D VAS aus RCT mit dem zu bewertenden Arzneimittel nach Alter

Baricitinib vs. Adalimumab			
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert
Veränderung der EQ-5D VAS nach Alter			
Subpopulation A2			Interaktionstest^a: p = 0,029
N	243	153	
< 65 Jahre			
N'	182	122	10,75 (5,21; 16,29); < 0,001
Baseline, MW (SD)	51,12 (21,02)	46,72 (20,72)	
Woche 52, MW (SD)	71,73 (22,51)	60,59 (25,83)	
LSM (SE)	22,17 (1,778)	11,42 (2,173)	
≥ 65 Jahre			
N'	55	26	-2,81 (-13,26; 7,64); 0,594
Baseline, MW (SD)	49,38 (20,80)	58,62 (20,80)	
Woche 52, MW (SD)	65,20 (20,94)	69,15 (22,95)	
LSM (SE)	13,22 (2,931)	16,03 (4,295)	
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.			

Subpopulation A2

Für den Endpunkt Veränderung des EQ-5D-5L (VAS) waren die Effektschätzer nach Alter nicht gleichgerichtet. In der Gruppe Patienten < 65 Jahre war Baricitinib statistisch signifikant überlegen zu Adalimumab. In der Gruppe Patienten ≥ 65 Jahre war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Veränderung der EQ-5D VAS nach Therapie mit Kortikosteroiden zu Baseline

Tabelle 4-136: Ergebnisse für Veränderung der EQ-5D VAS aus RCT mit dem zu bewertenden Arzneimittel nach Therapie mit Kortikosteroiden zu Baseline

Baricitinib vs. Adalimumab			
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert
Veränderung der EQ-5D VAS nach Therapie mit Kortikosteroiden zu Baseline			
mITT		Interaktionstest^a: p = 0,048	
N	487	330	
Ja			
N´	270	193	4,57 (0,43; 8,72); 0,031
Baseline, MW (SD)	50,03 (19,62)	49,24 (20,61)	
Woche 52, MW (SD)	70,60 (22,46)	65,83 (23,06)	
LSM (SE)	20,79 (1,362)	16,22 (1,611)	
Nein			
N´	209	127	11,40 (5,88; 16,92); < 0,001
Baseline, MW (SD)	52,41 (20,83)	51,54 (22,68)	
Woche 52, MW (SD)	71,42 (22,65)	59,84 (29,08)	
LSM (SE)	19,27 (1,725)	7,87 (2,213)	
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.			

mITT

Für den Endpunkt Veränderung der EQ-5D VAS waren die Effektschätzer nach Therapie mit Kortikosteroiden zu Baseline gleichgerichtet und statistisch signifikant zugunsten von Baricitinib. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm nach Seropositivität (ACPA)

Tabelle 4-137: Ergebnisse für Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität (ACPA)

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm nach Seropositivität (ACPA)					
mITT Interaktionstest^a: p = 0,0449					
N	487	330			
ACPA positiv					
N´	420	287	1,68 (1,23; 2,29); 0,001	1,22 (1,08; 1,38); 0,002	12,22 (4,93; 19,51); 0,001
Anzahl Patienten mit Ereignis n (%)	284 (67,6)	159 (55,4)			
ACPA negativ					
N´	60	34	0,67 (0,27; 1,64); 0,376	0,86 (0,63; 1,19); 0,360	-9,31 (-29,39; 10,76); 0,508
Anzahl Patienten mit Ereignis n (%)	35 (58,3)	23 (67,6)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm waren die Effektschätzer nach Seropositivität (ACPA) nicht gleichgerichtet. In der Gruppe der ACPA positiven Patienten war Baricitinib statistisch signifikant überlegen zu Adalimumab. In der Gruppe der ACPA negativen Patienten war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

4.3.1.3.5.9.1 Schwerste Müdigkeit (Worst Tiredness) - RCT**Veränderung der schwersten Müdigkeit nach Geschlecht**

Tabelle 4-138: Ergebnisse für Veränderung der schwersten Müdigkeit aus RCT mit dem zu bewertenden Arzneimittel nach Geschlecht

Baricitinib vs. Adalimumab				
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert	Hedges' g (95% KI)
Veränderung der schwersten Müdigkeit nach Geschlecht				
mITT			Interaktionstest^a: p = 0,017	
N	487	330		
Männlich				
N'	70	47	-1,48 (-2,25; -0,70); < 0,001	-0,70 (-1,08; -0,33)
Baseline, MW (SD)	5,83 (2,13)	5,98 (2,07)		
Woche 52, MW (SD)	3,05 (2,19)	4,00 (2,35)		
LSM (SE)	-2,97 (0,249)	-1,49 (0,304)		
Weiblich				
N'	207	143	-0,27 (-0,78; 0,25); 0,309	-0,11 (-0,32; 0,10)
Baseline, MW (SD)	6,30 (2,43)	6,46 (2,51)		
Woche 52, MW (SD)	3,40 (2,42)	3,68 (2,55)		
LSM (SE)	-2,91 (0,167)	-2,65 (0,201)		
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.				

mITT

Für den Endpunkt Veränderung der schwersten Müdigkeit waren die Effektschätzer nach Geschlecht gleichgerichtet. Es wird nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Veränderung der schwersten Müdigkeit nach Therapie mit Kortikosteroiden zu Baseline

Tabelle 4-139: Ergebnisse für Veränderung der schwersten Müdigkeit aus RCT mit dem zu bewertenden Arzneimittel nach Therapie mit Kortikosteroiden zu Baseline

Baricitinib vs. Adalimumab				
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der schwersten Müdigkeit nach Therapie mit Kortikosteroiden zu Baseline				
mITT Interaktionstest^a: p = 0,029				
N	487	330		
Ja				
N'	157	118	-0,17 (-0,75; 0,40); 0,554	-0,07 (-0,31; 0,17)
Baseline, MW (SD)	6,18 (2,37)	6,43 (2,38)		
Woche 52, MW (SD)	3,51 (2,45)	3,76 (2,48)		
LSM (SE)	-2,76 (0,191)	-2,59 (0,220)		
Nein				
N'	120	72	-1,15 (-1,82; -0,49); < 0,001	-0,51 (-0,81; -0,22)
Baseline, MW (SD)	6,18 (2,36)	6,19 (2,48)		
Woche 52, MW (SD)	3,07 (2,25)	3,75 (2,53)		
LSM (SE)	-3,15 (0,206)	-1,99 (0,266)		
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.				

mITT

Die Effektschätzer nach Begleiterscheinung waren für den Endpunkt Veränderung der schwersten Müdigkeit gleichgerichtet. Es wird nicht von einer Modifikation des Behandlungseffektes ausgegangen.

4.3.1.3.5.10 Short Form (36 item) Health Survey Version 2 akut (SF-36 v2 akut) – RCT**Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS \geq 5 nach spezifischer Region II**

Tabelle 4-140: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS \geq 5 aus RCT mit dem zu bewertenden Arzneimittel nach spezifischer Region II

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS \geq 5 nach spezifischer Region II					
mITT			Interaktionstest^a: p = 0,0128		
N	487	330			
Westeuropa					
N´	29	20	9,43	3,03	50,86
Anzahl Patienten mit Ereignis n (%)	22 (75,9)	5 (25,0)	(2,43; 36,62); 0,002	(1,35; 6,80); 0,008	(26,31; 75,41); 0,001
Andere					
N´	458	310	1,25	1,10	5,40
Anzahl Patienten mit Ereignis n (%)	270 (59,0)	166 (53,5)	(0,93; 1,67); 0,139	(0,97; 1,25); 0,144	(-1,75; 12,55); 0,158
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS \geq 5 waren die Effektschätzer nach spezifischer Region gleichgerichtet. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Veränderung der SF-36 v2 akut MCS nach Therapie mit Kortikosteroiden zu Baseline

Tabelle 4-141: Ergebnisse für Veränderung der SF-36 v2 akut MCS aus RCT mit dem zu bewertenden Arzneimittel nach Therapie mit Kortikosteroiden zu Baseline

Baricitinib vs. Adalimumab			
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert
Veränderung der SF-36 v2 akut MCS nach Therapie mit Kortikosteroiden zu Baseline			
mITT		Interaktionstest^a: p = 0,029	
N	487	330	
Ja			
N´	270	193	-0,35 (-2,11; 1,41); 0,693
Baseline, MW (SD)	46,05 (11,18)	45,63 (11,33)	
Woche 52, MW (SD)	49,65 (10,70)	49,73 (10,87)	
LSM (SE)	3,72 (0,578)	4,07 (0,684)	
Nein			
N´	209	127	2,52 (0,59; 4,45); 0,010
Baseline, MW (SD)	47,62 (11,81)	45,92 (12,84)	
Woche 52, MW (SD)	52,17 (9,77)	48,78 (12,03)	
LSM (SE)	4,86 (0,601)	2,34 (0,771)	
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.			

mITT

Für den Endpunkt Veränderung des SF-36 v2 akut MCS waren die Effektschätzer nach Therapie mit Kortikosteroiden zu Baseline nicht gleichgerichtet. Für die Gruppe der Patienten unter Kortikosteroidtherapie war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten, die keine Kortikosteroide erhielten, war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 nach Erkrankungsdauer

Tabelle 4-142: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel nach Erkrankungsdauer

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 nach Erkrankungsdauer					
Subpopulation A2			Interaktionstest^a: p = 0,0260		
N	243	153			
< 1 Jahr					
N'	48	35	2,06	1,62	15,95
Anzahl Patienten mit Ereignis n (%)	20 (41,7)	9 (25,7)	(0,79; 5,42); 0,139	(0,83; 3,15); 0,153	(-4,15; 36,06); 0,165
≥ 1 Jahr bis < 5 Jahre					
N'	81	55	2,87	2,04	22,63
Anzahl Patienten mit Ereignis n (%)	36 (44,4)	12 (21,8)	(1,31; 6,27); 0,009	(1,16; 3,57); 0,013	(7,26; 38,00); 0,010
≥ 5 Jahr bis ≥ 10 Jahre					
N'	51	23	1,50	1,35	7,67
Anzahl Patienten mit Ereignis n (%)	15 (29,4)	5 (21,7)	(0,46; 4,88); 0,495	(0,55; 3,33); 0,505	(-13,32; 28,66); 0,580
≥ 10 Jahre					
N'	63	40	0,53	0,67	-14,84
Anzahl Patienten mit Ereignis n (%)	19 (30,2)	18 (45,0)	(0,23; 1,21); 0,131	(0,40; 1,12); 0,126	(-33,98; 4,29); 0,144
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert.					
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 waren die Effektschätzer nach Erkrankungsdauer weitgehend gleichgerichtet. In der Gruppe der Patienten mit einer Erkrankungsdauer von ≥ 1 Jahr bis < 5 Jahre war Baricitinib statistisch signifikant überlegen zu Adalimumab. In den übrigen Gruppen bestand kein statistisch signifikanter Unterschied. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 nach Gelenkerosionsstatus

Tabelle 4-143: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 nach Gelenkerosionsstatus					
Subpopulation A2			Interaktionstest^a: p = 0,0482		
N	243	153			
1-2 Gelenkerosionen + Seropositivität					
N'	59	40	3,33 (1,34; 8,29); 0,010	2,18 (1,15; 4,14); 0,017	26,65 (8,48; 44,82); 0,011
Anzahl Patienten mit Ereignis n (%)	29 (49,2)	9 (22,5)			
Mindestens 3 Gelenkerosionen					
N'	183	111	1,09 (0,65; 1,80); 0,750	1,06 (0,75; 1,49); 0,751	1,80 (-9,21; 12,82); 0,798
Anzahl Patienten mit Ereignis n (%)	61 (33,3)	35 (31,5)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 waren die Effektschätzer nach Gelenkerosionsstatus gleichgerichtet. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 nach Seropositivität und Gelenkerosionsstatus

Tabelle 4-144: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 nach Seropositivität und Gelenkerosionsstatus					
Subpopulation A2			Interaktionstest^a: p = 0,0487		
N	243	153			
Seropositiv und mindestens 3 Gelenkerosionen					
N'	168	103			
Anzahl Patienten mit Ereignis n (%)	56 (33,3)	33 (32,0)	1,06 (0,63; 1,79); 0,826	1,04 (0,73; 1,48); 0,826	1,29 (-10,20; 12,78); 0,894
Andere					
N'	75	50			
Anzahl Patienten mit Ereignis n (%)	34 (45,3)	11 (22,0)	2,94 (1,30; 6,66); 0,010	2,06 (1,15; 3,69); 0,016	23,33 (7,25; 39,42); 0,008
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus gleichgerichtet. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 nach Art der Begleitmedikation

Tabelle 4-145: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel nach Art der Begleitmedikation

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 nach Art der Begleitmedikation					
mITT Interaktionstest^a: p = 0,0370					
N	487	330			
Nur MTX					
N'	413	277	1,08 (0,78; 1,50); 0,629	1,06 (0,85; 1,31); 0,630	1,77 (-5,38; 8,92); 0,680
Anzahl Patienten mit Ereignis n (%)	140 (33,9)	89 (32,1)			
MTX + andere cDMARDs					
N'	74	53	3,05 (1,24; 7,49); 0,016	2,33 (1,14; 4,77); 0,021	20,04 (5,51; 34,57); 0,015
Anzahl Patienten mit Ereignis n (%)	26 (35,1)	8 (15,1)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt Veränderung des SF-36 v2 akut MCS ≥ 5 waren die Effektschätzer nach Art der Begleitmedikation gleichgerichtet. Es wird nicht von einer Modifikation des Behandlungseffekts ausgegangen.

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 nach mTSS zu Baseline

Tabelle 4-146: Ergebnisse für Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5 nach mTSS zu Baseline					
Subpopulation A2			Interaktionstest^a: p = 0,0244		
N	243	153			
\leq Median					
N'	139	82	2,24 (1,21; 4,15); 0,011	1,74 (1,11; 2,71); 0,015	17,12 (4,87; 29,36); 0,012
Anzahl Patienten mit Ereignis n (%)	56 (40,3)	19 (23,2)			
$>$ Median					
N'	96	63	0,80 (0,41; 1,54); 0,498	0,87 (0,57; 1,31); 0,495	-5,31 (-20,68; 10,06); 0,506
Anzahl Patienten mit Ereignis n (%)	33 (34,4)	25 (39,7)			
mITT			Interaktionstest^a: p = 0,0439		
N	487	330			
\leq Median					
N'	246	153	1,63 (1,05; 2,54); 0,030	1,40 (1,02; 1,90); 0,034	10,60 (1,34; 19,86); 0,030
Anzahl Patienten mit Ereignis n (%)	92 (37,4)	41 (26,8)			
$>$ Median					
N'	227	160	0,86 (0,56; 1,33); 0,500	0,91 (0,68; 1,21); 0,498	-3,28 (-12,84; 6,27); 0,512
Anzahl Patienten mit Ereignis n (%)	72 (31,7)	56 (35,0)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

Subpopulation A2

Für den Endpunkt Veränderung des SF-36 v2 akut MCS ≥ 5 waren die Effektschätzer nach mTSS zu Baseline nicht gleichgerichtet. Für Patienten mit einem mTSS \leq Median war Baricitinib statistisch signifikant überlegen zu Adalimumab. Für Patienten mit einem mTSS $>$ Median war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

mITT

Für den Endpunkt Veränderung des SF-36 v2 akut MCS ≥ 5 waren die Effektschätzer nach mTSS zu Baseline nicht gleichgerichtet. Für Patienten mit einem mTSS \leq Median war Baricitinib statistisch signifikant überlegen zu Adalimumab. Für Patienten mit einem mTSS $>$ Median war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Veränderung der SF-36 v2 akut Domäne - Körperliche Funktionsfähigkeit nach mTSS zu Baseline

Tabelle 4-147: Ergebnisse für Veränderung der SF-36 v2 akut Domäne - Körperliche Funktionsfähigkeit aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline

Baricitinib vs. Adalimumab				
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert	Hedges' g (95% KI)
Veränderung der SF-36 v2 akut Domäne - Körperliche Funktionsfähigkeit nach mTSS zu Baseline				
mITT Interaktionstest^a: p = 0,031				
N	487	330		
\leq Median				
N'	245	152	-0,23 (-2,10; 1,63); 0,805	-0,03 (-0,23; 0,18)
Baseline, MW (SD)	32,76 (10,27)	33,04 (11,08)		
Woche 52, MW (SD)	42,86 (11,04)	43,11 (10,50)		
LSM (SE)	10,05 (0,588)	10,28 (0,747)		
$>$ Median				
N'	227	160	2,56 (0,74; 4,38); 0,006	0,28 (0,08; 0,49)
Baseline, MW (SD)	32,20 (10,22)	30,72 (10,31)		
Woche 52, MW (SD)	41,19 (10,65)	37,66 (11,97)		
LSM (SE)	9,20 (0,595)	6,64 (0,709)		
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.				

mITT

Für den Endpunkt Veränderung der SF-36 v2 akut Domäne - Körperliche Funktionsfähigkeit waren die Effektschätzer nach mTSS zu Baseline nicht gleichgerichtet. Für die Gruppe der Patienten mit einem mTSS \leq Median war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten mit einem mTSS $>$ Median war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Veränderung der SF-36 v2 akut Domäne – Allgemeiner Gesundheitszustand nach Gewicht zu Baseline

Tabelle 4-148: Ergebnisse für Veränderung der SF-36 v2 akut Domäne – Allgemeiner Gesundheitszustand aus RCT mit dem zu bewertenden Arzneimittel nach Gewicht zu Baseline

Baricitinib vs. Adalimumab				
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der SF-36 v2 akut Domäne – Allgemeiner Gesundheitszustand nach Gewicht zu Baseline				
mITT Interaktionstest^a: p = 0,046				
N	487	330		
< 60 kg				
N'	143	91	-0,82 (-3,03; 1,39); 0,466	-0,10 (-0,36; 0,16)
Baseline, MW (SD)	37,60 (7,65)	36,68 (9,69)		
Woche 52, MW (SD)	42,86 (9,14)	43,23 (9,56)		
LSM (SE)	5,44 (0,699)	6,26 (0,876)		
≥ 60 kg bis < 100 kg				
N'	308	216	2,05 (0,62; 3,49); 0,005	0,25 (0,07; 0,42)
Baseline, MW (SD)	37,41 (8,34)	36,36 (7,86)		
Woche 52, MW (SD)	43,66 (9,21)	40,99 (9,48)		
LSM (SE)	6,41 (0,469)	4,36 (0,560)		
≥ 100 kg				
N'	28	13	4,52 (-1,46; 10,49); 0,134	0,50 (-0,15; 1,16)
Baseline, MW (SD)	35,41 (8,53)	35,29 (14,24)		
Woche 52, MW (SD)	43,74 (10,17)	39,15 (12,11)		
LSM (SE)	8,35 (1,661)	3,83 (2,438)		
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.				

mITT

Für den Endpunkt Veränderung der SF-36 v2 akut Domäne – Allgemeiner Gesundheitszustand waren die Effektschätzer nach Gewicht zu Baseline nicht gleichgerichtet. Für Patienten mit einem Gewicht < 60 kg war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten mit einem Gewicht ≥ 60 kg bis < 100 kg war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Veränderung der SF-36 v2 akut Domäne - Emotionale Rollenfunktion nach Therapie mit Kortikosteroiden zu Baseline

Tabelle 4-149: Ergebnisse für Veränderung der SF-36 v2 akut Domäne - Emotionale Funktionsfähigkeit aus RCT mit dem zu bewertenden Arzneimittel nach Therapie mit Kortikosteroiden zu Baseline

Baricitinib vs. Adalimumab				
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert	Hedges'g (95% KI)
Veränderung der SF-36 v2 akut Domäne - Emotionale Rollenfunktion nach Therapie mit Kortikosteroiden zu Baseline				
Subpopulation A3		Interaktionstest^a: p = 0,038		
N	170	124		
Ja				
N'	100	73	-1,30 (-3,78; 1,17); 0,300	-0,16 (-0,46; 0,14)
Baseline, MW (SD)	43,15 (11,44)	40,43 (12,88)		
Woche 52, MW (SD)	48,67 (8,65)	49,14 (9,19)		
LSM (SE)	6,32 (0,812)	7,63 (0,951)		
Nein				
N'	69	48	2,63 (-1,04; 6,30); 0,158	0,27 (-0,10; 0,64)
Baseline, MW (SD)	46,08 (10,03)	40,22 (13,36)		
Woche 52, MW (SD)	49,53 (8,69)	44,48 (12,95)		
LSM (SE)	4,86 (1,165)	2,24 (1,405)		
mITT		Interaktionstest^a: p = 0,023		
N	487	330		
Ja				
N'	270	193	-0,05 (-1,87; 1,76); 0,953	-0,01 (-0,19; 0,18)
Baseline, MW (SD)	40,89 (12,41)	39,97 (12,77)		
Woche 52, MW (SD)	46,30 (10,20)	46,00 (11,08)		
LSM (SE)	5,72 (0,596)	5,77 (0,705)		
Nein				
N'	209	127	3,04 (1,04; 5,04); 0,003	0,34 (0,11; 0,56)
Baseline, MW (SD)	42,18 (12,60)	40,98 (12,91)		
Woche 52, MW (SD)	48,52 (8,90)	45,04 (11,98)		
LSM (SE)	6,63 (0,625)	3,58 (0,802)		
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.				

Subpopulation A3

Für den Endpunkt Veränderung der SF-36 v2 akut Domäne - Emotionale Rollenfunktion waren die Effektschätzer nach Therapie mit Kortikosteroiden zu Baseline nicht gleichgerichtet. Diese Ergebnisse waren jedoch statistisch nicht signifikant. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

mITT

Für den Endpunkt Veränderung der SF-36 v2 akut Domäne - Emotionale Rollenfunktion waren die Effektschätzer nach Therapie mit Kortikosteroiden zu Baseline nicht gleichgerichtet. Für die Gruppe der Patienten unter Kortikosteroidtherapie war Baricitinib nicht statistisch signifikant verschieden zu Adalimumab. Für Patienten, die keine Kortikosteroide erhielten, war Baricitinib statistisch signifikant überlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

4.3.1.3.5.11 Sicherheit – RCT**Unerwünschte Ereignisse****UE nach Gelenkerosionsstatus**

Tabelle 4-150: Ergebnisse für UE aus RCT mit dem zu bewertenden Arzneimittel nach Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
UE nach Gelenkerosionsstatus					
mITT			Interaktionstest^a: p = 0,0068		
N	487	330			
1-2 Gelenkerosionen + Seropositivität					
N'	115	82	0,49	0,86	-11,97
Anzahl Patienten mit Ereignis n (%)	83 (72,2)	69 (84,1)	(0,24; 1,01); 0,052	(0,74; 0,99); 0,043	(-23,36; -0,59); 0,059
Mindestens 3 Gelenkerosionen					
N'	371	245	1,46	1,09	6,58
Anzahl Patienten mit Ereignis n (%)	300 (80,9)	182 (74,3)	(0,99; 2,15); 0,054	(1,00; 1,19); 0,062	(-0,20; 13,36); 0,058
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert.					
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt Unerwünschte Ereignisse waren die Effektschätzer nach Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten mit 1-2 Gelenkerosionen + Seropositivität war Baricitinib statistisch signifikant überlegen zu Adalimumab. Für Patienten mit mindestens 3 Gelenkerosionen war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

UE nach Seropositivität und Gelenkerosionsstatus

Tabelle 4-151: Ergebnisse für UE aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
UE nach Seropositivität und Gelenkerosionsstatus					
mITT			Interaktionstest^a: p = 0,0061		
N	487	330			
Seropositivität und mind. 3 Gelenkerosionen					
N´	344	233	1,55	1,10	7,44
Anzahl Patienten mit Ereignis n (%)	281 (81,7)	173 (74,2)	(1,03; 2,31); 0,033	(1,00; 1,20); 0,040	(0,49; 14,38); 0,038
Andere					
N´	143	97	0,55	0,87	-10,45
Anzahl Patienten mit Ereignis n (%)	103 (72,0)	80 (82,5)	(0,29; 1,04); 0,065	(0,76; 1,00); 0,054	(-21,00; 0,11); 0,066
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt Unerwünschte Ereignisse waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten, die seropositiv waren und mind. 3 Gelenkerosionen hatten, war Baricitinib statistisch signifikant unterlegen. Für die Gruppe der Patienten mit anderem Status war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

Schwerwiegende Unerwünschte Ereignisse**SUE nach BMI zu Baseline**

Tabelle 4-152: Ergebnisse für SUE aus RCT mit dem zu bewertenden Arzneimittel nach BMI zu Baseline

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
SUE nach BMI zu Baseline					
mITT			Interaktionstest^a: p = 0,0191		
N	487	330			
< 25 kg/m²					
N´	211	148	7,22 (1,65; 31,67); 0,009	6,66 (1,57; 28,31); 0,010	7,65 (3,37; 11,94); 0,002
Anzahl Patienten mit Ereignis n (%)	19 (9,0)	2 (1,4)			
≥ 25 und < 30 kg/m²					
N´	161	108	2,28 (0,72; 7,24); 0,160	2,18 (0,73; 6,54); 0,164	4,37 (-1,14; 9,88); 0,203
Anzahl Patienten mit Ereignis n (%)	13 (8,1)	4 (3,7)			
≥ 30 kg/m²					
N´	114	74	0,53 (0,17; 1,66); 0,276	0,56 (0,19; 1,60); 0,275	-4,20 (-12,02; 3,63); 0,378
Anzahl Patienten mit Ereignis n (%)	6 (5,3)	7 (9,5)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt SUE waren die Effektschätzer nach BMI zu Baseline nicht gleichgerichtet. In der Gruppe der Patienten mit BMI zu Baseline < 25 kg/m² war Baricitinib statistisch signifikant unterlegen zu Adalimumab. In den anderen beiden BMI-Kategorien war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden. Hierbei ist zu beachten, dass die Ereignishäufigkeiten sehr niedrig waren.

SUE nach Erkrankungsdauer

Tabelle 4-153: Ergebnisse für SUE aus RCT mit dem zu bewertenden Arzneimittel nach Erkrankungsdauer

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
SUE nach Erkrankungsdauer					
mITT			Interaktionstest^a: p = 0,0475		
N	487	330			
< 5 Jahre					
N´	220	150	0,82	0,83	-1,00
Anzahl Patienten mit Ereignis n (%)	11 (5,0)	9 (6,0)	(0,33; 2,05); 0,677	(0,35; 1,97); 0,677	(-5,77; 3,77); 0,815
≥ 5 Jahr bis < 10 Jahre					
N´	108	66	4,00	3,67	8,08
Anzahl Patienten mit Ereignis n (%)	12 (11,1)	2 (3,0)	(0,86; 18,67); 0,078	(0,84; 16,04); 0,084	(0,85; 15,31); 0,083
≥ 10 Jahre					
N´	159	114	5,83	5,38	7,68
Anzahl Patienten mit Ereignis n (%)	15 (9,4)	2 (1,8)	(1,30; 26,21); 0,022	(1,25; 23,21); 0,024	(2,54; 12,82); 0,010
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt SUE waren die Effektschätzer nach Erkrankungsdauer nicht gleichgerichtet. Für Patienten mit einer Erkrankungsdauer < 5 Jahre, sowie ≥ 5 Jahre bis < 10 Jahre war Baricitinib statistisch nicht signifikant verschieden zu Adalimumab. Für Patienten mit einer Erkrankungsdauer ≥ 10 Jahre war Baricitinib statistisch signifikant unterlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden. Hierbei ist zu beachten, dass die Ereignishäufigkeiten sehr niedrig waren.

SUE nach Seropositivität und Gelenkerosionsstatus

Tabelle 4-154: Ergebnisse für SUE aus RCT mit dem zu bewertenden Arzneimittel nach Seropositivität und Gelenkerosionsstatus

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
SUE nach Seropositivität und Gelenkerosionsstatus					
mITT			Interaktionstest^a: p = 0,0292		
N	487	330			
Seropositiv und mind. 3 Gelenkerosionen					
N´	344	233	3,31	3,10	6,30
Anzahl Patienten mit Ereignis n (%)	32 (9,3)	7 (3,0)	(1,43; 7,65); 0,005	(1,39; 6,91); 0,006	(2,53; 10,07); 0,004
Andere					
N´	143	97	0,66	0,68	-1,99
Anzahl Patienten mit Ereignis n (%)	6 (4,2)	6 (6,2)	(0,21; 2,14); 0,491	(0,22; 2,05); 0,491	(-7,80; 3,82); 0,553
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt SUE waren die Effektschätzer nach Seropositivität und Gelenkerosionsstatus nicht gleichgerichtet. Für die Gruppe der Patienten, die seropositiv waren und mind. 3 Gelenkerosionen hatten, war Baricitinib statistisch signifikant unterlegen zu Adalimumab. Für die Gruppe der Patienten mit anderem Status war Baricitinib statistisch nicht signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden. Hierbei ist zu beachten, dass die Ereignishäufigkeiten sehr niedrig waren.

Schwere Unerwünschte Ereignisse (ohne Laborereignisse)**Schwere UE (ohne ausschließliche Laborereignisse) nach mTSS zu Baseline**

Tabelle 4-155: Ergebnisse für schwere UE (ohne ausschließliche Laborereignisse) aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
schwere UE (ohne ausschließliche Laborereignisse) nach mTSS zu Baseline					
mITT			Interaktionstest^a: p = 0,0148		
N	487	330			
≤ Median					
N´	246	153	0,70 (0,25; 1,98); 0,502	0,71 (0,26; 1,93); 0,501	-1,32 (-5,31; 2,66); 0,591
Anzahl Patienten mit Ereignis n (%)	8 (3,3)	7 (4,6)			
> Median					
N´	227	160	6,80 (1,55; 29,89); 0,011	6,34 (1,49; 27,08); 0,013	6,68 (2,77; 10,59); 0,004
Anzahl Patienten mit Ereignis n (%)	18 (7,9)	2 (1,3)			
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt schwere UE (ohne ausschließliche Laborereignisse) waren die Effektschätzer nach mTSS zu Baseline nicht gleichgerichtet. Für die Gruppe der Patienten mit mTSS ≤ Median war Baricitinib statistisch nicht signifikant verschieden zu Adalimumab. Für die Gruppe der Patienten mit mTSS > Median war Baricitinib statistisch signifikant unterlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden. Hierbei ist zu beachten, dass die Ereignishäufigkeiten sehr niedrig waren.

Therapieabbruch aufgrund von UE nach mTSS zu Baseline

Tabelle 4-156: Ergebnisse für Therapieabbruch aufgrund von UE aus RCT mit dem zu bewertenden Arzneimittel nach mTSS zu Baseline

Baricitinib vs. Adalimumab					
RA-BEAM (JADV)	Baricitinib	Adalimumab	OR (95% KI); p-Wert	RR (95% KI); p-Wert	ARD (95% KI); p-Wert
Therapieabbruch aufgrund von UE nach mTSS zu Baseline					
mITT			Interaktionstest^a: p = 0,0449		
N	487	330			
≤ Median					
N´	246	153	0,88	0,89	-0,51
Anzahl Patienten mit Ereignis n (%)	10 (4,1)	7 (4,6)	(0,33; 2,38); 0,806	(0,34; 2,29); 0,806	(-4,64; 3,62); 0,804
> Median					
N´	227	160	5,59	5,29	5,36
Anzahl Patienten mit Ereignis n (%)	15 (6,6)	2 (1,3)	(1,25; 24,91); 0,024	(1,22; 22,90); 0,026	(1,70; 9,02); 0,011
a: Logistisches Regressionsmodell für RR; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.					

mITT

Für den Endpunkt Therapieabbruch aufgrund von UE waren die Effektschätzer nach mTSS zu Baseline nicht gleichgerichtet. Für die Gruppe der Patienten mit mTSS ≤ Median war Baricitinib statistisch nicht signifikant verschieden zu Adalimumab. Für die Gruppe der Patienten mit mTSS > Median war Baricitinib nicht statistisch signifikant unterlegen. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden. Hierbei ist zu beachten, dass die Ereignishäufigkeiten sehr niedrig waren.

UE von besonderem Interesse**QIDS-SR16 nach Gewicht zu Baseline**

Tabelle 4-157: Ergebnisse für UE von besonderem Interesse – QIDS-SR16 aus RCT mit dem zu bewertenden Arzneimittel nach Gewicht zu Baseline

Baricitinib vs. Adalimumab				
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert	Hedges' g (95% KI)
QIDS-SR16 nach Gewicht zu Baseline				
mITT		Interaktionstest^a: p = 0,031		
N	487	330		
< 60 kg				
N´	140	89	0,27 (-0,61; 1,14); 0,547	0,08 (-0,18; 0,35)
LSM (SE)	-1,80 (0,276)	-2,06 (0,347)		
≥ 60 kg bis < 100 kg				
N´	311	208	0,18 (-0,46; 0,82); 0,584	0,05 (-0,13; 0,22)
LSM (SE)	-2,59 (0,205)	-2,77 (0,251)		
≥ 100 kg				
N´	28	12	-3,55 (-6,41; -0,69); 0,016	-0,86 (-1,55; -0,18)
LSM (SE)	-4,24 (0,758)	-0,69 (1,172)		
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.				

mITT

Für den Endpunkt QIDS-SR16 waren die Effektschätzer nach Gewicht zu Baseline nicht gleichgerichtet. Für Patienten mit einem Gewicht ≥ 100 kg war Baricitinib statistisch signifikant überlegen zu Adalimumab. Für Patienten mit einem Gewicht < 60 kg sowie ≥ 60 kg bis < 100 kg war Baricitinib nicht statistisch signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden. Hierbei ist zu beachten, dass die Anzahl der Patienten in der Gruppe ≥ 100 kg sehr niedrig war.

QIDS-SR16 nach Krankheitsaktivität zu Baseline

Tabelle 4-158: Ergebnisse für UE von besonderem Interesse – QIDS-SR16 aus RCT mit dem zu bewertenden Arzneimittel nach Krankheitsaktivität zu Baseline

Baricitinib vs. Adalimumab				
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert	Hedges'g (95% KI)
QIDS-SR16 nach Krankheitsaktivität zu Baseline				
mITT				
Interaktionstest^a: p = 0,043				
N	487	330		
≤ 5,1				
N'	114	76	-0,78	-0,26 (-0,56; 0,03)
LSM (SE)	-1,85 (0,275)	-1,07 (0,337)	(-1,64; 0,08); 0,075	
> 5,1				
N'	365	233	0,34	0,09 (-0,07; 0,26)
LSM (SE)	-2,63 (0,194)	-2,97 (0,243)	(-0,27; 0,95); 0,274	
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.				

mITT

Für den Endpunkt QIDS-SR16 waren die Effektschätzer nach DAS28-hsCRP nicht gleichgerichtet, jedoch auch nicht statistisch signifikant verschieden zu Adalimumab. Es wird daher nicht von einer Modifikation des Behandlungseffekts ausgegangen.

QIDS-SR16 nach Therapie mit Kortikosteroiden zu Baseline

Tabelle 4-159: Ergebnisse für UE von besonderem Interesse – QIDS-SR16 aus RCT mit dem zu bewertenden Arzneimittel nach Therapie mit Kortikosteroiden zu Baseline

Baricitinib vs. Adalimumab				
RA-BEAM (JADV)	Baricitinib	Adalimumab	LSMD (95% KI); p-Wert	Hedges'g (95% KI)
QIDS-SR16 nach Therapie mit Kortikosteroiden zu Baseline				
mITT				
Interaktionstest^a: p = 0,007				
N	487	330		
Ja				
N'	269	186	0,67 (0,01; 1,33); 0,047	0,19 (0,00; 0,38)
LSM (SE)	-2,02 (0,214)	-2,69 (0,258)		
Nein				
N'	210	123	-0,75 (-1,54; 0,05); 0,067	-0,21 (-0,43; 0,01)
LSM (SE)	-2,98 (0,246)	-2,23 (0,322)		
a: p-Wert wurde über ein ANCOVA Modell berechnet; Interaktions-p-Wert. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.				

mITT

Für den Endpunkt QIDS-SR16 waren die Effektschätzer nach Therapie mit Kortikosteroiden zu Baseline nicht gleichgerichtet. Für die Patienten mit Kortikosteroidbehandlung war Baricitinib statistisch signifikant unterlegen zu Adalimumab. Für die Patienten ohne Kortikosteroidbehandlung war Baricitinib statistisch nicht signifikant verschieden. Eine Modifikation des Behandlungseffekts kann nicht ausgeschlossen werden.

4.3.1.3.6 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.

Im Anwendungsgebiet A werden ausschließlich Informationen und Ergebnisse zur Zielpopulation Patienten mit vorangegangene Behandlung mit einem oder mehreren **cDMARDs**, einschließlich Methotrexat, abgebildet und der medizinische Nutzen und der medizinische Zusatznutzen vergleichend zu der vom G-BA genannten zVT im Anwendungsgebiet A bewertet.

Die Zielpopulation im Anwendungsgebiet A wurde vom G-BA in folgende Subpopulationen unterteilt:

- Subpopulation A1: Patienten, bei denen **keine ungünstigen Prognosefaktoren**¹ vorliegen und die unzureichend auf eine vorangegangene Behandlung mit **einem** cDMARD (inklusive MTX) ansprechen.
- Subpopulation A2: Patienten **mit ungünstigen Prognosefaktoren**¹, die unzureichend auf eine vorangegangene Behandlung mit **einem** cDMARD (inklusive MTX) ansprechen.
- Subpopulation A3: Patienten, die unzureichend auf eine vorangegangene Behandlung mit **mehreren** cDMARDs (inklusive MTX) ansprechen.

Für Patienten der Subpopulation A1 hat der G-BA die zVT alternative cDMARDs, sofern geeignet (z. B. MTX, Leflunomid) als Mono- oder Kombinationstherapie, bestimmt. Derzeit liegt keine direkte Evidenz zum Vergleich von Baricitinib mit der vom G-BA genannten zVT für diese Subpopulation vor.

Für die Subpopulationen A2 und A3 hat der G-BA die zVT bDMARDs in Kombination mit MTX (Adalimumab oder Etanercept oder Certolizumab-Pegol oder Golimumab oder Abatacept oder Tocilizumab); ggf. als Monotherapie unter Berücksichtigung des jeweiligen Zulassungsstatus bei MTX-Unverträglichkeit benannt.

¹ Beispiele für ungünstige Prognosefaktoren:

- Nachweis von Autoantikörpern (z. B. Rheumafaktoren, hohe Spiegel von Antikörpern gegen citrullinierte Peptid Antigene)
- Hohe Krankheitsaktivität (nachgewiesen durch DAS- bzw. DAS28-Bewertungssystem, geschwollene Gelenke, Parameter der Akute-Phase-Reaktion wie z. B. C-Reaktives Protein, Erythrozytensedimentationsrate)
- Frühes Auftreten von Gelenkerosionen.

Quelle: [2].

Aus Sicht der *Lilly Deutschland GmbH* besteht aufgrund der Leitlinienempfehlungen und der Bestimmung der zVT durch den G-BA die Notwendigkeit, für den Nachweis des Zusatznutzens von Baricitinib in den Subpopulationen A2 und A3 die gesamte Studienpopulation (mITT Population) der Zulassungsstudie RA-BEAM (JADV) heranzuziehen.

Bei den Subpopulationen A2 und A3 handelt es sich um Patienten, die gemäß den Leitlinien für eine Zweitlinientherapie mit bDMARDs in Frage kommen. Die Leitlinien geben dabei eine Empfehlung, wann die Eskalation indiziert ist. So ist bei Patienten mit mehreren cDMARDs in der Vortherapie oder einem cDMARD in der Vortherapie und gleichzeitigem Vorliegen ungünstiger Prognosefaktoren eine Zweitlinientherapie mit bDMARDs indiziert. Konsequenterweise hat der G-BA für diese Patienten die gleiche zVT (bDMARDs in Kombination mit MTX (Adalimumab oder Etanercept oder Certolizumab-Pegol oder Golimumab oder Abatacept oder Tocilizumab); ggf. als Monotherapie unter Berücksichtigung des jeweiligen Zulassungsstatus bei MTX-Unverträglichkeit) festgelegt.

Dabei empfiehlt die Leitlinie in jedem Fall ein individuelles Vorgehen, d. h. neben der aktuellen Therapiesituation und dem Vorliegen von ungünstigen Prognosefaktoren ist immer die jeweilige individuelle Patientensituation zu berücksichtigen [4]. Um das Therapieziel Remission oder niedrige Krankheitsaktivität zu erreichen, gibt die Leitlinie somit eine Entscheidungshilfe bei der Fragestellung des Beginns der Zweitlinientherapie.

Die Studie RA-BEAM (JADV) vergleicht Baricitinib mit der vom G-BA bestimmten zVT (in diesem Fall Adalimumab) bei jenen Patienten, für die gemäß der S1-Leitlinie eine erstmalige Behandlung mit Biologika vorgesehen ist [5]. Die Studie umfasst Eskalations-würdige Patienten und ist somit in ihrer Gesamtheit geeignet, die Fragestellung nach einem Zusatznutzen von Baricitinib gegenüber dem *Standard of Care* in der Zweitlinie zu beantworten. Ein Auftrennen der Studie generiert für die Nutzenbewertung keine zusätzliche Information für die Bewertung der Zweitlinientherapie.

Vor diesem Hintergrund wurde für jeden Endpunkt der RA-BEAM (JADV) untersucht, ob der Behandlungseffekt zwischen den verschiedenen Subpopulationen A2 und A3 heterogen, d. h. unterschiedlich ist (Abschnitt 4.3.1.3). Da diese Heterogenitätstests an Woche 52 unauffällig waren und demnach keine Heterogenität zeigten, gehen wir davon aus, dass die über die gesamte Studienpopulation zu beobachtenden Effekte auch auf die verschiedenen Subpopulationen übertragbar sind. Demnach können die Ergebnisse der mITT Population herangezogen werden, um den Zusatznutzen von Baricitinib in der Zweitlinientherapie abzuleiten.

Für die Subpopulationen A2 und A3 wurden lediglich die Patienten berücksichtigt, die gemäß Label Baricitinib in Kombination mit MTX erhalten hatten. Die mITT Population schließt neben diesen Patienten auch Patienten ein, die während der Studie neben MTX noch andere cDMARDs erhalten haben. Der prozentuale Anteil dieser Patienten in der Studienpopulation (15,5%) liegt unter dem vom IQWiG definierten Schwellenwert zum Einschluss von Studien (20%) (IQWiG Methodenpapier; Abschnitt 8.1.1 [6]) und erfüllt somit dieses Kriterium.

Daher kann bei der Ableitung eines Zusatznutzens unter Heranziehung der mITT Population die vollständige Studienpopulation verwendet werden, ohne der Notwendigkeit Patienten herauszurechnen, die zusätzlich zu MTX noch weitere cDMARDs erhalten hatten.

Im Folgenden werden die Ergebnisse von Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 aus der randomisierten, aktiv kontrollierten Phase-III-Studie RA-BEAM (JADV) zusammenfassend dargestellt. Dabei erfolgt, der oben dargestellten Rationale folgend, zunächst die Darstellung der Ergebnisse der mITT, dann der Subpopulationen A2 und A3.

Mortalität

Da Mortalität nicht als primärer oder sekundärer Endpunkt vorgesehen war und nicht im Rahmen einer Überlebenszeitanalyse untersucht wurde, wurde die Anzahl der Todesfälle im Rahmen der schwerwiegenden unerwünschten Ereignisse als Annäherung zur Beurteilung dieses patientenrelevanten Endpunktes ausgewertet (Abschnitt 4.3.1.3.4).

Morbidität

Druckschmerzhafte Gelenke und geschwollene Gelenke

Veränderung der Anzahl der druckschmerzhaften Gelenke und geschwollenen Gelenke (basierend auf 28 Gelenken)

In der mITT nahm die Anzahl der druckschmerzhaften Gelenke und auch die Anzahl der geschwollenen Gelenke unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab stärker ab. Dieser Unterschied war von Baseline bis Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung der Symptome Gelenkentzündung und der damit einhergehenden Schwellung.

Baricitinib versus Adalimumab – mITT

Veränderung der Anzahl der druckschmerzhaften Gelenke von Baseline bis Woche 52:

LSMD (95% KI): -0,9 (-1,6; -0,1); p = 0,032

Veränderung der Anzahl der geschwollenen Gelenke von Baseline bis Woche 52:

LSMD (95% KI): -0,9 (-1,5; -0,2); p = 0,007

In der mITT nahm die Anzahl der druckschmerzhaften und der geschwollenen Gelenke im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter ab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-25, Tabelle 4-26).

In den Subpopulationen A2 nahm die Anzahl der druckschmerzhaften Gelenke und auch die Anzahl der geschwollenen Gelenke unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab stärker ab. In der Subpopulation A3 nahm die Anzahl der druckschmerzhaften Gelenke von Baseline bis Woche 24 unter Adalimumab und von Baseline bis Woche 52 unter Baricitinib stärker ab. Die Anzahl der geschwollenen Gelenke nahm in der Subpopulation A3 unter Baricitinib von Baseline bis Woche 24 und Woche 52 im Vergleich zu Adalimumab stärker ab. Diese Unterschiede waren in den Subpopulationen A2 und A3 statistisch nicht signifikant.

In der Subpopulation A2 nahm die Anzahl der geschwollenen Gelenke im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter ab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-25, Tabelle 4-26).

Disease Activity Score 28 High Sensitivity C-Reactive Protein (DAS28-hsCRP)

Veränderung des DAS28-hsCRP

In der mITT Population konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Dieser Unterschied war sowohl von Baseline bis Woche 24 als auch bis Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Reduktion der Krankheitsaktivität (Tabelle 4-29).

Baricitinib versus Adalimumab – mITT

Veränderung des DAS28-hsCRP (mBOCF) von Baseline bis Woche 24:

LSMD (95% KI): -0,21 (-0,39; -0,03); $p = 0,022$

Veränderung des DAS28-hsCRP (mBOCF) von Baseline bis Woche 52:

LSMD (95% KI): -0,29 (-0,49; -0,08); $p = 0,007$

In den Subpopulationen A2 und A3 konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant. (Tabelle 4-29).

Die Ergebnisse der Sensitivitätsanalyse (mLOCF) haben dieselbe Richtung wie die Ergebnisse der Hauptanalyse (mBOCF). Sie unterstützen daher die Aussage und die Robustheit der Ergebnisse der Hauptanalyse.

Baricitinib versus Adalimumab – mITT

Veränderung des DAS28-hsCRP (mLOCF) von Baseline bis Woche 52:

LSMD (95% KI): -0,35 (-0,54; -0,16); $p < 0,001$

Veränderung des DAS28-hsCRP (MMRM) von Baseline bis Woche 24:

LSMD (95% KI): -0,19 (-0,36; -0,02); $p = 0,029$

Veränderung des DAS28-hsCRP (MMRM) von Baseline bis Woche 52:

LSMD (95% KI): -0,23 (-0,41; -0,05); p = 0,014

Baricitinib versus Adalimumab – A3

LSMD (95% KI): -0,35 (-0,64; -0,06); p = 0,020

Anteil der Patienten mit DAS28-hsCRP \leq 3,2

In der mITT Population erreichte ein höherer Anteil der Patienten von Baseline bis Woche 24 und Woche 52 eine niedrige Krankheitsaktivität unter Baricitinib als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-32).

In der Subpopulation A2 erreichte ein höherer Anteil der Patienten von Baseline bis Woche 24, jedoch von Baseline bis Woche 52 ein niedrigerer Anteil der Patienten unter Baricitinib eine Remission als unter Adalimumab. In der Subpopulation A3 erreichte ein höherer Anteil der Patienten von Baseline bis Woche 24 und Woche 52 unter Baricitinib eine Remission als unter Adalimumab. Der Unterschied war für die Subpopulationen A2 und A3 zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-32).

Anteil der Patienten mit DAS28-hsCRP < 2,6

In der mITT Population erreichte ein höherer Anteil der Patienten von Baseline bis Woche 24 und Woche 52 unter Baricitinib eine Remission als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-33).

In der Subpopulation A2 erreichte ein höherer Anteil der Patienten von Baseline bis Woche 24, jedoch von Baseline bis Woche 52 ein niedrigerer Anteil der Patienten unter Baricitinib eine Remission als unter Adalimumab. In der Subpopulation A3 erreichte ein höherer Anteil der Patienten von Baseline bis Woche 24 und Woche 52 unter Baricitinib eine Remission als unter Adalimumab. Der Unterschied war für die Subpopulationen A2 und A3 zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-33).

Simplified Disease Activity Index (SDAI)

Veränderung des SDAI

In der mITT Population konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Es lag ein statistisch signifikanter Unterschied zu beiden Zeitpunkten vor. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Reduktion der Krankheitsaktivität.

Baricitinib versus Adalimumab – mITT

Veränderung des SDAI von Baseline bis Woche 24:

LSMD (95% KI): -2,43 (-4,28; -0,59); p = 0,010

Veränderung des SDAI von Baseline bis Woche 52:

LSMD (95% KI): -3,18 (-5,04; -1,32); p < 0,001

In der mITT Population nahm die Krankheitsaktivität im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter ab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-36).

In den Subpopulationen A2 und A3 konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Dieser Unterschied zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-36).

In den Subpopulationen A2 und A3 nahm die Krankheitsaktivität im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter ab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-36).

Anteil der Patienten mit SDAI $\leq 3,3$

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine Remission als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-37).

In den Subpopulationen A2 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine Remission als unter Adalimumab. In der Subpopulation A3 erreichte ein niedrigerer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24, jedoch bis Woche 52 ein höherer Anteil Patienten unter Baricitinib eine Remission. Der Unterschied war für die Subpopulationen A2 und A3 zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-37).

Clinical Disease Activity Index (CDAI)

Veränderung des CDAI

In der mITT Population konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Es lag ein statistisch signifikanter Unterschied zu beiden Zeitpunkten vor. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Reduktion der Krankheitsaktivität.

Baricitinib versus Adalimumab – mITT

Veränderung des CDAI von Baseline bis Woche 24:

LSMD (95% KI): -2,06 (-3,83; -0,29); p = 0,023

Veränderung des SDAI von Baseline bis Woche 52:

LSMD (95% KI): -2,85 (-4,62; -1,09); p = 0,002

In der mITT nahm die Krankheitsaktivität im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter ab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-40).

In den Subpopulationen A2 und A3 konnte unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität beobachtet werden. Dieser Unterschied war für die Subpopulationen A2 von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Für die Subpopulationen A3 war der Unterschied zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-40).

Baricitinib versus Adalimumab – Subpopulation A2

Veränderung des CDAI von Baseline bis Woche 52:

LSMD (95% KI): -2,70 (-5,37; -0,02); p = 0,048

In den Subpopulationen A2 und A3 nahm die Verbesserung der Krankheitsaktivität im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-40).

Anteil der Patienten mit CDAI ≤ 10

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine niedrige Krankheitsaktivität als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-41).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine niedrige Krankheitsaktivität als unter Adalimumab. Der Unterschied war für die Subpopulationen A2 und A3 zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-41).

EULAR Responsekriterien („good Response“) (DAS28-basiert)

Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 ein gutes Therapieansprechen gemäß EULAR-Kriterien als unter Adalimumab. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-44).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 ein gutes Therapieansprechen gemäß EULAR-Kriterien als unter Adalimumab. Der Unterschied war in beiden Subpopulationen zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-44).

Patient Assessment of Pain (PAP) (VAS)***Veränderung des PAP (VAS)***

In der mITT Population wurde von den Patienten unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 weniger Schmerzen berichtet. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant.

Baricitinib versus Adalimumab – mITT

Veränderung des PAP von Baseline bis Woche 24:

LSMD (95% KI): -4,8 (-8,0; -1,6); p = 0,003

Veränderung des PAP von Baseline bis Woche 52:

LSMD (95% KI): -5,9 (-9,1; -2,6); p < 0,001

In der mITT Population kam es im Studienverlauf, d. h. von Woche 24 bis Woche 52, zu einer weiteren Verringerung der Schmerzen unter Baricitinib im Vergleich zu Adalimumab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-47).

In den Subpopulationen A2 und A3 wurden von den Patienten unter Baricitinib im Vergleich zu Adalimumab von Baseline bis Woche 24 und Woche 52 weniger Schmerzen berichtet. Für die Subpopulation A2 war der Unterschied zu beiden Zeitpunkten statistisch nicht signifikant. Für die Subpopulation A3 war dieser Unterschied von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies zeigt eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen.

Baricitinib versus Adalimumab – Subpopulation A3

Veränderung des PAP von Baseline bis Woche 52:

LSMD (95% KI): -5,9 (-11,0; -0,8); p = 0,023

In den Subpopulationen A2 und A3 kam es im Studienverlauf, d. h. von Woche 24 bis Woche 52, zu einer weiteren Verringerung der Schmerzen unter Baricitinib im Vergleich zu Adalimumab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-47).

Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Dieser Unterschied war von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen (Tabelle 4-48).

Baricitinib versus Adalimumab – mITT

Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 10 mm von Baseline bis Woche 52:

83,2% vs. 73,5%; RR (95% KI): 1,13 (1,05; 1,22); p = 0,001

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Dieser Unterschied war in beiden Subpopulationen von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen (Tabelle 4-48).

Baricitinib versus Adalimumab – Subpopulation A2

Anteil Patienten mit PAP (VAS) ≥ 10 mm von Baseline bis Woche 52:

82,0% vs. 73,3%; RR (95% KI): 1,14 (1,02; 1,27); p = 0,018

Baricitinib versus Adalimumab – Subpopulation A3

Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 10 mm von Baseline bis Woche 52:

84,1% vs. 75,6%; RR (95% KI): 1,14 (1,01; 1,28); p = 0,031

Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. Der Unterschied war in der mITT Population von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen (Tabelle 4-49).

Baricitinib versus Adalimumab – mITT

Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 20 mm von Baseline bis Woche 52:

73,3% vs. 62,5%; RR (95% KI): 1,18 (1,07; 1,30); p = 0,001

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Schmerzen als unter Adalimumab. In der Subpopulation A2 war dieser Unterschied zu beiden Zeitpunkten statistisch nicht signifikant. In der Subpopulation A3 war der Unterschied von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verringerung der Schmerzen (Tabelle 4-49).

Baricitinib versus Adalimumab – Subpopulation A3

Anteil Patienten mit einer Veränderung des (VAS) ≥ 20 mm von Baseline bis Woche 52:

75,9% vs. 64,2%; RR (95% KI): 1,23 (1,05; 1,44); p = 0,012

Patient's Global Assessment of Disease Activity (PtGADA) (VAS)

Veränderung des PtGADA (VAS)

In der mITT Population wurde aus Sicht des Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität berichtet als unter Adalimumab. In der mITT Population war dieser Unterschied zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Reduktion der Krankheitsaktivität.

Baricitinib versus Adalimumab – mITT

Veränderung des PtGADA von Baseline bis Woche 24:

LSMD (95% KI): -4,0 (-7,1; -0,8); p = 0,013

Veränderung des PtGADA von Baseline bis Woche 52:

LSMD (95% KI): -6,0 (-9,2; -2,8); p < 0,001

In der mITT Population kam es im Studienverlauf, d. h. von Woche 24 bis Woche 52, zu einer zunehmenden Verbesserung der Krankheitsaktivität aus Sicht des Patienten unter Baricitinib im Vergleich zu Adalimumab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-52).

In den Subpopulationen A2 und A3 wurde aus Sicht des Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine niedrigere Krankheitsaktivität berichtet als unter Adalimumab. Für die Subpopulationen A2 und A3 war der Unterschied von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant.

Baricitinib versus Adalimumab – Subpopulation A2

Veränderung des PtGADA von Baseline bis Woche 52:

LSMD (95% KI): -6,0 (-10,8; -1,2); p = 0,015

Baricitinib versus Adalimumab – Subpopulation A3

Veränderung des PtGADA von Baseline bis Woche 52:

LSMD (95% KI): -5,2 (-10,4; -0,1); p = 0,046

In den Subpopulationen A2 und A3 kam es im Studienverlauf, d. h. von Woche 24 bis Woche 52, zu einer zunehmenden Verbesserung der Krankheitsaktivität aus Sicht des Patienten unter Baricitinib im Vergleich zu Adalimumab. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-52).

Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 10 mm

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-53).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Dieser Unterschied war in beiden Subpopulationen zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-53).

Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Dieser Unterschied war zu Woche 52, jedoch nicht zu Woche 24, statistisch signifikant (Tabelle 4-54).

Baricitinib versus Adalimumab – mITT

Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm von Baseline bis Woche 52:

72,9% vs. 65,2%; RR (95% KI): 1,10 (1,00; 1,21); p = 0,043

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verringerung der Krankheitsaktivität aus Sicht des Patienten als unter Adalimumab. Dieser Unterschied war in beiden Subpopulationen zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-54).

Health Assessment Questionnaire-Disability Index (HAQ-DI)

Veränderung des HAQ-DI

In der mITT Population waren die krankheitsassoziierten Einschränkungen der physischen Funktionalität unter Baricitinib von Baseline bis Woche 24 und Woche 52 niedriger als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung der physischen Funktionalität (Tabelle 4-57).

Baricitinib versus Adalimumab – mITT

Veränderung des HAQ-DI von Baseline bis Woche 24:

LSMD (95% KI): -0,11 (-0,18; -0,03); p = 0,008

Veränderung des HAQ-DI von Baseline bis Woche 52:

LSMD (95% KI): -0,09 (-0,17; -0,00); p = 0,039

In den Subpopulationen A2 und A3 waren die krankheitsassoziierten Einschränkungen der physischen Funktionalität unter Baricitinib von Baseline bis Woche 24 und Woche 52 niedriger als unter Adalimumab. In den beiden Subpopulationen war dieser Unterschied zu beiden Zeitpunkten statistisch nicht signifikant. (Tabelle 4-57).

Die Ergebnisse der Sensitivitätsanalyse (mLOCF) haben dieselbe Richtung wie die Ergebnisse der Hauptanalyse (mBOCF). Sie unterstützen daher die Aussage und die Robustheit der Ergebnisse der Hauptanalyse.

Baricitinib versus Adalimumab – mITT

Veränderung des HAQ-DI (mLOCF) von Baseline bis Woche 52:

LSMD (95% KI): -0,11 (-0,19; -0,03); p = 0,007

Veränderung des HAQ-DI (MMRM) von Baseline bis Woche 24:

LSMD (95% KI): -0,12 (-0,2; -0,0); p = 0,002

Veränderung des HAQ-DI (MMRM) von Baseline bis Woche 52:

LSMD (95% KI): -0,08 (-0,2; -0,0); p = 0,044

Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine bedeutsame Verbesserung der körperlichen Funktionen als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung der physischen Funktionalität (Tabelle 4-60).

Baricitinib versus Adalimumab – mITT

Anteil Patienten mit HAQ-DI $\geq 0,22$ von Baseline bis Woche 24:

72,5% vs. 63,9%; RR (95% KI): 1,12 (1,02; 1,23); p = 0,021

Anteil Patienten mit HAQ-DI $\geq 0,22$ von Baseline bis Woche 52:

67,6% vs. 58,2%; RR (95% KI): 1,14 (1,03; 1,27); p = 0,016

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine bedeutsame Verbesserung der physischen Funktionen als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten für beide Subpopulationen statistisch nicht signifikant (Tabelle 4-60).

Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine bedeutsame Verbesserung der körperlichen Funktionen als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-61).

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine bedeutsame Verbesserung der physischen Funktionen als unter Adalimumab. Dieser Unterschied war in beiden Subpopulationen zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-61).

Dauer der Morgensteifigkeit (min)

Veränderung der Dauer der Morgensteifigkeit (min)

In der mITT Population verkürzte sich die Dauer der Morgensteifigkeit unter Baricitinib von Baseline bis Woche 24 um 55 min und bis Woche 52 um 61 min, hingegen unter Adalimumab nur um 34 min (beide Zeitpunkte). Der Unterschied zwischen den Behandlungsarmen Baricitinib und Adalimumab betrug 21 min zu Woche 24 und 27 min zu Woche 52 mit statistisch signifikantem Effekt zugunsten von Baricitinib zu Woche 52. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verkürzung der Dauer der Morgensteifigkeit.

Baricitinib versus Adalimumab – mITT

Veränderung der Dauer der Morgensteifigkeit von Baseline bis Woche 52:

LSMD (95% KI): -27,34 (-51,35; -3,32); p = 0,026

In der mITT Populationen verkürzte sich die Dauer der Morgensteifigkeit im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-64).

In der Subpopulation A2 verkürzte sich die Dauer der Morgensteifigkeit unter Baricitinib von Baseline bis Woche 24 um 72 min und bis Woche 52 um 71 min, hingegen unter Adalimumab nur um 17 min bzw. um 29 min. Der Unterschied zwischen den Behandlungsarmen Baricitinib und Adalimumab betrug 54 min zu Woche 24 und 41 min zu Woche 52 mit statistisch signifikantem Effekt zugunsten von Baricitinib zu beiden Zeitpunkten. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verkürzung der Dauer der Morgensteifigkeit (Tabelle 4-64).

Baricitinib versus Adalimumab – Subpopulation A2

Veränderung der Dauer der Morgensteifigkeit von Baseline bis Woche 24:

LSMD (95% KI): -54,3 (-91,08; -17,66); $p = 0,004$

Veränderung der Dauer der Morgensteifigkeit von Baseline bis Woche 52:

LSMD (95% KI): -41,49 (-80,56; -2,42); $p = 0,038$

In der Subpopulation A3 verkürzte sich die Dauer der Morgensteifigkeit unter Baricitinib von Baseline bis Woche 24 um 55 min und bis Woche 52 um 68 min und unter Adalimumab um 50 min bzw. 44 min. Der Unterschied zwischen den Behandlungsarmen Baricitinib und Adalimumab betrug 6 min zu Woche 24 zugunsten von Adalimumab und 24 min zu Woche 52 zugunsten von Baricitinib. Diese Unterschiede waren statistisch nicht signifikant.

In der Subpopulation A3 verkürzte sich die Dauer der Morgensteifigkeit im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-64).

European Quality of Life-5 Dimensions (EQ-5D) VAS*Veränderung der EQ-5D VAS*

In der mITT Population verbesserte sich der allgemeine Gesundheitszustand von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung des allgemeinen Gesundheitszustandes.

Baricitinib versus Adalimumab – mITT

Veränderung der EQ-5D VAS von Baseline bis Woche 24:

LSMD (95% KI): 5,6 (2,4; 8,8); $p < 0,001$

Veränderung der EQ-5D VAS von Baseline bis Woche 52:

LSMD (95% KI): 7,4 (4,2; 10,7); $p < 0,001$

In der mITT Population nahm die Verbesserung der EQ-5D VAS im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-67).

In den Subpopulationen A2 und A3 verbesserte sich der allgemeine Gesundheitszustand von Baseline bis Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung des allgemeinen Gesundheitszustandes.

Baricitinib versus Adalimumab – Subpopulation A2

Veränderung der EQ-5D VAS von Baseline bis Woche 24:

LSMD (95% KI): 6,4 (1,7; 11,2); p = 0,008

Veränderung der EQ-5D VAS von Baseline bis Woche 52:

LSMD (95% KI): 8,3 (3,5; 13,1); p < 0,001

Baricitinib versus Adalimumab – Subpopulation A3

Veränderung der EQ-5D VAS von Baseline bis Woche 24:

LSMD (95% KI): 5,4 (0,2; 10,6); p = 0,042

Veränderung der EQ-5D VAS von Baseline bis Woche 52:

LSMD (95% KI): 5,6 (0,1; 11,2); p = 0,046

In den Subpopulationen A2 und A3 nahm die Verbesserung der EQ-5D VAS im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-67).

Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung des allgemeinen Gesundheitszustandes als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung des allgemeinen Gesundheitszustandes (Tabelle 4-68).

Baricitinib versus Adalimumab – mITT

Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm von Baseline bis Woche 24:

64,4% vs. 56,1%; RR (95% KI): 1,15 (1,02; 1,29); p = 0,020

Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm von Baseline bis Woche 52:

66,5% vs. 56,7%; RR (95% KI): 1,16 (1,04; 1,30); p = 0,009

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung des allgemeinen Gesundheitszustandes als unter Adalimumab. Für die Subpopulation A2 war dieser Unterschied von Baseline bis Woche 24, jedoch nicht bis Woche 52, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung des allgemeinen Gesundheitszustandes. Für die Subpopulation A3 waren die Unterschiede zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-68).

Baricitinib versus Adalimumab – Subpopulation A2

Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm von Baseline bis Woche 24:

63,4% vs. 51,0%; RR (95% KI): 1,25 (1,04; 1,5); p = 0,020

Schwerste Müdigkeit (Worst Tiredness)*Veränderung der schwersten Müdigkeit*

In der mITT Population verbesserte sich das Symptom Müdigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib im Vergleich zu Adalimumab. In der mITT Population lag ein statistisch signifikanter Unterschied von Baseline bis Woche 52, jedoch nicht bis Woche 24, vor. Dies spricht für eine Überlegenheit von Baricitinib in hinsichtlich der Verbesserung des Symptoms Müdigkeit.

Baricitinib versus Adalimumab – mITT

Veränderung der schwersten Müdigkeit von Baseline bis Woche 52:

LSMD (95% KI): -0,59 (-1,02; -0,16); p = 0,007

In der mITT Population nahm die Verbesserung des Symptoms Müdigkeit im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-71).

In den Subpopulationen A2 und A3 verbesserte sich das Symptom Müdigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Für die Subpopulation A2 war dieser Unterschied zu beiden Zeitpunkten statistisch nicht signifikant. In der Subpopulation A3 lag ein statistisch signifikanter Unterschied von Baseline bis Woche 52, jedoch nicht bis Woche 24, vor. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Verbesserung des Symptoms Müdigkeit.

Baricitinib versus Adalimumab – Subpopulation A3

Veränderung der schwersten Müdigkeit von Baseline bis Woche 52:

LSMD (95% KI): -0,77 (-1,50; -0,04); p = 0,039

In den Subpopulationen A2 und A3 nahm die Verbesserung des Symptoms Müdigkeit im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-71).

Lebensqualität

Short Form (36 item) Health Survey Version 2 akut (SF-36 v2 akut)

Veränderung des SF-36 v2 akut PCS

In der mITT Population verbesserte sich die physische Komponente des allgemeinen Gesundheitszustandes sowohl von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes und der Lebensqualität (Tabelle 4-74).

Baricitinib versus Adalimumab – mITT

Veränderung des SF-36 v2 akut PCS von Baseline bis Woche 24:

LSMD (95% KI): 1,7 (0,7; 2,8); $p = 0,001$

Veränderung des SF-36 v2 akut PCS von Baseline bis Woche 52:

LSMD (95% KI): 1,7 (0,6; 2,9); $p = 0,004$

In den Subpopulationen A2 und A3 verbesserte sich die physische Komponente des allgemeinen Gesundheitszustandes sowohl von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. In der Subpopulation A2 war dieser Unterschied von Baseline bis Woche 52, jedoch bis Woche 24, statistisch signifikant. In der Subpopulation A3 lag ein statistisch signifikanter Unterschied von Baseline bis Woche 24, jedoch nicht bis Woche 52, vor. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich der Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes und der Lebensqualität (Tabelle 4-74).

Baricitinib versus Adalimumab – Subpopulation A2

Veränderung des SF-36 v2 akut PCS von Baseline bis Woche 52:

LSMD (95% KI): 1,8 (0,1; 3,5); $p = 0,033$

Baricitinib versus Adalimumab – Subpopulation A3

Veränderung des SF-36 v2 akut PCS von Baseline bis Woche 24:

LSMD (95% KI): 1,8 (0,0; 3,6); $p = 0,046$

Anteil der Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine Verbesserung der physischen Komponente des allgemeine Gesundheitszustandes (Tabelle 4-75).

Baricitinib versus Adalimumab – mITT

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5

von Baseline bis Woche 24:

65,1% vs. 56,7%; RR (95% KI): 1,15 (1,02; 1,28); p = 0,018

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5

von Baseline bis Woche 52:

60,0% vs. 51,8%; RR (95% KI): 1,14 (1,00; 1,29); p = 0,047

In den Subpopulationen A2 und A3 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der physischen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. In beiden Subpopulationen war dieser Unterschied zu keinem Zeitpunkt statistisch signifikant (Tabelle 4-75).

Veränderung des SF-36 v2 akut MCS

In der mITT Population verbesserte sich die mentale Komponente des allgemeinen Gesundheitszustandes von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant.

In der mITT Population nahm die Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-76).

In der Subpopulationen A2 verbesserte sich die mentale Komponente des allgemeinen Gesundheitszustandes von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Der Unterschied war für diese Population zu beiden Zeitpunkten statistisch nicht signifikant. In der Subpopulation A3 verbesserte sich die mentale Komponente des allgemeinen Gesundheitszustandes von Baseline bis Woche 24 zugunsten von Adalimumab. Dieser Unterschied war statistisch nicht signifikant. Zu Woche 52 gab es in der Subpopulation A3 keine Unterschiede zwischen den Behandlungsarmen.

In der Subpopulation A2 nahm die Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes im Verlauf der Studie, d. h. von Woche 24 bis Woche 52, unter Baricitinib im Vergleich zu Adalimumab weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-76).

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5

In der mITT Population erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-77).

In der Subpopulation A2 erreichte ein höherer Anteil der Patienten unter Baricitinib von Baseline bis Woche 24 und Woche 52 eine klinisch relevante Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Dieser Unterschied war zu Woche 24, jedoch nicht zu Woche 52, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib in Hinblick auf eine klinisch relevante Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes. In der Subpopulation A3 erreichte ein höherer Anteil der Patienten unter Adalimumab von Baseline bis Woche 24 und bis Woche 52 eine klinisch relevante Verbesserung der mentalen Komponente des allgemeinen Gesundheitszustandes als unter Adalimumab. Dieser Unterschied war zu keinem Zeitpunkt statistisch signifikant (Tabelle 4-77).

Baricitinib versus Adalimumab – Subpopulation A2

Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5

von Baseline bis Woche 24:

41,6% vs. 30,1%; RR (95% KI): 1,38 (1,05; 1,82); p = 0,021

*Veränderung der SF-36 v2 akut Domänen**SF-36 v2 akut Domäne – Körperliche Funktionsfähigkeit*

In der mITT Population verbesserte sich die Domäne körperliche Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Funktionsfähigkeit und somit der Lebensqualität (Tabelle 4-78).

Baricitinib versus Adalimumab – mITT

Veränderung der SF-36 v2 akut Domäne – Körperliche Funktionsfähigkeit von Baseline bis Woche 24:

LSMD (95% KI): 1,7 (0,5; 3,0); p = 0,006

Veränderung der SF-36 v2 akut Domäne – Körperliche Funktionsfähigkeit von Baseline bis Woche 52:

LSMD (95% KI): 1,4 (0,1; 2,7); p = 0,029

In den Subpopulationen A2 und A3 verbesserte sich die Domäne körperliche Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu keinem Zeitpunkt signifikant (Tabelle 4-78).

SF-36 v2 akut Domäne – Körperliche Rollenfunktion

In der mITT Population verbesserte sich die Domäne körperliche Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Rollenfunktion und somit der Lebensqualität.

Baricitinib versus Adalimumab – mITT

Veränderung der SF-36 v2 akut Domäne – Körperliche Rollenfunktion von Baseline bis Woche 24:

LSMD (95% KI): 1,3 (0,2; 2,5); $p = 0,025$

Veränderung der SF-36 v2 akut Domäne – Körperliche Rollenfunktion von Baseline bis Woche 52:

LSMD (95% KI): 1,9 (0,7; 3,1); $p = 0,002$

In der mITT Population nahm die Verbesserung der körperlichen Rollenfunktion im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-79).

In den Subpopulationen A2 und A3 verbesserte sich die Domäne körperliche Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in der Subpopulation A2 von Baseline bis Woche 52, jedoch nicht bis Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Rollenfunktion und somit der Lebensqualität. In der Subpopulation A3 war dieser Unterschied zu keinem Zeitpunkt statistisch signifikant.

Baricitinib versus Adalimumab – Subpopulation A2

Veränderung der SF-36 v2 akut Domäne – Körperliche Rollenfunktion von Baseline bis Woche 52:

LSMD (95% KI): 2,3 (0,6; 4,1); $p = 0,010$

In den Subpopulationen A2 und A3 nahm die Verbesserung der körperlichen Rollenfunktion im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-79).

SF-36 v2 akut Domäne – Körperliche Schmerzen

In der mITT Population verbesserte sich die Domäne körperlichen Schmerzen von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war für beide Zeitpunkte statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Schmerzen und somit der Lebensqualität (Tabelle 4-80).

Baricitinib versus Adalimumab – mITT

Veränderung der SF-36 v2 akut Domäne – Körperliche Schmerzen von Baseline bis Woche 24:

LSMD (95% KI): 1,8 (0,6; 3,0); $p = 0,002$

Veränderung der SF-36 v2 akut Domäne – Körperliche Schmerzen von Baseline bis Woche 52:

LSMD (95%-KI): 1,5 (0,2; 2,8); $p = 0,028$

In den Subpopulationen A2 und A3 verbesserte sich die Domäne körperlichen Schmerzen von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in der Subpopulation A2 von Baseline bis Woche 52, jedoch nicht bis Woche 24, und in der Subpopulation A3 von Baseline bis Woche 24, jedoch nicht bis Woche 52, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Schmerzen und somit der Lebensqualität.

Baricitinib versus Adalimumab – Subpopulation A2

Veränderung der SF-36 v2 akut Domäne – Körperliche Schmerzen von Baseline bis Woche 52:

LSMD (95% KI): 2,1 (0,2; 4,0); $p = 0,032$

Baricitinib versus Adalimumab – Subpopulation A3

Veränderung der SF-36 v2 akut Domäne – Körperliche Schmerzen von Baseline bis Woche 24:

LSMD (95% KI): 2,3 (0,3; 4,2); $p = 0,023$

In der Subpopulation A2 nahm die Verbesserung der körperlichen Schmerzen im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-80).

SF-36 v2 akut Domäne – Allgemeiner Gesundheitsstatus

In der mITT Population verbesserte sich die Domäne allgemeiner Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu Woche 52 statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung des allgemeinen Gesundheitszustandes und somit der Lebensqualität.

Baricitinib versus Adalimumab – mITT

Veränderung der SF-36 v2 akut Domäne – Allgemeiner Gesundheitszustand von Baseline bis Woche 52:

LSMD (95% KI): 1,3 (0,2; 2,5); $p = 0,023$

In der mITT Population nahm die Verbesserung des allgemeinen Gesundheitszustandes im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-81).

In den Subpopulationen A2 verbesserte sich die Domäne allgemeiner Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in der Subpopulation A2 zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung des allgemeinen Gesundheitszustandes und somit der Lebensqualität. In der Subpopulation A3 verbesserte sich der allgemeine Gesundheitszustand zu Woche 24 zugunsten von Adalimumab und zu Woche 52 zugunsten von Baricitinib. Dieser Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant.

Baricitinib versus Adalimumab – Subpopulation A2

Veränderung der SF-36 v2 akut Domäne – Allgemeiner Gesundheitszustand von Baseline bis Woche 24:

LSMD (95% KI): 1,7 (0,1; 3,3); $p = 0,037$

Veränderung der SF-36 v2 akut Domäne – Allgemeiner Gesundheitszustand von Baseline bis Woche 52:

LSMD (95% KI): 2,4 (0,7; 4,0); $p = 0,006$

In der Subpopulation A2 nahm die Verbesserung des allgemeinen Gesundheitszustandes im Studienverlauf, d. h. von Woche 24 bis Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-81).

SF-36 v2 akut Domäne – Vitalität

In der mITT Population verbesserte sich die Domäne Vitalität von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu Woche 52, jedoch nicht zu Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der Vitalität und somit der Lebensqualität.

Baricitinib versus Adalimumab – mITT

Veränderung der SF-36 v2 akut Domäne – Vitalität von Baseline bis Woche 52:

LSMD (95% KI): 1,3 (0,0; 2,6); p = 0,045

In der mITT Population nahm die Verbesserung der Vitalität im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-82).

In den Subpopulationen A2 und A3 verbesserte sich die Domäne Vitalität von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in beiden Subpopulationen zu keinem Zeitpunkt statistisch signifikant.

In der Subpopulation A2 nahm die Verbesserung der Vitalität im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-82).

SF-36 v2 akut Domäne – Soziale Funktionsfähigkeit

In der mITT Population verbesserte sich die Domäne soziale Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Funktionsfähigkeit und somit der Lebensqualität (Tabelle 4-83).

Baricitinib versus Adalimumab – mITT

Veränderung der SF-36 v2 akut Domäne – Soziale Funktionsfähigkeit von Baseline bis Woche 24:

LSMD (95% KI): 2,0 (0,7; 3,3); p = 0,002

Veränderung der SF-36 v2 akut Domäne – Soziale Funktionsfähigkeit von Baseline bis Woche 52:

LSMD (95% KI): 1,9 (0,7; 3,2); p = 0,003

In den Subpopulationen A2 und A3 verbesserte sich die Domäne soziale Funktionsfähigkeit von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Für die Subpopulation A2 lag zu beiden Zeitpunkten ein statistisch signifikanter Unterschied vor. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der körperlichen Funktionsfähigkeit und somit der Lebensqualität. Der Unterschied war für die Subpopulation A3 zu beiden Zeitpunkten statistisch nicht signifikant.

Baricitinib versus Adalimumab – Subpopulation A2

Veränderung der SF-36 v2 akut Domäne – Soziale Funktionsfähigkeit von Baseline bis Woche 24:

LSMD (95% KI): 2,3 (0,4; 4,2); p = 0,020

Veränderung der SF-36 v2 akut Domäne – Soziale Funktionsfähigkeit von Baseline bis Woche 52:

LSMD (95% KI): 3,1 (1,1; 5,0); p = 0,002

In der Subpopulation A2 nahm die Verbesserung der sozialen Funktionsfähigkeit im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-83).

SF-36 v2 akut Domäne – Emotionale Rollenfunktion

In der mITT Population verbesserte sich die Domäne emotionale Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu Woche 52, jedoch nicht zu Woche 24, statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der emotionalen Rollenfunktion und somit der Lebensqualität.

Baricitinib versus Adalimumab – mITT

Veränderung der SF-36 v2 akut Domäne – Emotionale Rollenfunktion von Baseline bis Woche 52:

LSMD (95% KI): 1,4 (0,0; 2,7); p = 0,048

In der mITT Population nahm die Verbesserung der emotionalen Rollenfunktion im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-84).

In der Subpopulationen A2 verbesserte sich die Domäne emotionalen Rollenfunktion von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in den Subpopulationen A2 zu beiden Zeitpunkten statistisch signifikant. Dies spricht für eine Überlegenheit von Baricitinib hinsichtlich einer Verbesserung der emotionalen Rollenfunktion und somit der Lebensqualität. In der Subpopulation A3 verbesserte sich die Domäne emotionale Rollenfunktion von Baseline bis Woche 24 zugunsten von Adalimumab und von Baseline bis Woche 52 zugunsten von Baricitinib. Dieser Unterschied war zu keinem Zeitpunkt statistisch signifikant (Tabelle 4-84).

Baricitinib versus Adalimumab – Subpopulation A2

Veränderung der SF-36 v2 akut Domäne – Emotionale Rollenfunktion von Baseline bis Woche 24:

LSMD (95% KI): 2,5 (0,4; 4,6); p = 0,021

Veränderung der SF-36 v2 akut Domäne – Emotionale Rollenfunktion von Baseline bis Woche 52:

LSMD (95% KI): 2,6 (0,6; 4,7); p = 0,012

SF-36 v2 akut Domäne – Mentaler Gesundheitszustand

In der mITT Population verbesserte sich die Domäne mentaler Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war zu keinem Zeitpunkt statistisch signifikant.

In der mITT Population nahm die Verbesserung des mentalen Gesundheitszustandes im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-85).

In den Subpopulationen A2 und A3 verbesserte sich die Domäne mentaler Gesundheitszustand von Baseline bis Woche 24 und Woche 52 zugunsten von Baricitinib gegenüber Adalimumab. Dieser Unterschied war in beiden Subpopulationen zu keinem Zeitpunkt statistisch signifikant

In der Subpopulation A2 nahm die Verbesserung des mentalen Gesundheitszustandes im Studienverlauf, d. h. von Woche 24 zu Woche 52, unter Baricitinib weiter zu. Dies spricht für die anhaltende Wirkung von Baricitinib (Tabelle 4-85).

Sicherheit*Todesfälle*

In der mITT Population traten an Woche 24 und Woche 52 vergleichbar wenige Todesfälle unter Baricitinib (Gesamtzahl Todesfälle bis Woche 52: n = 2) und unter Adalimumab (Gesamtzahl Todesfälle bis Woche 52: 1) auf. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant (Tabelle 4-88).

In der Subpopulation A2 traten an Woche 24 und Woche 52 vergleichbar wenige Todesfälle unter Baricitinib (Gesamtzahl Todesfälle bis Woche 52: n = 2) und unter Adalimumab (Gesamtzahl Todesfälle bis Woche 52: n = 1) auf. Der Unterschied war zu beiden Zeitpunkten statistisch nicht signifikant.

In der Subpopulation A3 traten an Woche 24 und Woche 52 keine Todesfälle unter Baricitinib und unter Adalimumab auf (Tabelle 4-88).

Schwerwiegende Unerwünschte Ereignisse (SUE)

In der mITT Population traten mehr schwerwiegende unerwünschte Ereignisse an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab auf. Diese Unterschiede waren zu beiden Zeitpunkten statistisch signifikant (Tabelle 4-89).

An Woche 24 war die Rate an schwerwiegenden unerwünschten Ereignissen unter Baricitinib auf dem gleichen Niveau wie unter Placebo [9]. Das geringe Auftreten von schwerwiegenden unerwünschten Ereignissen unter Adalimumab im Vergleich zu Placebo ist überraschend angesichts vorangegangener Adalimumab Studien und der bekannten Datenlage [58]. Die aufgetretenen SUE waren gleichmäßig über die SOCs verteilt. In keiner SOC konnte ein statistisch signifikanter Unterschied zwischen der Baricitinib- und der Adalimumab-Behandlungsgruppe beobachtet werden.

Baricitinib versus Adalimumab – mITT

Anteil Patienten mit SUE von Baseline bis Woche 24:

4,72% vs. 1,82%; RR (95% KI): 2,60 (1,069; 6,310); p = 0,033

Anteil Patienten mit SUE von Baseline bis Woche 52:

7,80% vs. 3,94%; RR (95% KI): 1,98 (1,072; 3,660); p = 0,027

In der Subpopulation A2 traten mehr schwerwiegende unerwünschte Ereignisse an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab auf. Diese Unterschiede waren zu beiden Zeitpunkten statistisch signifikant (Tabelle 4-89).

Baricitinib versus Adalimumab – Subpopulation A2

Anteil Patienten mit SUE von Baseline bis Woche 24:

5,76% vs. 1,31%; RR (95% KI): 4,41 (1,016; 19,125); p = 0,034

Anteil Patienten mit SUE von Baseline bis Woche 52:

8,64% vs. 2,61%; RR (95% KI): 3,31 (1,157; 9,446); p = 0,019

In der Subpopulation A3 unterschieden sich die schwerwiegenden unerwünschten Ereignisse in den beiden Behandlungsarmen an Woche 24 und Woche 52 nicht statistisch signifikant voneinander (Tabelle 4-89).

Unerwünschte Ereignisse (UE)

In der mITT Population unterschied sich der Anteil der Patienten mit unerwünschten Ereignissen an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-90).

In den Subpopulationen A2 und A3 unterschied sich der Anteil der Patienten mit unerwünschten Ereignissen an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-90).

Schwere UE (ohne ausschließliche Laborereignisse)

Alle in der vorliegenden Nutzenbewertung aufgeführten schweren unerwünschten Ereignisse wurden ohne Laborereignisse ausgewertet, in deren Zusammenhang keine Symptome beschrieben wurden oder dies angenommen werden musste und die somit asymptomatisch waren (Abschnitt 4.2.5.2). Diese schweren UE waren ausschließlich Laborereignisse ohne assoziierte Symptomatik und daher für den Patienten nicht zu spüren, also nicht patientenrelevant. Sollte eine mit diesen Laborereignissen in Zusammenhang stehende Symptomatik aufgetreten sein, dann wurde diese Symptomatik innerhalb der unerwünschten Ereignisse erfasst.

In der mITT Population unterschied sich die Anzahl der Patienten mit mindestens einem schweren unerwünschten Ereignis in den beiden Behandlungsarmen zu Woche 24 und Woche 52 nicht statistisch signifikant voneinander.

In den Subpopulationen A2 und A3 unterschied sich die Anzahl der Patienten mit mindestens einem schweren unerwünschten Ereignis an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-91).

Studienabbruch aufgrund von UE

In der mITT Population unterschieden sich die Studienabbrüche aufgrund von unerwünschten Ereignissen an Woche 24 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander. An Woche 52 traten unter Baricitinib signifikant mehr Studienabbrüche aufgrund von unerwünschten Ereignissen auf als unter Adalimumab (Tabelle 4-92).

Baricitinib versus Adalimumab – mITT

Studienabbrüche aufgrund von UE von Baseline bis Woche 52:

6,98% vs. 3,64%; RR (95% KI): 1,92 (1,009; 3,652); 0,045

In Subpopulation A2 und Subpopulation A3 unterschieden sich die Studienabbrüche aufgrund von unerwünschten Ereignissen an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-92).

Therapieabbruch aufgrund von UE

In der mITT Population unterschieden sich die Therapieabbrüche aufgrund von unerwünschten Ereignissen an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-93).

In den Subpopulationen A2 und A3 unterschieden sich die Therapieabbrüche aufgrund von unerwünschten Ereignissen an Woche 24 und Woche 52 unter Baricitinib im Vergleich zu Adalimumab nicht statistisch signifikant voneinander (Tabelle 4-93).

UE von besonderem Interesse

In der mITT Population konnte für keines der unerwünschten Ereignisse von besonderem Interesse an Woche 24 oder bis Woche 52 unter Baricitinib im Vergleich zu Adalimumab ein statistisch signifikanter Unterschied festgestellt werden (Tabelle 4-94).

In den Subpopulationen A2 und A3 konnte für keines der unerwünschten Ereignisse von besonderem Interesse an Woche 24 oder bis Woche 52 unter Baricitinib im Vergleich zu Adalimumab ein statistisch signifikanter Unterschied festgestellt werden (Tabelle 4-94).

Effektmodifikation in Subgruppen

Bei der Beurteilung der Ergebnisse von Subgruppenanalysen sollten die mit multiplem Testen in Zusammenhang stehende α -Fehler-Kumulierung, die die Zunahme der Wahrscheinlichkeit für Fehler 1. Art (falsch positive Testergebnisse) beschreibt, sowie bei der Interpretation der Resultate kleine Fallzahlen innerhalb der Subgruppen berücksichtigt werden.

Es wurden für 43 (Unter-) Endpunkte bis zu 18 Subgruppenanalysen durchgeführt. Unter der Annahme, dass keine Subgruppeneffekte vorliegen, wären daher zufallsbedingt ca. 38 – 39 Belege ($p < 0,05$) pro Population für eine Effektmodifikation zu erwarten.

In der mITT Population zeigten sich 37 Belege, in der Subpopulation A2 18 Belege sowie in der Subpopulation A3 9 Belege für eine Effektmodifikation.

Die Anzahl der beobachteten Effektmodifikationen liegt somit im oder deutlich unterhalb des Bereiches der zu erwarteten Ergebnisse. Auch sind keine auffälligen Häufungen über die Endpunkte oder Subgruppen hinweg zu beobachten. Auf Grund dessen wird insgesamt nicht von einer Modifikation des Behandlungseffekts ausgegangen (Abschnitt 4.3.1.3.5).

4.3.2 Weitere Unterlagen

4.3.2.1 Indirekte Vergleiche auf Basis randomisierter kontrollierter Studien

Hinweis: Die nachfolgenden Unterabschnitte sind nur dann auszufüllen, wenn indirekte Vergleiche als Nachweis für einen Zusatznutzen herangezogen werden sollen. Das ist dann möglich, wenn keine direkten Vergleichsstudien für das zu bewertende Arzneimittel gegenüber der zweckmäßigen Vergleichstherapie vorliegen oder diese keine ausreichenden Aussagen über den Zusatznutzen zulassen.

4.3.2.1.1 Ergebnis der Informationsbeschaffung – Studien für indirekte Vergleiche

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 (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.). 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 4.3.1.2 und 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 4-160: Matrix der Endpunkte in den eingeschlossenen RCT für indirekte Vergleiche

Studie	<Mortalität>	<Gesundheitsbezogene Lebensqualität>	<Endpunkt>	<Endpunkt>	<Endpunkt>
Nicht zutreffend.					

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 4-161: Zusammenfassung der verfügbaren Vergleiche in den Studien, die für den indirekten Vergleich herangezogen wurden

Anzahl Studien	Studie	Intervention	<Vergleichstherapie 1>	<Vergleichstherapie 2>	<Vergleichstherapie 3>
Nicht zutreffend.					

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 4-162: Operationalisierung von <Endpunkt xxx>

Studie	Operationalisierung
Nicht zutreffend.	

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

Tabelle 4-163: Bewertung des Verzerrungspotenzials für <Endpunkt xxx> in RCT für indirekte Vergleiche

Studie	Verzerrungspotenzial auf Studienebene	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte	Verzerrungspotenzial Endpunkt
Nicht zutreffend.						
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.						

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 4-164: 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)
Nicht zutreffend.	

Nicht zutreffend.

Stellen Sie die Ergebnisse der indirekten Vergleiche in tabellarischer Form dar. Optional können die Ergebnisse zusätzlich auch grafisch illustriert werden. Orientieren Sie sich dabei an der üblichen Darstellung metaanalytischer Ergebnisse. Gliedern Sie die Ergebnisse nach folgenden Punkten:

- *Homogenität der Ergebnisse: 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.*

Machen Sie darüber hinaus Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Nicht zutreffend.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt, für den ein indirekter Vergleich vorgenommen wird, fortlaufend in einem eigenen Abschnitt dar.

4.3.2.1.3.2 Subgruppenanalysen – indirekte Vergleiche aus RCT

Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen auf Basis indirekter Vergleiche aus RCT. Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.2.

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 (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.). 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-A. 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-A.

Tabelle 4-165: Verzerrungsaspekte auf Studienebene – nicht randomisierte vergleichende Interventionsstudien

Studie	Zeitliche Parallelität der Gruppen	Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren	Verblindung		Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte
			Patient	Behandelnde Personen		
Nicht zutreffend.						

Beschreiben Sie zusammenfassend die Bewertungsergebnisse zu Verzerrungsaspekten auf Studienebene.

Nicht zutreffend.

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 4-166: Operationalisierung von <Endpunkt xxx>

Studie	Operationalisierung
Nicht zutreffend.	

Beschreiben Sie die Verzerrungsaspekte für den in diesem Abschnitt beschriebenen Endpunkt mithilfe des Bewertungsbogens in Anhang 4-A. 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-A.

Tabelle 4-167: Verzerrungsaspekte für <Endpunkt xxx> – nicht randomisierte vergleichende Studien

Studie	Verblindung Endpunkterheber	Adäquate Umsetzung des ITT-Prinzips	Ergebnisunabhängige Berichterstattung	Keine sonstigen Aspekte
Nicht zutreffend.				
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.				

Beschreiben Sie zusammenfassend die Bewertungsergebnisse zu Verzerrungsaspekten auf Endpunktebene.

Nicht zutreffend.

Stellen Sie die Ergebnisse der nicht randomisierten vergleichenden Studien gemäß den Anforderungen des TREND- bzw. des STROBE-Statements dar. Machen Sie dabei auch Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Nicht zutreffend.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus nicht randomisierten vergleichenden Studien fortlaufend in einem eigenen Abschnitt dar.

4.3.2.2.3.2 Subgruppenanalysen – nicht randomisierte vergleichende Studien

Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen aus nicht randomisierten vergleichenden Studien. Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.2.

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 (einschließlich tabellarischer Darstellungen, Angabe eines Flussdiagramms etc.). 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 Verzerrungsaspekte.

Ergebnisse nicht randomisierter Studien, die keine kontrollierten Interventionsstudien sind, gelten aufgrund ihres Studiendesigns generell als potenziell hoch verzerrt. Trifft das auf die von Ihnen vorgelegten Studien nicht zu, begründen Sie Ihre Einschätzung.

Strukturieren Sie diesen Abschnitt analog Abschnitt 4.3.1.2 und stellen Sie Informationen analog Abschnitt 4.3.1.2 zur Verfügung.

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 4-168: Operationalisierung von <Endpunkt xxx> – weitere Untersuchungen

Studie	Operationalisierung
Nicht zutreffend.	

Bewerten Sie die Verzerrungsaspekte 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. Machen Sie darüber hinaus Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext.

Nicht zutreffend.

Stellen Sie die in diesem Abschnitt beschriebenen Informationen für jeden weiteren Endpunkt aus weiteren Untersuchungen fortlaufend in einem eigenen Abschnitt dar.

4.3.2.3.3.2 Subgruppenanalysen – weitere Untersuchungen

Beschreiben Sie nachfolgend die Ergebnisse von Subgruppenanalysen aus weiteren Untersuchungen. Berücksichtigen Sie dabei die Anforderungen gemäß Abschnitt 4.3.1.3.2.

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.

Methodische Herangehensweise zur Beurteilung der Aussagekraft der Nachweise

Das Ausmaß des Zusatznutzens soll im Dossier gemäß den Kategorien der AM-NutzenV beschrieben werden [3]. Zur Feststellung des Ausmaßes des Zusatznutzens für die im Dossier zu Baricitinib darzustellenden Endpunkte mittels inferenzstatistischer Schwellenwerte (Hypothesengrenzen) gemäß IQWiG-Methodenpapier sollen die Endpunkte in eine von drei Kategorien für die Qualität der Zielgröße eingeteilt werden: 1: *Gesamtmortalität*, 2: *schwerwiegende Symptome und Nebenwirkungen sowie gesundheitsbezogene Lebensqualität* und 3: *nicht-schwerwiegende Symptome und Nebenwirkungen* [6]. Ein klares methodisches Vorgehen zur Kategorisierung wird vom IQWiG nicht vorgegeben. Jedoch empfiehlt sich eine Kategorisierung anhand der Baselinewerte (und der Veränderung zu Baseline).

Im Dossier ist unter Angabe der Aussagekraft der Nachweise darzulegen, mit welcher Wahrscheinlichkeit und in welchem Ausmaß ein Zusatznutzen des zu bewertenden Arzneimittels im Vergleich zur zVT vorliegt. Diese Angaben sollen sowohl bezogen auf die Anzahl der Patientinnen und Patienten als auch bezogen auf die Größe des Zusatznutzens gemacht werden.

Das Ausmaß des Zusatznutzens soll im Dossier gemäß den Kategorien der AM-NutzenV beschrieben werden [3]:

- *erheblicher*,
- *beträchtlicher*,
- *geringer*,
- *nicht quantifizierbarer*,
- *kein Zusatznutzen belegt*, oder
- Nutzen des zu bewertenden Arzneimittels *geringer* als Nutzen der zweckmäßigen Vergleichstherapie).

Das IQWiG bewertet die beschriebenen Effekte unter Berücksichtigung ihrer Ergebnissicherheit. In der Bewertung werden die qualitative und die quantitative Ergebnissicherheit der vorgelegten Nachweise sowie die Größe der beobachteten Effekte und deren Konsistenz untersucht und bewertet (Tabelle 4-169).

Tabelle 4-169: Anforderungen an die Aussage zur Beleglage/Aussagesicherheit für verschiedene Evidenzsituationen beim Vorliegen von Studien derselben qualitativen Ergebnissicherheit

		Anzahl Studien				
		1 (mit statistisch signifikantem Effekt)	≥ 2			
			homogen	heterogen		
			Meta-Analyse statistisch signifikant	gleichgerichtete Effekte		
deutlich	mäßig	nein				
qualitative Ergebnis- sicherheit	hoch	Hinweis	Beleg	Beleg	Hinweis	-
	mäßig	Anhaltspunkt	Hinweis	Hinweis	Anhaltspunkt	-
	gering	-	Anhaltspunkt	Anhaltspunkt	-	-
Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [6].						

Qualitative Aussage

Prüfung der Wahrscheinlichkeit für das Vorliegen eines Effektes auf Endpunktebene anhand der Kriterien zur Ableitung von Aussagen zur Beleglage. Je nach Evidenzlage wird die Wahrscheinlichkeit demnach als:

- *Anhaltspunkt,*
- *Hinweis oder*
- *Beleg*

eingestuft.

Quantitative Aussage

Feststellung des Ausmaßes der Effektstärke auf Endpunktebene in:

- *erheblich,*
- *beträchtlich,*
- *gering,*
- *nicht quantifizierbar.*

Gesamtaussage zum Zusatznutzen

Unter Berücksichtigung aller Endpunkte unter Würdigung der Wahrscheinlichkeit und des Ausmaßes des Zusatznutzens auf Endpunktebene im Rahmen einer Gesamtschau und entsprechend der vorgegebenen Kategorien:

- *erheblicher,*
- *beträchtlicher,*
- *geringer,*
- *nicht quantifizierbarer Zusatznutzen,*
- *kein Zusatznutzen belegt,*
- *geringer als Nutzen der zweckmäßigen Vergleichstherapie.*

Zur Feststellung des Ausmaßes auf Endpunktebene sind die Qualität der Zielgröße und die Effektstärke maßgeblich. Das grundsätzliche Konzept sieht vor, für relative Effektmaße Schwellenwerte für Konfidenzintervalle in Abhängigkeit von anzustrebenden Effekten abzuleiten, die wiederum von der Qualität der Zielgrößen und den Ausmaßkategorien abhängen. Vom Fall eines quantifizierbaren Effekts ausgehend, richtet sich das weitere Vorgehen nach der Skala der Zielgröße. Es werden folgende Skalen unterschieden:

Binäre Zielgrößen

Zur Feststellung des Ausmaßes des Zusatznutzens bei binären Zielgrößen wird das zweiseitige 95%-Konfidenzintervall für das RR herangezogen. Falls mehrere Studien quantitativ zusammengefasst wurden, findet das meta-analytische Ergebnis für das relative Risiko Anwendung.

Zeit bis Ereignis

Zur Feststellung des Ausmaßes des Effekts bei Zielgrößen „Zeit bis zu einem Ereignis“ wird das zweiseitige 95%-Konfidenzintervall für das Hazard Ratio (HR) herangezogen. Falls mehrere Studien quantitativ zusammengefasst wurden, wird das meta-analytische Ergebnis für das Hazard Ratio herangezogen. Für die Feststellung des Ausmaßes werden dieselben Grenzen wie für das relative Risiko angelegt (Tabelle 4-170).

Stetige Zielgrößen mit jeweils vorliegenden Responderanalysen

Zur Feststellung des Ausmaßes des Zusatznutzens bei stetigen oder quasi-stetigen Zielgrößen werden Responderanalysen herangezogen. Dazu bedarf es eines validierten bzw. etablierten Responsekriteriums bzw. Cut-off-Werts. Anhand der Responderauswertungen (Vierfeldertafeln) werden die relativen Risiken daraus direkt berechnet (Tabelle 4-170).

Sonstige Zielgrößen

Für den Fall sonstiger Zielgrößen, für die auch keine Responderauswertungen mit daraus ableitbaren relativen Risiken vorliegen, ist im Einzelfall zu eruieren, ob relative Risiken geschätzt werden können.

Gesamtaussage zum Ausmaß des Zusatznutzens

Laut IQWiG-Methodenpapier ist für den dritten Schritt der Operationalisierung der Gesamtaussage zum Ausmaß des Zusatznutzens bei gemeinsamer Betrachtung aller Endpunkte eine strenge Formalisierung nicht möglich, da für die hierzu zutreffenden Werturteile gegenwärtig keine ausreichende Abstraktion bekannt ist. Das IQWiG stellt im Rahmen seiner Nutzenbewertung die Aussagen zur Wahrscheinlichkeit und zum Ausmaß der Effekte vergleichend gegen und unterbreitet einen Vorschlag für eine Gesamtaussage.

Je nach Qualität der Zielgröße muss das Konfidenzintervall vollständig unterhalb eines bestimmten Schwellenwertes liegen, um das Ausmaß als *gering*, *beträchtlich* oder *erheblich* anzusehen. Es werden folgende die Kategorien: Gesamtmortalität, schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen sowie gesundheitsbezogene Lebensqualität und nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen für die Qualität der Zielgröße gebildet.

Die Schwellenwerte sind für jede Kategorie separat festgelegt und umso größer, je schwerwiegender das Ereignis ist und umso kleiner, je höher das Ausmaß ist (Tabelle 4-170).

Tabelle 4-170: Inferenzstatistische Schwellenwerte (Hypothesengrenzen) für relative Effektmaße gemäß IQWiG-Methodenpapier

Zielgrößenkategorie				
Ausmaßkategorie		Gesamtmortalität	Schwerwiegende Symptome und Nebenwirkungen, und gesundheitsbezogene Lebensqualität ^a	Nicht-schwerwiegende Symptome und Nebenwirkungen
	Erheblich	0,85	0,75 und Risiko $\geq 5\%$ ^b	nicht besetzt
	Beträchtlich	0,95	0,90	0,80
	Gering	1,00	1,00	0,90

a: Voraussetzung ist wie für alle patientenberichtete Endpunkte die Verwendung eines validierten bzw. etablierten Instruments sowie eines validierten bzw. etablierten Responsekriteriums.

b: Risiko muss für mindestens eine der beiden zu vergleichenden Gruppen mindestens 5% betragen. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.

Quelle: [6].

Wahrscheinlichkeit des Zusatznutzens

Die Wahrscheinlichkeit des Zusatznutzens beschreibt die Aussagesicherheit zum Zusatznutzen. Im Rahmen einer Nutzenbewertung wird vom IQWiG zunächst festgestellt, ob das Vorliegen oder das Fehlen eines Zusatznutzens (bzw. Schadens) belegt ist oder das Vorliegen oder Fehlen eines Zusatznutzens (bzw. Schadens) nicht belegt und daher unklar ist, oder ob ein Zusatznutzen (bzw. Schaden) nachgewiesen wurde. Die Wahrscheinlichkeit des Zusatznutzens wird dabei als: *Beleg*, *Hinweis* oder *Anhaltspunkt* definiert.

Zur Feststellung krankheits- und behandlungsbedingter Veränderungen patientenrelevanter Zielgrößen gemäß SGB V (z. B. in § 35 Abs. 1b SGB V) sind vorrangig patientenrelevante Zielgrößen, wie:

- *Mortalität*,
- *Morbidität (Beschwerden und Komplikationen)* und
- *Gesundheitsbezogene Lebensqualität*

zu berücksichtigen.

Gemäß § 35b Abs. 1 Satz 4 SGB V sollen insbesondere:

- *die Verbesserung des Gesundheitszustandes*,
- *eine Verkürzung der Krankheitsdauer*,
- *eine Verlängerung der Lebensdauer*,
- *eine Verringerung der Nebenwirkungen sowie*
- *eine Verbesserung der Lebensqualität*

angemessen berücksichtigt werden. Allgemeine Nutzen- bzw. Schadenaussagen stützen sich in erster Linie auf Belege bezüglich der höher gewichteten Zielgrößen.

Feststellung des Ausmaßes des Zusatznutzens

Gemäß IQWiG-Methodenpapier werden die Zielgrößen ihrer Bedeutung nach wie folgt gruppiert:

- *Gesamtmortalität,*
- *schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) schwerwiegende (bzw. schwere) Nebenwirkungen,*
- *gesundheitsbezogene Lebensqualität und*
- *nicht-schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) nicht-schwerwiegende (bzw. schwere) Nebenwirkungen.*

Der gesundheitsbezogenen Lebensqualität wird die gleiche Bedeutung beigemessen wie schwerwiegenden (bzw. schweren) Symptomen, Folgekomplikationen und Nebenwirkungen. Die möglichen Ausmaßkategorien für die nicht-schwerwiegenden Zielgrößen bleiben auf „*beträchtlich*“ und „*gering*“ beschränkt (Tabelle 4-171).

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

Tabelle 4-171: Methodik zur Feststellung des Ausmaßes des Zusatznutzens – hierarchisierte Kriterien gemäß AM-NutzenV mit Ergänzungen^a

Ausmaßkategorie	Zielgrößenkategorie			
	Gesamtmortalität	<i>Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen</i>	<i>Gesundheitsbezogene Lebensqualität</i>	<i>Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen</i>
Erheblich nachhaltige und gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte große Verbesserung des therapierelevanten Nutzens	erhebliche Verlängerung der Überlebensdauer	langfristige Freiheit bzw. weitgehende Vermeidung	<i>erhebliche Verbesserung</i>	<i>nicht besetzt</i>
Beträchtlich gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte deutliche Verbesserung des therapierelevanten Nutzens	moderate Verlängerung der Überlebensdauer	Abschwächung bzw. relevante Vermeidung	<i>bedeutsame Verbesserung</i>	bedeutsame Vermeidung
Gering gegenüber der zweckmäßigen Vergleichstherapie bisher nicht erreichte moderate und nicht nur geringfügige Verbesserung des therapierelevanten Nutzens	<i>jegliche Verlängerung der Überlebensdauer</i>	<i>jegliche Verringerung</i>	<i>relevante Verbesserung</i>	relevante Vermeidung
<p>a: Ergänzungen gegenüber AM-NutzenV <i>kursiv</i> gesetzt. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [6].</p>				

Tabelle 4-172: Kategorisierung der im Dossier zu Baricitinib dargestellten Endpunkte in die Zielgrößenkategorien schwerwiegende und nicht-schwerwiegende Symptome (bzw. Folgekomplikationen) und Nebenwirkungen mit Ergänzungen^a

Nutzendimension Endpunkt Rationale zur Kategorisierung	Zielgrößenkategorie			
	Gesamt- mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	Gesundheitsbezogene Lebensqualität	Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
RA-BEAM (JADV)				
Morbidität				
Druckschmerzhafte Gelenke und geschwollene Gelenke	-	✓	-	-
<u>Rationale</u> Der Schweregrad des Endpunkts „Druckschmerzhafte Gelenke und geschwollene Gelenke“ wird anhand der Baselinewerte der Patienten abgeleitet. Die in die Studie eingeschlossenen Patienten haben eine vorherige Therapie mit MTX/cDMARDs durchlaufen. Eine Therapie mit cDMARDs führt zu einer Verringerung der Anzahl druckschmerzhafter und geschwollener Gelenke. Trotzdem handelt es sich bei der eingeschlossenen Studienpopulation um MTX-Therapieversager, in Folge dessen ist die Krankheitsaktivität (CDAI, SDAI, DAS28) zu Baseline hoch. Somit entspricht die bei Baseline vorhandene Anzahl druckschmerzhafter und geschwollener Gelenke der Anzahl an betroffenen Gelenken, die man nach Versagen einer MTX-Therapie und unter hoher Krankheitsaktivität erwarten würde, d. h. sie entspricht einer hohen Krankheitsaktivität unter cDMARD-Therapie. Folglich wird der Endpunkt „druckschmerzhafte und geschwollene Gelenke“ der Endpunktkategorie „Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet. Laut AM-NutzenV gilt jegliche Verbesserung eines Endpunkts dieser Kategorie als ein Zusatznutzen geringen Ausmaßes.				
Baselinewerte				
Anzahl der druckschmerzhaften Gelenke (basierend auf 28 Gelenken)	<i>mITT</i>	<i>Subpopulation A1</i>	<i>Subpopulation A2</i>	<i>Subpopulation A3</i>
Baricitinib, MW (SD)	14,0 (6,6)	Nicht zutreffend.	14,7 (6,7)	13,0 (6,3)
Adalimumab, MW (SD)	13,9 (6,9)	Nicht zutreffend.	14,8 (6,8)	13,0 (6,6)
Anzahl der geschwollenen Gelenke (basierend auf 28 Gelenken)	<i>mITT</i>	<i>Subpopulation A1</i>	<i>Subpopulation A2</i>	<i>Subpopulation A3</i>
Baricitinib, MW (SD)	11,1 (5,0)	Nicht zutreffend.	10,9 (5,0)	11,0 (4,8)
Adalimumab, MW (SD)	11,2 (5,6)	Nicht zutreffend.	11,8 (6,0)	10,6 (5,1)

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

Nutzendimension Endpunkt Rationale zur Kategorisierung	Zielgrößenkategorie			
	Gesamt- mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	Gesundheitsbezogene Lebensqualität	Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
DAS28-hsCRP	-	✓	-	-
<u>Rationale</u> Der Schweregrad dieses Endpunkts wird anhand der Baselinewerte der Patienten abgeleitet. Die Krankheitsaktivität gemessen mittels DAS28 wird gemäß ACR und EULAR in verschiedene Stufen eingeteilt. Bei einem DAS28-Score von < 2,6 liegt eine Remission vor [16]. Des Weiteren erfolgt die Einteilung der Krankheitsaktivität mittels DAS28 in die Stufen gering ($\geq 2,6$ bis < 3,2), mittelschwer ($\geq 3,2$ bis $\leq 5,1$) und hoch ($> 5,1$) [15]. Nach vorangegangener Therapie haben die Patienten immer noch einen hohen DAS28-hsCRP-Wert. Auch sind mehr als 50% der Patienten zu Baseline der DAS28-hsCRP-Kategorie $> 5,1$ zuzuordnen. Die Krankheitsaktivität und Krankheitslast zu Baseline ist demnach hoch. Folglich wird der Endpunkt DAS28-hsCRP der Endpunktkategorie „Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet. Eine statistisch signifikante Veränderung der hohen Krankheitslast entspricht einer bedeutsamen Verbesserung schwerwiegender Symptome und damit einer bedeutsamen Verbesserung des Gesundheitszustands.				
Baselinewerte				
<i>DAS28-hsCRP (mg/l)</i>	<i>mITT</i>	<i>Subpopulation A1</i>	<i>Subpopulation A2</i>	<i>Subpopulation A3</i>
<i>Baricitinib, MW (SD)</i>	5,76 (0,92)	Nicht zutreffend.	5,83 (0,93)	5,61 (0,91)
<i>Adalimumab, MW (SD)</i>	5,76 (0,94)	Nicht zutreffend.	5,88 (0,95)	5,65 (0,88)
<i>Anteil Patienten DAS28-hsCRP Kategorie > 5,1</i>	<i>mITT</i>	<i>Subpopulation A1</i>	<i>Subpopulation A2</i>	<i>Subpopulation A3</i>
<i>Baricitinib, n (%)</i>	367 (75,5)	Nicht zutreffend.	188 (77,7)	117 (68,8)
<i>Adalimumab, n (%)</i>	249 (75,7)	Nicht zutreffend.	120 (78,9)	91 (73,4)

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

Nutzendimension Endpunkt Rationale zur Kategorisierung	Zielgrößenkategorie			
	Gesamt- mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	Gesundheitsbezogene Lebensqualität	Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
SDAI	-	✓	-	-
<u>Rationale</u> Der Schweregrad dieses Endpunkts wird anhand der Baselinewerte der Patienten abgeleitet. Eine Remission liegt gemäß ARC und EULAR bei einem SDAI-Score von $\leq 3,3$ vor [15,20,27,32]. Des Weiteren wird die Krankheitsaktivität in die Stufen gering ($> 3,3$ bis 11), mittelschwer (> 11 bis ≤ 22) und hoch (> 26) eingeteilt [15]. Nach vorangegangener Therapie haben die Patienten immer noch einen hohen SDAI-Wert (> 26). Auch sind mehr als 50% der Patienten zu Baseline der SDAI-Kategorie ≥ 26 zuzuordnen. Die Krankheitsaktivität und Krankheitslast zu Baseline ist demnach hoch. Folglich wird der Endpunkt SDAI der Endpunktkategorie „Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet. Eine statistisch signifikante Veränderung der hohen Krankheitslast entspricht einer bedeutsamen Verbesserung schwerwiegender Symptome und damit einer bedeutsamen Verbesserung des Gesundheitszustands.				
Baselinewerte				
<i>SDAI</i>	<i>mITT</i>	<i>Subpopulation A1</i>	<i>Subpopulation A2</i>	<i>Subpopulation A3</i>
<i>Baricitinib, MW (SD)</i>	40,30 (12,66)	Nicht zutreffend.	41,04 (12,84)	38,30 (12,30)
<i>Adalimumab, MW (SD)</i>	40,12 (13,37)	Nicht zutreffend.	42,02 (13,77)	38,40 (12,05)
<i>Anteil Patienten SDAI ≥ 26</i>	<i>mITT</i>	<i>Subpopulation A1</i>	<i>Subpopulation A2</i>	<i>Subpopulation A3</i>
<i>Baricitinib, n (%)</i>	429 (89,0)	Nicht zutreffend.	218 (90,8)	145 (86,3)
<i>Adalimumab, n (%)</i>	278 (85,3)	Nicht zutreffend.	133 (88,7)	101 (82,1)

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

Nutzendimension Endpunkt Rationale zur Kategorisierung	Zielgrößenkategorie			
	Gesamt- mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	Gesundheitsbezogene Lebensqualität	Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
CDAI	-	✓	-	-
<u>Rationale</u> Der Schweregrad dieses Endpunkts wird anhand der Baselinewerte der Patienten abgeleitet. Eine Remission liegt gemäß ARC und EULAR bei einem CDAI-Score von $\leq 2,8$ vor [15,20,27,32]. Des Weiteren wird die Krankheitsaktivität in die Stufen gering ($> 2,8$ bis 10), mittelschwer (> 10 bis ≤ 22) und hoch (> 22) eingeteilt [15]. Nach vorangegangener Therapie haben die Patienten immer noch einen hohen CDAI-Wert (> 22). Auch sind mehr als 50% der Patienten zu Baseline der CDAI-Kategorie > 22 zuzuordnen. Die Krankheitsaktivität und Krankheitslast zu Baseline ist demnach hoch. Folglich wird der Endpunkt CDAI der Endpunktkategorie „Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet. Eine statistisch signifikante Veränderung der hohen Krankheitslast entspricht einer bedeutsamen Verbesserung schwerwiegender Symptome und damit einer bedeutsamen Verbesserung des Gesundheitszustands.				
Baselinewerte				
<i>CDAI</i>	<i>mITT</i>	<i>Subpopulation A1</i>	<i>Subpopulation A2</i>	<i>Subpopulation A3</i>
<i>Baricitinib, MW (SD)</i>	38,06 (11,99)	Nicht zutreffend.	38,78 (12,20)	36,44 (11,34)
<i>Adalimumab, MW (SD)</i>	37,95 (12,95)	Nicht zutreffend.	39,71 (13,37)	36,28 (11,96)
<i>Anteil Patienten CDAI > 22</i>	<i>mITT</i>	<i>Subpopulation A1</i>	<i>Subpopulation A2</i>	<i>Subpopulation A3</i>
<i>Baricitinib, n (%)</i>	444 (92,1)	Nicht zutreffend.	223 (92,9)	154 (91,7)
<i>Adalimumab, n (%)</i>	296 (90,8)	Nicht zutreffend.	139 (92,7)	109 (88,6)
EULAR Responsekriterien („good Response“) (DAS28-basiert)	-	-	-	✓
<u>Rationale</u> Die EULAR Responsekriterien stellen Veränderungen der Krankheitsaktivität dar und sind nicht an einen absoluten Baselinescore gekoppelt. Da somit eine Kategorisierung an Hand der Baselinewerte nicht vorgenommen werden kann, werden die EULAR Responsekriterien der Kategorie „Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet.				

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

Nutzendimension Endpunkt Rationale zur Kategorisierung	Zielgrößenkategorie			
	Gesamt- mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	Gesundheitsbezogene Lebensqualität	Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
PAP (VAS)	-	✓	-	-
<u>Rationale</u> Schmerz ist eines der Hauptsymptome der RA und ist aus Patientensicht eines der schwerwiegendsten Symptome [95]. Die meisten RA-Patienten bewerten den Schmerz als größtes Problem mit der höchsten Priorität [40,96-101]. Ca. 70% der Patienten würden eine Verbesserung der Behandlung von Schmerzen begrüßen [98]. Gemäß ACR <i>Pain Management Task Force</i> ist Schmerz wahrscheinlich der wichtigste durch Patienten berichtete Endpunkt in der Rheumatologie, der eventuell mehr zur Beeinträchtigung der RA-Patienten beiträgt als strukturelle Gelenkschäden [102-103]. Die QUEST-RA Studie [104], die 7.028 Patienten untersuchte, zeigte dass Schmerz auch der entscheidende Faktor bezüglich der globalen Einschätzung der Krankheitsaktivität durch den Patienten ist und ferner der wichtigste prädiktive Faktor der psychosozialen Gesundheit der RA-Patienten [105]. Das individuelle Schmerzempfinden beeinträchtigt das tägliche Wohlempfinden und hat somit einen erheblichen Einfluss auf die Lebensqualität der Betroffenen [41-44]. Durch den sehr hohen Stellenwert, den der Endpunkt Schmerz in dieser Indikation hat, wird er der Endpunktkategorie „Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet.				
PtGADA (VAS)	-	-	-	✓
<u>Rationale</u> Für den Endpunkt <i>Patient Assessment of Global Disease Activity</i> konnte in der Literatur keine Einteilung in Schweregrade identifiziert werden. Somit wird der Endpunkt der Kategorie „Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet.				
HAQ-DI	-	✓	-	-
<u>Rationale</u> Unter der Leitung der EULAR wurde der Patienten-abgeleitete RAID-Score entwickelt [95]. Dieser Score beinhaltet die sieben am wichtigsten durch die Patienten eingestuften Domänen inklusive ihrer Gewichtung mit der funktionellen Beeinträchtigung (16%) an zweiter Stelle [95]. Der Schweregrad dieses Endpunkts wird anhand der Baselinewerte der Patienten abgeleitet. Die Krankheitsaktivität oder Symptomatik wird in die Stufen sehr schwer: HAQ $\geq 1,68$; schwer: HAQ $\geq 1,37$; mittelschwer: HAQ $\leq 1,01$ und mild: HAQ $\leq 0,43$ unterteilt [162]. Nach vorangegangener Therapie haben die Patienten immer noch einen hohen HAQ-DI-Wert (Baseline Mean: 1,56; Median: 1,63 (über alle Behandlungsarme)). Somit liegen über 50% der Patienten in einem Bereich der von Patienten als mindestens schwer wahrgenommen wird ($> 1,37$). Die Krankheitsaktivität und Krankheitslast zu Baseline ist demnach hoch. Folglich wird der Endpunkt HAQ-DI der Endpunktkategorie „Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet. Eine statistisch signifikante Veränderung der hohen Krankheitslast entspricht einer bedeutsamen Verbesserung schwerwiegender Symptome und damit einer bedeutsamen Verbesserung des Gesundheitszustands.				

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

Nutzendimension Endpunkt Rationale zur Kategorisierung	Zielgrößenkategorie			
	Gesamt- mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	Gesundheitsbezogene Lebensqualität	Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
Baselinewerte				
<i>HAQ-DI</i>	<i>mITT</i>	<i>Subpopulation A1</i>	<i>Subpopulation A2</i>	<i>Subpopulation A3</i>
<i>Baricitinib, MW (SD)</i>	1,57 (0,68)	Nicht zutreffend.	1,68 (0,71)	1,41 (0,64)
<i>Adalimumab, MW (SD)</i>	1,59 (0,70)	Nicht zutreffend.	1,63 (0,69)	1,60 (0,72)
Dauer der Morgensteifigkeit (min)	-	✓	-	-
<p><u>Rationale</u></p> <p>Bei der Morgensteifigkeit handelt es sich um ein typisches Begleitsymptom der RA und die Dauer der Morgensteifigkeit korreliert mit der körperlichen Einschränkung des Patienten und dessen Lebensqualität [45]. Morgensteifigkeit beschreibt eine Einschränkung der Beweglichkeit der Gelenke, die direkt nach dem Aufwachen von den Patienten wahrgenommen wird. Sie führt zu einer starken Einschränkung der üblichen morgendlichen Tätigkeiten, wie beispielsweise dem Aufstehen und der Körperpflege. Aus Patientensicht ist die Dauer der Morgensteifigkeit ein extrem beeinträchtigendes Krankheitssymptom, welches starken Einfluss auf die Lebensqualität, den Funktionsstatus und die Arbeitsfähigkeit ausübt und welches ein wichtiges und klinisch relevantes Element der Krankheitsaktivität darstellt [117-119]. RA-Patienten beschreiben die Dauer der Morgensteifigkeit als wichtige Konsequenz der Erkrankung und sehen sie als relevantes Symptom, welches sich von anderen Symptomen der RA, wie Schmerz, unterscheidet. Viele der Patienten beschreiben, dass die Steifigkeit der Gelenke von Tag zu Tag in ihrer Dauer und Intensität variiert und einen großen Einfluss auf die Lebensqualität haben kann. Speziell die Mobilität, Balance, Geschicklichkeit und der Bewegungsradius seien negativ durch die RA-bedingte Steifigkeit der Gelenke beeinflussbar [120-122]. Eine Umfrage unter 518 Rheumatologen und mit 750 RA-Patienten in elf europäischen Ländern zeigte, dass 92% der Patienten die Morgensteifigkeit als mittelschwer bis schwer beeinträchtigend beschreiben und 82% der Patienten angeben, dass sie einen erheblichen negativen Einfluss auf ihre Lebensqualität ausübt [123]. 95% der befragten Rheumatologen stimmten diesem Ergebnis zu [123]. Der Schweregrad dieses Endpunkts wird anhand der Baselinewerte der Patienten abgeleitet. Khan et al., 2009 [124] nehmen für den patientenberichteten Endpunkt Morgensteifigkeit (Dauer) – „Morning Joint Stiffness (MJS)“ – eine Einteilung in: keine: 0 Minuten, mild: 1 – 30 Minuten, mittelschwer: 31 – 60 Minuten und schwer: > 60 Minuten vor. Nach vorangegangener Therapie haben die Patienten immer noch eine lange Dauer der Morgensteifigkeit (> 60 min). Die Krankheitsaktivität und Krankheitslast zu Baseline ist demnach hoch. Folglich wird der Endpunkt „Dauer der Morgensteifigkeit“ der Endpunktkategorie „Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet. Laut AM-NutzenV gilt jegliche Verbesserung eines Endpunkts dieser Kategorie als ein Zusatznutzen geringen Ausmaßes.</p>				
Baselinewerte				
<i>Dauer der Morgensteifigkeit (min)</i>	<i>mITT</i>	<i>Subpopulation A1</i>	<i>Subpopulation A2</i>	<i>Subpopulation A3</i>
<i>Baricitinib, MW (SD)</i>	120,53 (143,52)	Nicht zutreffend.	133,00 (174,60)	115,16 (118,38)
<i>Adalimumab, MW (SD)</i>	121,32 (150,37)	Nicht zutreffend.	126,92 (146,74)	124,04 (157,41)

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

Nutzendimension Endpunkt Rationale zur Kategorisierung	Zielgrößenkategorie			
	Gesamt- mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	Gesundheitsbezogene Lebensqualität	Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
EQ-5D VAS	-	✓	-	-
<u>Rationale</u> Für die Bestimmung der Endpunktkategorie des Endpunkts EQ-5D VAS wurden die Publikationen Huber et al. (2016) und Szende et al. (2004) herangezogen [163-164]. Beide Publikationen berichten EQ-5D VAS Werte der deutschen Durchschnittsbevölkerung. Da der Altersdurchschnitt der RA-BEAM Studienpopulation bei 53,3 Jahren (SD 12,1 Jahre) liegt, wurde der EQ-5D VAS Wert an Baseline mit EQ-5D VAS Werten der deutschen Durchschnittsbevölkerung im Alter von 50-59 Jahren verglichen. Der Durchschnitt (SD) des EQ-5D VAS an Baseline beträgt 50,9 (20,1) in der Baricitinib-Gruppe und 50,3 (21,5) in der Adalimumab-Gruppe. Diese Werte befinden sich deutlich unter denen der deutschen Bevölkerung im Alter zwischen 50 und 59 Jahren: der EQ-5D VAS beträgt hier im Durchschnitt (SD) 79,4 (18,8) (Szende et al.), respektive 81,9 (15,2) (Huber et al.). Somit liegen die EQ-5D VAS Werte der Studienpopulation an Baseline ca. 1,5 bis 2 SD unter den Werten der entsprechenden Alterskohorte, was einen erheblich schlechteren allgemeinen Gesundheitszustand im Vergleich zur deutschen Durchschnittsbevölkerung bedeutet. Vergleicht man die Durchschnittswerte an Baseline der Baricitinib- und Adalimumab-Gruppe mit denen anderer Altersstrata der deutschen Durchschnittsbevölkerung, so lässt sich feststellen, dass diese in der Größenordnung sogar noch unter denen der Alterskohorte mit dem schlechtesten Gesundheitszustand liegen – den über 80jährigen. Hier liegt der EQ-5D VAS bei 58,7 (19,6) (Huber et al.), bzw. 58,9 (27,1). Der Gesundheitszustand der RA-BEAM Studienpopulation liegt somit sogar noch unter einem Gesundheitszustand, wie er bei Personen anzutreffen ist, die 30 Jahre und mehr älter sind. Folglich wird der Endpunkt EQ-5D VAS der Endpunktkategorie „Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen und gesundheitsbezogene Lebensqualität“ zugeordnet.				
Schwerste Müdigkeit (Worst Tiredness)	-	-	-	✓
<u>Rationale</u> Für den Endpunkt Schwerste Müdigkeit (Worst Tiredness) konnte in der Literatur keine Einteilung in Schweregrade identifiziert werden. Somit wird der Endpunkt der Endpunktkategorie „Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet.				
Lebensqualität				
SF-36 v2 akut	-	-	✓	-
<u>Rationale</u> Der Endpunkte SF-36 v2 akut gehört der Endpunktkategorie „Gesundheitsbezogene Lebensqualität“ an.				

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

<i>Nutzendimension</i> Endpunkt Rationale zur Kategorisierung	Zielgrößenkategorie			
	<i>Gesamt-</i> <i>mortalität</i>	<i>Schwerwiegende</i> <i>(bzw. schwere) Symptome</i> <i>(bzw. Folgekomplikationen)</i> <i>und Nebenwirkungen</i>	<i>Gesundheitsbezogene</i> <i>Lebensqualität</i>	<i>Nicht-schwerwiegende</i> <i>(bzw. nicht schwere)</i> <i>Symptome</i> <i>(bzw. Folgekomplikationen)</i> <i>und Nebenwirkungen</i>
Sicherheit				
Todesfälle	✓	-	-	-
SUE	-	✓	-	-
<u>Rationale</u> Per Definition schwerwiegend.				
UE	-	-	-	✓
<u>Rationale</u> Die Gesamtrate der UE wird der Endpunktkategorie „Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet.				
Schwere UE (ohne ausschließliche Laborereignisse)	-	✓	-	-
<u>Rationale</u> Per Definition schwerwiegend.				
Studienabbruch aufgrund von UE	-	-	-	✓
<u>Rationale</u> Eine Einschätzung, um welche Endpunktkategorie es sich bei den Studienabbrüchen aufgrund von UE handelt, hängt von der Schwere der UE ab, die zum Abbruch geführt haben. Auf die Gesamtpopulation bezogen lag der Anteil der SUE an den Studienabbrüchen aufgrund von UE bei 47,2% für den Baricitinib-Arm und bei 15,4% für den Adalimumab-Arm. Somit liegt der Anteil der schwerwiegenden UE an den UE, die zum Studienabbruch führten, bei unter 50%. Weiterhin lag, bezogen auf die Gesamtpopulation, der Anteil der schweren (<i>severe</i>) UE an den Studienabbrüchen aufgrund von UE bei 33,3% für den Baricitinib-Arm und bei 38,5% für den Adalimumab-Arm. Somit liegt auch der Anteil der schweren (<i>severe</i>) UE an den UE, die zum Studienabbruch führten, bei unter 50%. Der Endpunkt Studienabbruch aufgrund von UE wird somit der Endpunktkategorie „Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet.				

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

Nutzendimension Endpunkt Rationale zur Kategorisierung	Zielgrößenkategorie			
	Gesamt- mortalität	Schwerwiegende (bzw. schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen	Gesundheitsbezogene Lebensqualität	Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen
Therapieabbruch aufgrund von UE	-	-	-	✓
<u>Rationale</u> Eine Einschätzung, um welche Endpunktkategorie es sich bei den Therapieabbrüchen aufgrund von UE handelt, hängt von der Schwere der UE ab, die zum Abbruch geführt haben. Auf die Gesamtpopulation bezogen lag der Anteil der SUE an den Therapieabbrüchen aufgrund von UE bei 47,2% für den Baricitinib-Arm und bei 20% für den Adalimumab-Arm. Somit liegt der Anteil der schwerwiegenden UE an den UE, die zum Therapieabbruch führten, bei unter 50%. Weiterhin lag, bezogen auf die Gesamtpopulation, der Anteil der schweren (<i>severe</i>) UE an den Therapieabbrüchen aufgrund von UE bei 33,3% für den Baricitinib-Arm und bei 33,3% für den Adalimumab-Arm. Somit liegt auch der Anteil der schweren (<i>severe</i>) UE an den UE, die zum Therapieabbruch führten, bei unter 50%. Der Endpunkt Therapieabbruch aufgrund von UE wird somit der Endpunktkategorie „Nicht-schwerwiegende (bzw. nicht schwere) Symptome (bzw. Folgekomplikationen) und Nebenwirkungen“ zugeordnet.				
AESI				
Infektionen	-	abhängig von Schweregrad	-	abhängig von Schweregrad
Schwere Infektionen	-	✓	-	-
Kardiovaskuläre UE	-	abhängig von Schweregrad	-	abhängig von Schweregrad
Malignität	-	abhängig von Schweregrad	-	abhängig von Schweregrad
Depressionen: QIDS-SR16	-	abhängig von Schweregrad	-	abhängig von Schweregrad
Allergische Reaktionen/Hypersensitivität^b	-	abhängig von Schweregrad	-	abhängig von Schweregrad
Gastrointestinale Perforation^c	-	✓	-	-
<u>Rationale</u> Da bei den oben genannten AESI kein Unterschied zwischen den Behandlungsarmen auftrat, wird auf eine Zuordnung zu einer Endpunktkategorie verzichtet (mit Ausnahme der Endpunkte schwere Infektionen und gastrointestinale Perforationen, von denen ausgegangen werden kann, dass sie schwerwiegend sind).				
a: Ergänzungen gegenüber AM-Nutzen ^V <i>kursiv</i> gesetzt. b: Für das UE von besonderem Interesse Allergische Reaktionen/Hypersensitivität trat ein Ereignis auf. c: Für das UE von besonderem Interesse Gastrointestinale Perforation trat kein Ereignis auf. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert. Quelle: [6].				

Studienqualität, Validität der herangezogenen Endpunkte und Evidenzstufe

Für den Nachweis des medizinischen Nutzens und medizinischen Zusatznutzens wurde eine randomisierte kontrollierte direkte Vergleichsstudie herangezogen, die die zweckmäßige Vergleichstherapie im Kontrollarm verwendet.

Die zur Bewertung herangezogene klinische Studie wurde anhand des CONSORT-Statements und der Cochrane-Kriterien zur Bewertung des Verzerrungspotenzials der Ergebnisse (Anhang 4-E und Anhang 4-E) auf die Planungs-, Durchführungs- und Auswertungsqualität überprüft. Für die Bewertung wurden die dazugehörigen Studiendokumente (Studienprotokoll und -bericht) herangezogen und für den Nachweis des medizinischen Nutzens und des medizinischen Zusatznutzens die relevanten Ergebnisse im Dossier dargestellt. Das Verzerrungspotenzial auf Studienebene wurde als *niedrig* bewertet.

Die Methodik der Studie entspricht den internationalen Standards der *Good Clinical Practice*. Es handelt sich somit um eine hochwertige Studie, die zudem von ausreichender Größe und daher grundsätzlich geeignet ist, Belege für einen Zusatznutzen zu liefern. Mit Blick auf die Patientencharakteristika entspricht die Zusammensetzung der Studienpopulation der Patientenpopulation unter Alltagsbedingungen (Abschnitt 4.3.1.2.1). Daher ist die Studie RA-BEAM (JADV) dem Evidenzlevel Ib zuzuordnen.

Für die Bewertung des medizinischen Zusatznutzens wurden die Nutzendimensionen *Mortalität*, *Morbidität*, *Lebensqualität* und *Sicherheit* berücksichtigt. Für die Operationalisierung der Nutzendimensionen wurden nur patientenrelevante Endpunkte verwendet. Die Erhebungsinstrumente der jeweiligen Endpunkte sind validiert (Abschnitt 4.2.5.2).

Insgesamt werden zum Nachweis des Zusatznutzens von Baricitinib folgende patientenrelevante Endpunkte im Dossier herangezogen:

Mortalität

Da Mortalität nicht als primärer oder sekundärer Endpunkt vorgesehen war und nicht im Rahmen einer Überlebenszeitanalyse untersucht wurde, wurde die Anzahl der Todesfälle im Rahmen der schwerwiegenden unerwünschten Ereignisse als Annäherung zur Beurteilung dieses patientenrelevanten Endpunktes ausgewertet (Abschnitt 4.3.1.3.4).

Morbidität

- Druckschmerzhafte Gelenke und geschwollene Gelenke (basierend auf 28 Gelenken)
 - Veränderung der Anzahl der druckschmerzhaften Gelenke
 - Veränderung der Anzahl der geschwollenen Gelenke
- *Disease Activity Score 28 High Sensitivity C-Reactive Protein (DAS28-hsCRP)*
 - Veränderung des DAS28-hsCRP
 - Anteil Patienten mit $\text{DAS28-hsCRP} \leq 3,2$
 - Anteil Patienten mit $\text{DAS28-hsCRP} < 2,6$
- *Simplified Disease Activity Index (SDAI)*
 - Veränderung des SDAI
 - Anteil Patienten mit $\text{SDAI} \leq 3,3$
- *Clinical Disease Activity Index (CDAI)*
 - Veränderung des CDAI
 - Anteil Patienten mit $\text{CDAI} \leq 10$
- EULAR Responsekriterien („*good Response*“) (DAS28-basiert)
 - Anteil Patienten mit „*good Response*“ gemäß EULAR Responsekriterien (DAS28-basiert)
- *Patient Assessment of Pain (PAP) (VAS)*
 - Veränderung des PAP (VAS)
 - Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 10 mm
 - Anteil Patienten mit einer Veränderung des PAP (VAS) ≥ 20 mm
- *Patient's Global Assessment of Disease Activity (PtGADA) (VAS)*
 - Veränderung des PtGADA (VAS)
 - Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 10 mm
 - Anteil Patienten mit einer Verbesserung der PtGADA (VAS) ≥ 20 mm
- *Health Assessment Questionnaire-Disability Index (HAQ-DI)*
 - Veränderung des HAQ-DI
 - Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$
 - Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,3$

- Dauer der Morgensteifigkeit (min)
 - Veränderung der Dauer der Morgensteifigkeit (min)
- *European Quality of Life-5 Dimensions (EQ-5D) VAS*
 - Veränderung der EQ-5D VAS
 - Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm
- Schwerste Müdigkeit (*Worst Tiredness*)
 - Veränderung der schwersten Müdigkeit

Lebensqualität

- *Short Form (36 item) Health Survey Version 2 akut (SF-36 v2 akut)*
 - Veränderung des SF-36 v2 akut PCS
 - Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5
 - Veränderung des SF-36 v2 akut MCS
 - Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS ≥ 5
 - Veränderung der SF-36 v2 akut Domänen

Sicherheit

- Todesfälle
- Schwerwiegende Unerwünschte Ereignisse (SUE)
- Unerwünschte Ereignisse (UE)
- Schwere UE (ohne ausschließliche Laborereignisse)
- Studienabbruch aufgrund von UE
- Therapieabbruch aufgrund von UE
- UE von besonderem Interesse
 - Infektionen
 - Schwere Infektionen
 - Kardiovaskuläre UE
 - Malignität
 - Depressionen: QIDS-SR16
 - Allergische Reaktionen/Hypersensitivität
 - Gastrointestinale Perforation

Die zum Nachweis des Zusatznutzens von Baricitinib herangezogenen patientenrelevanten Endpunkte sind nach Prüfung aller möglichen Aspekte bezüglich der Ergebnissicherheit, der Beleglage und der Zielgrößenkategorie wie folgt zu beurteilen (Tabelle 4-173):

Tabelle 4-173: Nutzendimensionen und deren Operationalisierung in Studien zur Bewertung des medizinischen Zusatznutzens von Baricitinib und Ergebnissicherheit, Einstufung der Beleglage und Zielgrößenkategorie

<i>Nutzendimension Endpunkt</i>	<i>Ergebnissicherheit hoch/mäßig/gering</i>	<i>Einstufung Beleglage Beleg/Hinweis/Anhaltspunkt</i>	<i>Zielgrößenkategorie</i>
RA-BEAM (JADV)			
<i>Mortalität^a</i>	-	-	<i>Gesamtmortalität</i>
<i>Morbidität</i>			
Druckschmerzhafte Gelenke und geschwollene Gelenke	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>
DAS28-hsCRP	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>
SDAI	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>
CDAI	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>
EULAR Responsekriterien („good Response“) (DAS28-basiert)	<i>hoch</i>	<i>Hinweis</i>	<i>Nicht-schwerwiegende Symptome und Nebenwirkungen</i>
PAP (VAS)	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>
PtGADA (VAS)	<i>hoch</i>	<i>Hinweis</i>	<i>Nicht-schwerwiegende Symptome und Nebenwirkungen</i>
HAQ-DI	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>
Dauer der Morgensteifigkeit (min)	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>
EQ-5D VAS	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>
Schwerste Müdigkeit	<i>hoch</i>	<i>Hinweis</i>	<i>Nicht-schwerwiegende Symptome und Nebenwirkungen</i>

<i>Nutzendimension Endpunkt</i>	<i>Ergebnissicherheit hoch/mäßig/gering</i>	<i>Einstufung Beleglage Beleg/Hinweis/Anhaltspunkt</i>	<i>Zielgrößenkategorie</i>
RA-BEAM (JADV)			
<i>Lebensqualität</i>			
SF-36 v2 akut	<i>hoch</i>	<i>Hinweis</i>	<i>Gesundheitsbezogene Lebensqualität</i>
<i>Sicherheit</i>			
Todesfälle	<i>hoch</i>	<i>Hinweis</i>	<i>Gesamtmortalität</i>
SUE	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>
UE	<i>hoch</i>	<i>Hinweis</i>	<i>Nicht- schwerwiegende Symptome und Nebenwirkungen</i>
Schwere UE (ohne ausschließliche Laborereignisse)	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>
Studienabbruch aufgrund von UE	<i>hoch</i>	<i>Hinweis</i>	<i>Nicht- schwerwiegende Symptome und Nebenwirkungen</i>
Therapieabbruch aufgrund von UE	<i>hoch</i>	<i>Hinweis</i>	<i>Nicht- schwerwiegende Symptome und Nebenwirkungen</i>
AESI			
Infektionen	<i>hoch</i>	<i>Hinweis</i>	<i>abhängig von Schweregrad</i>
Schwere Infektionen	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>
Kardiovaskuläre Ereignisse	<i>hoch</i>	<i>Hinweis</i>	<i>abhängig von Schweregrad</i>
Malignität	<i>hoch</i>	<i>Hinweis</i>	<i>abhängig von Schweregrad</i>
Depressionen: QIDS-SR16	<i>hoch</i>	<i>Hinweis</i>	<i>abhängig von Schweregrad</i>
Allergische Reaktionen/Hypersensitivität ^b	<i>hoch</i>	<i>Hinweis</i>	<i>abhängig von Schweregrad</i>
Gastrointestinale Perforation ^c	<i>hoch</i>	<i>Hinweis</i>	<i>Schwerwiegende Symptome und Nebenwirkungen</i>

<i>Nutzendimension Endpunkt</i>	<i>Ergebnissicherheit hoch/mäßig/gering</i>	<i>Einstufung Beleglage Beleg/Hinweis/Anhaltspunkt</i>	<i>Zielgrößenkategorie</i>
RA-BEAM (JADV)			
<p>a: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.</p> <p>b: Für das UE von besonderem Interesse Allergische Reaktionen/Hypersensitivität trat ein Ereignis auf.</p> <p>c: Für das UE von besonderem Interesse Gastrointestinale Perforation trat kein Ereignis auf.</p> <p>Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.</p> <p>Quelle: [7-10].</p>			

4.4.2 Beschreibung des Zusatznutzens einschließlich dessen Wahrscheinlichkeit und Ausmaß

Führen Sie die in den Abschnitten 4.3.1 und 4.3.2 beschriebenen Ergebnisse zum Zusatznutzen auf Ebene einzelner Endpunkte zusammen und leiten Sie ab, ob sich aus der Zusammenschau der Ergebnisse zu den einzelnen Endpunkten insgesamt ein Zusatznutzen des zu bewertenden Arzneimittels im Vergleich zur zweckmäßigen Vergleichstherapie ergibt. Berücksichtigen Sie dabei auch die Angaben zur Übertragbarkeit der Studienergebnisse auf den deutschen Versorgungskontext. Liegt ein Zusatznutzen vor, beschreiben Sie, worin der Zusatznutzen besteht.

Stellen Sie die Wahrscheinlichkeit des Zusatznutzens dar, 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.

Aus Sicht der *Lilly Deutschland GmbH* besteht aufgrund der Leitlinienempfehlungen und der Bestimmung der zVT durch den G-BA die Notwendigkeit, für den Nachweis des Zusatznutzens von Baricitinib in den Subpopulationen A2 und A3 die gesamte Studienpopulation (mITT Population) der Zulassungsstudie RA-BEAM (JADV) heranzuziehen.

Bei den Subpopulationen A2 und A3 handelt es sich um Patienten, die gemäß den Leitlinien für eine Zweitlinientherapie mit bDMARDs in Frage kommen. Die Leitlinien geben dabei eine Empfehlung, wann die Eskalation indiziert ist. So ist bei Patienten mit mehreren cDMARDs in der Vortherapie oder einem cDMARD in der Vortherapie und gleichzeitigem Vorliegen ungünstiger Prognosefaktoren eine Zweitlinientherapie mit bDMARDs indiziert. Konsequenterweise hat der G-BA für diese Patienten die gleiche zVT (bDMARDs in Kombination mit MTX (Adalimumab oder Etanercept oder Certolizumab-Pegol oder Golimumab oder Abatacept oder Tocilizumab); ggf. als Monotherapie unter Berücksichtigung des jeweiligen Zulassungsstatus bei MTX-Unverträglichkeit) festgelegt.

Dabei empfiehlt die Leitlinie in jedem Fall ein individuelles Vorgehen, d. h. neben der aktuellen Therapiesituation und dem Vorliegen von ungünstigen Prognosefaktoren ist immer die jeweilige individuelle Patientensituation zu berücksichtigen [4]. Um das Therapieziel Remission oder niedrige Krankheitsaktivität zu erreichen, gibt die Leitlinie somit eine Entscheidungshilfe bei der Fragestellung des Beginns der Zweitlinientherapie.

Die Studie RA-BEAM (JADV) vergleicht Baricitinib mit der vom G-BA bestimmten zVT (in diesem Fall Adalimumab) bei jenen Patienten, für die gemäß der S1-Leitlinie eine erstmalige Behandlung mit Biologika vorgesehen ist [5]. Die Studie umfasst Eskalations-würdige Patienten und ist somit in ihrer Gesamtheit geeignet, die Fragestellung nach einem Zusatznutzen von Baricitinib gegenüber dem *Standard of Care* in der Zweitlinie zu beantworten. Ein Auftrennen der Studie generiert für die Nutzenbewertung keine zusätzliche Information für die Bewertung der Zweitlinientherapie.

Vor diesem Hintergrund wurde für jeden Endpunkt der RA-BEAM (JADV) untersucht, ob der Behandlungseffekt zwischen den verschiedenen Subpopulationen A2 und A3 heterogen, d. h. unterschiedlich ist (Abschnitt 4.3.1.3). Da diese Heterogenitätstests an Woche 52 unauffällig waren und demnach keine Heterogenität zeigten, gehen wir davon aus, dass die über die gesamte Studienpopulation zu beobachtenden Effekte auch auf die verschiedenen Subpopulationen übertragbar sind. Demnach können die Ergebnisse der mITT Population herangezogen werden, um den Zusatznutzen von Baricitinib in der Zweitlinientherapie abzuleiten.

Für die mITT Population konnte für folgende Endpunkte eine signifikante Überlegenheit entweder konsistent zu beiden Zeitpunkten oder zum ausschlaggebenden Zeitpunkt von 52 Wochen gegenüber der zVT Adalimumab festgestellt werden:

- Veränderung der Anzahl der druckschmerzhaften Gelenke(basierend auf 28 Gelenken)
- Veränderung der Anzahl der geschwollenen Gelenke
- Veränderung des DAS28-hsCRP (mBOCF, mLOCF, MMRM)
- Veränderung des SDAI
- Veränderung des CDAI
- Veränderung des PAP (VAS)
- Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm
- Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 20 mm
- Veränderung des PtGADA (VAS)
- Anteil Patienten mit einer Verbesserung des PtGADA (VAS) ≥ 20 mm
- Veränderung des HAQ-DI (mBOCF, mLOCF, MMRM)
- Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$
- Veränderung der Dauer der Morgensteifigkeit (min)
- Veränderung der EQ-5D VAS
- Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm
- Veränderung der schwersten Müdigkeit
- Veränderung des SF-36 v2 akut PCS
- Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5
- Veränderung der SF-36 v2 akut-Domänen
 - Körperliche Funktionsfähigkeit
 - Körperliche Rollenfunktion
 - Körperliche Schmerzen
 - Allgemeiner Gesundheitszustand
 - Vitalität

- Soziale Funktionsfähigkeit
- Emotionale Rollenfunktion

Statistisch signifikante Ergebnisse zuungunsten von Baricitinib konnten für folgende Endpunkte entweder konsistent zu beiden Zeitpunkten oder zum ausschlaggebenden Zeitpunkt von 52 Wochen nachgewiesen werden:

- SUE
- Studienabbruch aufgrund von UE

Bezüglich der anderen patientenrelevanten Endpunkte der Nutzendimension *Sicherheit*, Todesfälle, Unerwünschte Ereignisse, Therapieabbruch aufgrund von UE und UE von besonderem Interesse, erwiesen sich Baricitinib und Adalimumab als vergleichbar.

Gemäß der in Abschnitt 4.4.1 beschriebenen Methodik wurde der Zusatznutzen auf Endpunktebene für statistisch signifikante Ergebnisse für die mITT Population wie folgt abgeleitet (Tabelle 4-174).

Tabelle 4-174: Zusammenfassende Darstellung zum Ausmaß des Zusatznutzens auf Endpunktebene für die mITT Population

Endpunkt	Baricitinib vs. Adalimumab		Ausmaß des Zusatznutzens ^f (Aussagesicherheit)
	Effektschätzer ^a (95% KI); p-Wert		
	Woche 24	Woche 52	
Mortalität^b	-	-	
Morbidität			
Druckschmerzhafte Gelenke und geschwollene Gelenke			
Veränderung der Anzahl der druckschmerzhaften Gelenke (basierend auf 28 Gelenken)	-0,6 (-1,5; 0,2); 0,135	-0,9 (-1,6; -0,1); 0,032 Hedges´g: -0,1 (-0,3; 0,0)	geringer Zusatznutzen (hoch)
Veränderung der Anzahl der geschwollenen Gelenke (basierend auf 28 Gelenken)	-0,5 (-1,2; 0,2); 0,159	-0,9 (-1,5; -0,2); 0,007 Hedges´g: -0,2 (-0,3; -0,0)	geringer Zusatznutzen (hoch)
DAS28-hsCRP			
Veränderung des DAS28-hsCRP (mBOCF)	-0,21 (-0,39; -0,03); 0,022	-0,29 (-0,49; -0,08); 0,007	kein Zusatznutzen (hoch)
Veränderung des DAS28-hsCRP (mLOCF)	-	-0,35 (-0,54; -0,16); < 0,001	kein Zusatznutzen (hoch)
Veränderung des DAS28-hsCRP (MMRM)	-0,19 (-0,36; -0,02); 0,029	-0,23 (-0,41; -0,05); 0,014	kein Zusatznutzen (hoch)
Anteil Patienten mit DAS28-hsCRP ≤ 3,2	1,08 (0,95; 1,24); 0,251	1,14 (1,00; 1,30); 0,059	kein Zusatznutzen (hoch)
Anteil Patienten mit DAS28-hsCRP < 2,6	1,09 (0,90; 1,33); 0,385	1,01 (0,85; 1,20); 0,904	kein Zusatznutzen (hoch)
SDAI			
Veränderung des SDAI	-2,43 (-4,28; -0,59); 0,010	-3,18 (-5,04; -1,32); < 0,001	kein Zusatznutzen (hoch)
Anteil Patienten mit SDAI ≤ 3,3	1,22 (0,86; 1,71); 0,261	1,26 (0,96; 1,67); 0,101	kein Zusatznutzen (hoch)
CDAI			
Veränderung des CDAI	-2,06 (-3,83; -0,29); 0,023	-2,85 (-4,62; -1,09); 0,002	kein Zusatznutzen (hoch)
Anteil Patienten mit CDAI ≤ 10	1,04 (0,91; 1,20); 0,567	1,14 (1,00; 1,29); 0,055	kein Zusatznutzen (hoch)
EULAR Responsekriterien („good Response“, DAS28-basiert)			
Anteil Patienten mit „good Response“ gemäß EULAR Responsekriterien (DAS28-basiert)	1,07 (0,93; 1,23); 0,323	1,15 (1,00; 1,31); 0,051	kein Zusatznutzen (hoch)

Endpunkt	Baricitinib vs. Adalimumab		Ausmaß des Zusatznutzens ^f (Aussagesicherheit)
	Effektschätzer ^a (95% KI); p-Wert		
	Woche 24	Woche 52	
PAP (VAS)			
Veränderung des PAP (VAS)	-4,8 (-8,0; -1,6); 0,003	-5,9 (-9,1; -2,6) < 0,001	nicht quantifizierbarer Zusatznutzen (hoch)
Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 10 mm	1,06 (0,99; 1,14); 0,111	1,13 (1,05; 1,22); 0,001	geringer Zusatznutzen (hoch)
Anteil Patienten mit einer Verbesserung des PAP (VAS) \geq 20 mm	1,09 (0,99; 1,20); 0,090	1,18 (1,07; 1,30); 0,001	geringer Zusatznutzen (hoch)
PtGADA (VAS)			
Veränderung des PtGADA (VAS)	-4,0 (-7,1; -0,8); 0,013	-6,0 (-9,2; -2,8); < 0,001	kein Zusatznutzen (hoch)
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 10 mm	1,04 (0,97; 1,11); 0,298	1,05 (0,98; 1,13); 0,167	kein Zusatznutzen (hoch)
Anteil Patienten mit einer Verbesserung des PtGADA (VAS) \geq 20 mm	1,08 (0,98; 1,18); 0,124	1,10 (1,00; 1,21); 0,043	kein Zusatznutzen (hoch)
HAQ-DI			
Veränderung des HAQ-DI (mBOCF)	-0,11 (-0,18; -0,03); 0,008	-0,09 (-0,17; -0,00); 0,039	nicht quantifizierbarer Zusatznutzen (hoch)
Veränderung des HAQ-DI (mLOCF)	-	-0,11 (-0,19; -0,03); 0,007	nicht quantifizierbarer Zusatznutzen (hoch)
Veränderung des HAQ-DI (MMRM)	-0,12 (-0,2; -0,0); 0,002	-0,08 (-0,2; -0,0); 0,044	nicht quantifizierbarer Zusatznutzen (hoch)
Anteil Patienten mit einer Verbesserung des HAQ-DI \geq 0,22	1,12 (1,02; 1,23); 0,021	1,14 (1,03; 1,27); 0,016	geringer Zusatznutzen (hoch)
Anteil Patienten mit einer Verbesserung des HAQ-DI \geq 0,3	1,11 (1,00; 1,23); 0,060	1,10 (0,98; 1,24); 0,103	kein Zusatznutzen (hoch)
Dauer der Morgensteifigkeit (min)			
Veränderung der Dauer der Morgensteifigkeit (min)	-20,65 (-45,26; 3,95); 0,100	-27,34 (-51,35; -3,32); 0,026 Hedges' g: -0,18 (-0,36; 0,01)	geringer Zusatznutzen (hoch)

Endpunkt	Baricitinib vs. Adalimumab Effektschätzer ^a (95% KI); p-Wert		Ausmaß des Zusatznutzens ^f (Aussagesicherheit)
	Woche 24	Woche 52	
	EQ-5D VAS		
Veränderung der EQ-5D VAS	5,6 (2,4; 8,8); < 0,001	7,4 (4,2; 10,7); < 0,001	<i>nicht quantifizierbarer Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung der EQ-5D VAS \geq 10 mm	1,15 (1,02; 1,29); 0,020	1,16 (1,04; 1,30); 0,009	<i>geringer Zusatznutzen (hoch)</i>
Müdigkeit			
Veränderung der schwersten Müdigkeit	-0,27 (-0,70; 0,16); 0,222	-0,59 (-1,02; -0,16); 0,007 Hedges'g: -0,22 (-0,40; -0,03)	<i>kein Zusatznutzen (hoch)</i>
Lebensqualität			
SF-36 v2 akut			
Veränderung des SF-36 v2 akut PCS	1,7 (0,7; 2,8); 0,001	1,7 (0,6; 2,9); 0,004	<i>nicht quantifizierbarer Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS \geq 5	1,15 (1,02; 1,28); 0,018	1,14 (1,00; 1,29); 0,047	<i>geringer Zusatznutzen (hoch)</i>
Veränderung des SF-36 v2 akut MCS	0,5 (0,9; 1,8); 0,503	0,9 (0,4; 0,2); 0,169	<i>kein Zusatznutzen (hoch)</i>
Anteil Patienten mit einer Verbesserung des SF-36 v2 akut MCS \geq 5	1,13 (0,94; 1,36); 0,202	1,14 (0,93; 1,39); 0,219	<i>kein Zusatznutzen (hoch)</i>
Veränderung der SF-36 v2 akut Domänen			
Körperliche Funktionsfähigkeit	1,7 (0,5; 3,0); 0,006	1,4 (0,1; 2,7); 0,029 Hedges'g: 0,14 (-0,00; 0,28)	<i>kein Zusatznutzen (hoch)</i>
Körperliche Rollenfunktion	1,3 (0,2; 2,5); 0,025	1,9 (0,7; 3,1); 0,002 Hedges'g: 0,20 (0,06; 0,34)	<i>kein Zusatznutzen (hoch)</i>
Körperliche Schmerzen	1,8 (0,6; 3,0); 0,002	1,5 (0,2; 2,8); 0,028 Hedges'g: 0,14 (-0,00; 0,28)	<i>kein Zusatznutzen (hoch)</i>
Allgemeiner Gesundheitszustand	0,7 (0,4; 1,8); 0,209	1,3 (0,2; 2,5); 0,023 Hedges'g: 0,14 (0,00; 0,28)	<i>kein Zusatznutzen (hoch)</i>

Endpunkt	Baricitinib vs. Adalimumab Effektschätzer ^a (95% KI); p-Wert		Ausmaß des Zusatznutzens ^f (Aussagesicherheit)
	Woche 24	Woche 52	
	Vitalität	0,8 (0,4; 2,1); 0,195	1,3 (0,0; 2,6); 0,045 Hedges´g: 0,13 (-0,02; 0,27)
Soziale Funktionsfähigkeit	2,0 (0,7; 3,3); 0,002	1,9 (0,7; 3,2); 0,003 Hedges´g: 0,19 (0,05; 0,33)	kein Zusatznutzen (hoch)
Emotionale Rollenfunktion	0,7 (0,7; 2,0); 0,355	1,4 (0,0; 2,7); 0,048 Hedges´g: 0,12 (-0,02; 0,27)	kein Zusatznutzen (hoch)
Mentaler Gesundheitszustand	0,5 (0,8; 1,9); 0,441	0,7 (0,7; 2,0); 0,327 Hedges´g: 0,06 (-0,08; 0,20)	kein Zusatznutzen (hoch)
Sicherheit			
Todesfälle	5,36 (0,318; 90,546); 0,518	1,34 (0,133; 13,502); 1,000	kein höheres/geringeres Schadenspotenzial (hoch)
SUE	2,60 (1,069; 6,310); 0,033	1,98 (1,072; 3,660); 0,027	höheres Schadenspotenzial: Ausmaß: gering (hoch)
UE	1,05 (0,956; 1,152); 0,302	1,03 (0,954; 1,109); 0,460	kein höheres/geringeres Schadenspotenzial (hoch)
Schwere UE (ohne ausschließliche Laborereignisse) ^c	2,44 (0,915; 6,506); 0,084	1,72 (0,871; 3,416); 0,133	kein höheres/geringeres Schadenspotenzial (hoch)
Studienabbruch aufgrund von UE	2,13 (0,920; 4,928); 0,083	1,92 (1,009; 3,652); 0,045	kein höheres/geringeres Schadenspotenzial (hoch)
Therapieabbruch aufgrund von UE	2,23 (0,967; 5,129); 0,059	1,65 (0,897; 3,018); 0,129	kein höheres/geringeres Schadenspotenzial (hoch)

Endpunkt	Baricitinib vs. Adalimumab		Ausmaß des Zusatznutzens ^f (Aussagesicherheit)
	Effektschätzer ^a (95% KI); p-Wert		
	Woche 24	Woche 52	
UE von besonderem Interesse			
Infektionen	1,08 (0,894; 1,315); 0,409	1,09 (0,934; 1,269); 0,272	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
Schwere Infektionen	1,02 (0,170; 6,093); 1,000	1,36 (0,411; 4,464); 0,771	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
Kardiovaskuläre Ereignisse	5,35 (0,099; 290,613); 1,000	1,34 (0,133; 13,502); 1,000	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
Malignität	5,36 (0,318; 90,546); 0,518	5,37 (0,534; 54,094); 0,277	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
Depressionen: QIDS-SR16	-0,0 (-0,5; 0,5); 0,937	0,2 (-0,4; 0,8); 0,449	<i>kein höheres/geringeres Schadenspotenzial (hoch)</i>
Allergische Reaktionen/Hypersensitivität ^d	-	-	<i>kein höheres/geringeres Schadenspotenzial</i>
Gastrointestinale Perforation ^e	-	-	<i>kein höheres/geringeres Schadenspotenzial</i>

a: Für kontinuierliche Endpunkte wird der LSMD dargestellt. Für binäre Endpunkte wird das RR dargestellt. Für Todesfälle und UE von besonderem Interesse wird das Peto OR dargestellt.

b: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet.

c: Alle in der vorliegenden Nutzenbewertung aufgeführten schweren unerwünschten Ereignisse wurden ohne Laborereignisse ausgewertet, in deren Zusammenhang keine Symptome beschrieben wurden oder dies angenommen werden musste und die somit asymptomatisch waren (Abschnitt 4.2.5.2). Diese schweren UE waren ausschließlich Laborereignisse ohne assoziierte Symptomatik und daher für den Patienten nicht zu spüren, also nicht patientenrelevant. Sollte eine mit diesen Laborereignissen in Zusammenhang stehende Symptomatik aufgetreten sein, dann wurde diese Symptomatik innerhalb der unerwünschten Ereignisse erfasst.

d: Für das UE von besonderem Interesse Allergische Reaktionen/Hypersensitivität trat ein Ereignis auf.

e: Für das UE von besonderem Interesse Gastrointestinale Perforation trat kein Ereignis auf.

f: Sofern eine Responderanalyse für einen Endpunkt einen Zusatznutzen ergeben hat, wurde das Ausmaß des Zusatznutzens für die kontinuierliche Auswertung dieses Endpunkts auf *nicht quantifizierbar* gesetzt.

Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.

Quelle: [59-60].

Schlussfolgerungen zum Zusatznutzen und zum therapeutisch bedeutsamen Zusatznutzen

Bei der der Bewertung zugrunde liegenden Studie handelt es sich um eine randomisierte, aktiv-kontrollierte, klinische Studie der Evidenzstufe Ib. Die Methodik der Studie entspricht den internationalen Standards der *Good Clinical Practice* [61]. Es handelt sich somit um eine hochwertige Studie, die zudem von ausreichender Größe und daher grundsätzlich geeignet ist, *Hinweise* für einen Zusatznutzen zu liefern.

Subpopulation A1

Für Patienten der Subpopulation A1, für Patienten für die **keine ungünstigen Prognosefaktoren**¹ vorliegen und die unzureichend auf eine vorangegangene Behandlung mit **einem** cDMARD, inklusive MTX, ansprachen, hat der G-BA die zVT alternative cDMARDs, sofern geeignet als Mono- oder Kombinationstherapie, bestimmt. Derzeit liegt keine direkte Evidenz zum Vergleich von Baricitinib mit der vom G-BA genannten zVT für diese Subpopulation vor und somit kann **kein Zusatznutzen** gegenüber einer der zweckmäßigen Vergleichstherapien festgestellt werden.

Subpopulationen A2 und Subpopulation A3

Wie oben ausgeführt, werden für die Ableitung des Zusatznutzens für die Subpopulationen A2 und A3 die Ergebnisse der mITT Population herangezogen.

Für die mITT Population wird basierend auf einer direkt vergleichenden Studie RA-BEAM (JADV) (Baricitinib vs. Adalimumab vs. Placebo) ein **Hinweis** auf einen **geringen Zusatznutzen** gegenüber der zweckmäßigen Vergleichstherapie Adalimumab festgestellt (Tabelle 4-175).

Ein Zusatznutzen zeigt sich durch die im Vergleich zur zVT statistisch signifikant besseren Ergebnisse der **Morbiditätsendpunkte Anzahl druckschmerzhafter und geschwollener Gelenke, Schmerz** (Anteil Patienten mit einer Verbesserung des PAP (VAS) ≥ 10 mm und mit einer Verbesserung des PAP (VAS) ≥ 20 mm, **körperliche Funktion** (Anteil Patienten mit einer Verbesserung des HAQ-DI $\geq 0,22$), **Dauer der Morgensteifigkeit** und **allgemeiner Gesundheitszustand** (Anteil Patienten mit einer Verbesserung der EQ-5D VAS ≥ 10 mm) (Tabelle 4-174). Der Zusatznutzen zeigt sich zudem durch die im Vergleich zur zVT statistisch signifikant besseren Ergebnisse des Endpunktes der **Lebensqualität** (Anteil Patienten mit einer Verbesserung des SF-36 v2 akut PCS ≥ 5) (Tabelle 4-174).

¹ Beispiele für ungünstige Prognosefaktoren:

- Nachweis von Autoantikörpern (z. B. Rheumafaktoren, hohe Spiegel von Antikörpern gegen citrullinierte Peptid Antigene)
- Hohe Krankheitsaktivität (nachgewiesen durch DAS- bzw. DAS28-Bewertungssystem, geschwollene Gelenke, Parameter der Akute-Phase-Reaktion wie z. B. C-Reaktives Protein, Erythrozytensedimentationsrate)
- Frühes Auftreten von Gelenkerosionen.

Quelle: [2].

Zusätzlich zeigen sich statistisch signifikante Vorteile für Baricitinib im Vergleich zu Adalimumab für die Endpunkte. Auch wenn diese Endpunkte nach den formalen Kriterien des IQWiG nicht zu der Ableitung eines Zusatznutzens auf Endpunktebene führen, so legen sie doch in ihrer Gesamtsumme einen geringen Zusatznutzen für Baricitinib nahe.

Bezogen auf die *Sicherheitsendpunkte* insgesamt ergab sich für Baricitinib ein höheres Schadenspotenzial durch den Endpunkt **schwerwiegende unerwünschte Ereignisse** (Tabelle 4-174). Hierzu ist zu bemerken, dass die SUE unter Baricitinib an Woche 24 auf Placeboniveau waren. Weiterhin waren die SUE der Adalimumab-Behandlungsgruppe, vor dem Hintergrund der bekannten Datenlage [58], auf unerwartet niedrigem Niveau. Die aufgetretenen SUE waren gleichmäßig über die SOC's verteilt. In keiner SOC konnte ein statistisch signifikanter Unterschied zwischen der Baricitinib- und der Adalimumab-Behandlungsgruppe beobachtet werden.

In der Gesamtschau überwiegen die positiven Effekte der patientenrelevanten Endpunkte der Nutzendimensionen *Morbidität* und *Lebensqualität* den Nachteil der Nutzendimensionen *Sicherheit*. Daher wird der Zusatznutzen von Baricitinib als eine bisher nicht erreichte moderate und nicht nur geringfügige Verbesserung des therapielevanten Nutzens bewertet.

Tabelle 4-175: Gesamtaussage zum medizinischen Zusatznutzen für die mITT Population

Studie RA-BEAM (JADV)	
Nutzendimension Aussage zum Zusatznutzen	Gesamtaussage zum Zusatznutzen
Mortalität^a	Hinweis auf geringen Zusatznutzen
-	
Morbidität	
Hinweis auf einen geringen Zusatznutzen	
Lebensqualität	
Hinweis auf einen geringen Zusatznutzen	
Sicherheit	
Hinweis auf ein höheres Schadenspotenzial: Ausmaß: gering	
a: Mortalität wurde nicht im Rahmen einer Überlebenszeitanalyse untersucht. Die Anzahl der Todesfälle wurde jedoch im Rahmen der schwerwiegenden unerwünschten Ereignisse ausgewertet. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.	

Fazit

Die Rheumatoide Arthritis (RA) ist eine systemische entzündliche Autoimmunerkrankung. Die Krankheit hat unterschiedliche Verlaufsformen und Schweregrade, die sich von einer milden Erkrankung bis hin zu einer schwerwiegenden Erkrankung, assoziiert mit zunehmender Gelenkerstörung, bedeutend eingeschränkter Lebensqualität und reduziertem Überleben, erstrecken können [62-63]. Die stärkste Belastung für Patienten stellen Schmerzen, Fatigue, Morgensteifigkeit sowie Funktionseinschränkungen und Depressionen dar. Die negativen psychologischen Effekte der RA scheinen der Kombination aus Schmerzen und Funktionseinschränkungen zu folgen. Chronische Schmerzen sind eine der Hauptkomponenten der Krankheitslast für RA-Patienten. Die Patienten leiden an stärkeren Schmerzen als die Allgemeinbevölkerung und erfahren ein Schmerzniveau, das mit dem von Schmerzpatienten vergleichbar ist. Höhere Schmerzniveaus korrelieren mit der Funktionseinschränkung sowie mit Depressionen, was sich negativ auf die Lebensqualität der Patienten auswirkt.

In den vergangenen Jahren sind weitere Optionen zur Therapie der RA verfügbar geworden. Trotz einer Vielzahl zugelassener Medikamente für die Behandlung der RA ist die komplette oder anhaltende Krankheitsremission noch immer die Ausnahme. Ein Großteil der Patienten in klinischen Studien erreicht keine klinische Remission [64]. Selbst in Patienten, die eine scheinbar ausreichende Kontrolle der Anzeichen und Symptome der RA durch eine Therapie mit cDMARDs und/oder Biologika erreichen, kann die Erkrankung voranschreiten [65-66]. Bei derzeit auf dem Markt befindlichen Therapeutika zur Behandlung der RA muss zudem mit einem verzögerten Wirkeintritt gerechnet werden.

Auch wenn eine Kontrolle der Schmerzen ein wichtiges Kriterium einer erfolgreichen Therapie darstellt, erfährt ein Großteil der Patienten trotz einer Therapie und der aktuellen Therapieoptionen noch immer bedeutsame Schmerzen [36-40] und eine damit einhergehende Einschränkungen der Lebensqualität [40,67]. Dementsprechend besteht weiterhin ein erheblicher, ungedeckter Bedarf für alternative Therapieoptionen zur Behandlung der RA.

Baricitinib ist ein oral verfügbarer, selektiver Januskinase 1 (JAK1)-/Januskinase 2 (JAK2)-Inhibitor zur Behandlung der mittelschweren bis schweren aktiven rheumatoiden Arthritis, der gezielt in Signalkaskaden eingreift und der nicht auf Signalwege, die für ein funktionierendes Immunsystem wichtig sind, wirkt.

Durch die einmal tägliche orale Gabe, ohne die Notwendigkeit subkutaner oder intravenöser Injektionen/Infusionen, ist Baricitinib zudem patientenfreundlich und kann zu einer verbesserten *Compliance* führen. Baricitinib zeichnet sich durch einen vergleichsweise schnelleren Wirkeintritt aus und führt zu einer schnelleren Verbesserung patientenrelevanter Endpunkte (wie z. B. Schmerz, Anzahl schmerzhafter Gelenke,...). Das Risiko, dass die Erkrankung im kritischen Zeitfenster bedeutend voranschreitet, verringert sich. Die Gefahr des sekundären Wirkverlustes besteht nicht, da keine Antikörper gegenüber Baricitinib gebildet werden.

Damit kann durch Baricitinib als Alternative zu Adalimumab eine „*Verbesserung des Gesundheitszustands und der Lebensqualität*“ für die Patienten erreicht werden [3].

Aufgrund der Strukturgleichheit, v. a. bezüglich der klinischen Parameter, zwischen der Studienpopulation und der Zielpopulation im Anwendungsgebiet A wird davon ausgegangen, dass die in der Studie RA-BEAM (JADV) beobachteten klinischen Effekte auch in der medizinischen Versorgung unter Alltagsbedingungen auftreten und die Studienergebnisse somit auf den deutschen Versorgungskontext übertragbar sind.

4.4.3 Angabe der Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht

Geben Sie auf Basis der in den Abschnitten 4.3.1 und 4.3.2 beschriebenen Ergebnisse und unter Berücksichtigung des in Abschnitt 4.4.2 dargelegten Zusatznutzens sowie dessen Wahrscheinlichkeit und Ausmaß in der nachfolgenden Tabelle an, für welche Patientengruppen ein therapeutisch bedeutsamer Zusatznutzen besteht. Benennen Sie das Ausmaß des Zusatznutzens in Patientengruppen mit therapeutisch bedeutsamem Zusatznutzen. Fügen Sie für jede Patientengruppe mit therapeutisch bedeutsamem Zusatznutzen eine neue Zeile ein.

Tabelle 4-176: Patientengruppen, für die ein therapeutisch bedeutsamer Zusatznutzen besteht, einschließlich Ausmaß des Zusatznutzens

Bezeichnung der Patientengruppen	Ausmaß des Zusatznutzens
Subpopulation A1 Erwachsene Patienten mit mittelschwerer bis schwerer aktiver rheumatoider Arthritis (RA), bei denen keine ungünstigen Prognosefaktoren^a vorliegen und die unzureichend auf eine vorangegangene Behandlung mit einem cDMARD (inklusive MTX) ansprechen.	<i>kein Zusatznutzen</i>
Subpopulation A2 Erwachsene Patienten mit mittelschwerer bis schwerer aktiver rheumatoider Arthritis (RA), mit ungünstigen Prognosefaktoren^a und die unzureichend auf eine vorangegangene Behandlung mit einem cDMARD inklusive MTX, ansprechen.	<i>geringer Zusatznutzen^b</i>
Subpopulation A3 Erwachsene Patienten mit mittelschwerer bis schwerer aktiver rheumatoider Arthritis (RA), die unzureichend auf eine vorangegangene Behandlung mit mehreren cDMARDs, inklusive MTX, ansprechen.	<i>geringer Zusatznutzen^b</i>
a: Beispiele für ungünstige Prognosefaktoren: <ul style="list-style-type: none"> • Nachweis von Autoantikörpern (z. B. Rheumafaktoren, hohe Spiegel von Antikörpern gegen citrullinierte Peptid Antigene) • Hohe Krankheitsaktivität (nachgewiesen durch DAS- bzw. DAS28-Bewertungssystem, geschwollene Gelenke, Parameter der Akute-Phase-Reaktion wie z. B. C-Reaktives Protein, Erythrozytensedimentationsrate) • Frühes Auftreten von Gelenkerosionen. b: basierend auf den Ergebnissen der mITT Population. Quelle: [2].	

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 nicht randomisierter vergleichender Studien und weiterer Untersuchungen

Sofern mit dem Dossier nicht randomisierte vergleichende Studien (Abschnitt 4.3.2.2) oder weitere Untersuchungen (Abschnitt 4.3.2.3) eingereicht wurden, nennen Sie die Gründe, nach denen es unmöglich oder unangemessen ist, zu den in diesen Studien 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 2005¹, Molenberghs 2010²). 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 2006³) 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 2006⁴) 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:

¹ Burzykowski T (Ed.): The evaluation of surrogate endpoints. New York: Springer; 2005.

² 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.

³ Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharm Stat* 2006; 5(3): 173-186.

⁴ Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med* 2006; 25(2): 183-203.

- 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.

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 Quellen (z. B. Publikationen, Studienberichte, Studienregistereinträge).

Studie	Quelle
RA-BEAM (JADV)	<u>Studienunterlagen</u> Studienbericht [9] Studienprotokoll [7] <u>Registereinträge</u> ClinicalTrials.gov [155] ICTRP [156] EU-CTR [157] PharmNet.Bund [158] <u>Publikation</u> Genovese et al., 2016 [159]

4.7 Referenzliste

Listen Sie nachfolgend alle Quellen (z. B. Publikationen, Studienberichte, Studienregister-einträge), die Sie im vorliegenden Dokument angegeben haben (als fortlaufend nummerierte Liste). Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard). Geben Sie bei Fachinformationen immer den Stand des Dokuments an.

1. Eli Lilly Nederland B.V. Fachinformation. Baricitinib. Olumiant® 2 mg/4 mg Filmtabletten. Zulassungsnummer(n): 2 mg EU/1/16/1170/001-008; 4 mg EU/1/16/1170/009-016. (Stand: Februar 2017). 2017.
2. Gemeinsamer Bundesausschuss. Niederschrift zum Beratungsgespräch gemäß § 8 AM-NutzenV, Beratungsanforderung 2015-B-127 zu LY3009104 (Baricitinib) zur Behandlung der moderaten bis schweren aktiven rheumatoiden Arthritis. 2016.
3. Bundesministerium der Justiz und für Verbraucherschutz. Verordnung über die Nutzenbewertung über die Nutzenbewertung Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V. (Arzneimittel-Nutzenbewertungsverordnung - AM-NutzenV). (Inkraftgetreten: 01.01.2011). 2014.
4. Smolen JS, Landewe R, Breedveld FC, Buch M, Burmester G, Dougados M, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs: 2013 update. *Ann Rheum Dis.* 2014;73(3):492-509.
5. Krüger K, Wollenhaupt J, Albrecht K, Alten R, Backhaus M, Baerwald C, et al. S1-Leitlinie der Deutschen Gesellschaft für Rheumatologie. Handlungsempfehlungen der DGRh zur sequenziellen medikamentösen Therapie der rheumatoiden Arthritis 2012: adaptierte EULAR Empfehlungen und aktualisierter Therapiealgorithmus. 2012.
6. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Allgemeine Methoden. Version 4.2. 2015.
7. Eli Lilly and Company. Protocol I4V-MC-JADV (c). A Randomized, Double-Blind, Placebo- and Active-Controlled, Phase 3 Study Evaluating the Efficacy and Safety of Baricitinib in Patients with Moderately to Severely Active Rheumatoid Arthritis Who Have Had an Inadequate Response to Methotrexate Therapy. 2013.
8. Eli Lilly and Company. Statistical Analysis Plan: I4V-MC-JADV. A Randomized, Double-Blind, Placebo- and Active-Controlled, Phase 3 Study Evaluating the Efficacy and Safety of Baricitinib in Patients with Moderately to Severely Active Rheumatoid Arthritis Who Have Had an Inadequate Response to Methotrexate Therapy. 2015.
9. Eli Lilly and Company. JADV Clinical Study Report. A Randomized, Double-Blind, Placebo- and Active-Controlled, Phase 3 Study Evaluating the Efficacy and Safety of Baricitinib in Patients with Moderately to Severely Active Rheumatoid Arthritis Who Have Had an Inadequate Response to Methotrexate Therapy. 2015.

10. Eli Lilly and Company. Post-hoc Analyseplan. RA-BEAM (JADV). 2017.
11. Eli Lilly and Company. Post-hoc Analysen RA-BEAM (JADV) -Subgruppenanalysen. 2017.
12. Singh JA, Saag KG, Bridges SL, Jr., Akl EA, Bannuru RR, Sullivan MC, et al. 2015 American College of Rheumatology Guideline for the Treatment of Rheumatoid Arthritis. *Arthritis Rheumatol.* 2016;68(1):1-26.
13. Smolen JS, Landewe R, Breedveld FC, Dougados M, Emery P, Gaujoux-Viala C, et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. *Ann Rheum Dis.* 2010;69(6):964-975.
14. Smolen JS, Breedveld FC, Burmester GR, Bykerk V, Dougados M, Emery P, et al. Treating rheumatoid arthritis to target: 2014 update of the recommendations of an international task force. *Ann Rheum Dis.* 2016;75(1):3-15.
15. Smolen JS, Aletaha D, McInnes IB. Rheumatoid arthritis. *Lancet.* 2016;388:2023-2038.
16. Wells G, Becker JC, Teng J, Dougados M, Schiff M, Smolen J, et al. Validation of the 28-joint Disease Activity Score (DAS28) and European League Against Rheumatism response criteria based on C-reactive protein against disease progression in patients with rheumatoid arthritis, and comparison with the DAS28 based on erythrocyte sedimentation rate. *Ann Rheum Dis.* 2009;68(6):954-960.
17. Schneider M. Interdisziplinäre Leitlinie. Management der frühen rheumatoiden Arthritis. 2011.
18. Anderson JK, Zimmerman L, Caplan L, Michaud K. Measures of rheumatoid arthritis disease activity: Patient (PtGA) and Provider (PrGA) Global Assessment of Disease Activity, Disease Activity Score (DAS) and Disease Activity Score with 28-Joint Counts (DAS28), Simplified Disease Activity Index (SDAI), Clinical Disease Activity Index (CDAI), Patient Activity Score (PAS) and Patient Activity Score-II (PASII), Routine Assessment of Patient Index Data (RAPID), Rheumatoid Arthritis Disease Activity Index (RADAI) and Rheumatoid Arthritis Disease Activity Index-5 (RADAI-5), Chronic Arthritis Systemic Index (CASI), Patient-Based Disease Activity Score With ESR (PDAS1) and Patient-Based Disease Activity Score without ESR (PDAS2), and Mean Overall Index for Rheumatoid Arthritis (MOI-RA). *Arthritis Care Res (Hoboken).* 2011;63 Suppl 11:S14-36.
19. van Gestel AM, Haagsma CJ, van Riel PL. Validation of rheumatoid arthritis improvement criteria that include simplified joint counts. *Arthritis Rheum.* 1998;41:1845-1850.
20. Aletaha D, Smolen J. The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): a review of their usefulness and validity in rheumatoid arthritis. *Clin Exp Rheumatol.* 2005;23(5 Suppl 39):S100-108.

21. Fransen J, van Riel PL. The Disease Activity Score and the EULAR response criteria. *Rheum Dis Clin North Am.* 2009;35(4):745-757, vii-viii.
22. Fransen J, van Riel PL. The Disease Activity Score and the EULAR response criteria. *Clin Exp Rheumatol.* 2005;23(5 Suppl 39):S93-99.
23. Fransen J, Visser K, van Dongen H, Huizinga T, van Riel PL, van der Heijde DM. Validity of the disease activity score in undifferentiated arthritis. *Arthritis Care Res (Hoboken).* 2010;62(10):1392-1398.
24. Uhlig T, Kvien TK, Pincus T. Test-retest reliability of disease activity core set measures and indices in rheumatoid arthritis. *Ann Rheum Dis.* 2009;68(6):972-975.
25. Prevoo ML, van Gestel AM, van THMA, van Rijswijk MH, van de Putte LB, van Riel PL. Remission in a prospective study of patients with rheumatoid arthritis. American Rheumatism Association preliminary remission criteria in relation to the disease activity score. *British journal of rheumatology.* 1996;35(11):1101-1105.
26. Radovits BJ, Kievit W, Fransen J, van de Laar MA, Jansen TL, van Riel PL, et al. Influence of age on the outcome of antitumour necrosis factor alpha therapy in rheumatoid arthritis. *Ann Rheum Dis.* 2009;68(9):1470-1473.
27. Aletaha D, Nell VP, Stamm T, Uffmann M, Pflugbeil S, Machold K, et al. Acute phase reactants add little to composite disease activity indices for rheumatoid arthritis: validation of a clinical activity score. *Arthritis Res Ther.* 2005;7:R796-806.
28. Wells GA, Boers M, Shea B, Brooks PM, Simon LS, Strand CV, et al. Minimal disease activity for rheumatoid arthritis: a preliminary definition. *J Rheumatol.* 2005;32(10):2016-2024.
29. Abe A, Ishikawa H, Murasawa A, Nakazono K. Disease activity and the course of elbow joint deterioration over 10 years in the patients with early rheumatoid arthritis. *Clin Rheumatol.* 2008;27(7):867-872.
30. Aletaha D, Landewe R, Karonitsch T, Bathon J, Boers M, Bombardier C, et al. Reporting disease activity in clinical trials of patients with rheumatoid arthritis: EULAR/ACR collaborative recommendations. *Annals of the Rheumatic Diseases.* 2008;67:1360-1364.
31. Fransen J, van Riel PL. DAS remission cut points. *Clin Exp Rheumatol.* 2006;24(6 Suppl 43):S-29-32.
32. Felson DT, Smolen JS, Wells G, Zhang B, van Tuyl LH, Funovits J, et al. American College of Rheumatology/European League against Rheumatism provisional definition of remission in rheumatoid arthritis for clinical trials. *Ann Rheum Dis.* 2011;70(3):404-413.

33. Kosinski M, Zhao SZ, Dedhiya S, Osterhaus JT, Ware JE, Jr. Determining minimally important changes in generic and disease-specific health-related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis Rheum.* 2000;43:1478-1487.
34. Schiff M, Luijckens K, Davies O, Yazici Y. AB0317 Association of ACR clinical responses with CDAI (clinical disease activity index) and RAPID3 (routine assessment of patient index data 3) indices of disease activity in rheumatoid arthritis patients treated with certolizumab pegol plus methotrexate. *Annals of the Rheumatic Diseases.* 2013;71(Suppl 3):655-655.
35. van Gestel AM, Prevoo ML, van 't Hof MA, van Rijswijk MH, van de Putte LB, van Riel PL. Development and validation of the European League Against Rheumatism response criteria for rheumatoid arthritis. Comparison with the preliminary American College of Rheumatology and the World Health Organization/International League Against Rheumatism Criteria. *Arthritis Rheum.* 1996;39(1):34-40.
36. Fifield J, Tennen H, Reisine S, McQuillan J. Depression and the long-term risk of pain, fatigue, and disability in patients with rheumatoid arthritis. *Arthritis Rheum.* 1998;41(10):1851-1857.
37. Hagen KB, Kvien TK, Bjorndal A. Musculoskeletal pain and quality of life in patients with noninflammatory joint pain compared to rheumatoid arthritis: a population survey. *J Rheumatol.* 1997;24(9):1703-1709.
38. Hakkinen A, Kautiainen H, Hannonen P, Ylinen J, Arkela-Kautiainen M, Sokka T. Pain and joint mobility explain individual subdimensions of the health assessment questionnaire (HAQ) disability index in patients with rheumatoid arthritis. *Ann Rheum Dis.* 2005;64(1):59-63.
39. Sokka T, Krishnan E, Hakkinen A, Hannonen P. Functional disability in rheumatoid arthritis patients compared with a community population in Finland. *Arthritis Rheum.* 2003;48(1):59-63.
40. Taylor PC, Moore A, Vasilescu R, Alvir J, Tarallo M. A structured literature review of the burden of illness and unmet needs in patients with rheumatoid arthritis: a current perspective. *Rheumatol Int.* 2016;36(5):685-695.
41. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis (A-10-01). 2013.
42. Pope J, Bingham CO, 3rd, Fleischmann RM, Dougados M, Massarotti EM, Wollenhaupt J, et al. Impact of certolizumab pegol on patient-reported outcomes in rheumatoid arthritis and correlation with clinical measures of disease activity. *Arthritis Res Ther.* 2015;17:343.

43. Wolfe F, Michaud K. Assessment of pain in rheumatoid arthritis: minimal clinically significant difference, predictors, and the effect of anti-tumor necrosis factor therapy. *J Rheumatol.* 2007;34(8):1674-1683.
44. Stauffer ME, Taylor SD, Watson DJ, Peloso PM, Morrison A. Definition of nonresponse to analgesic treatment of arthritic pain: an analytical literature review of the smallest detectable difference, the minimal detectable change, and the minimal clinically important difference on the pain visual analog scale. *Int J Inflam.* 2011;2011:231926.
45. Yazici Y, Pincus T, Kautiainen H, Sokka T. Morning stiffness in patients with early rheumatoid arthritis is associated more strongly with functional disability than with joint swelling and erythrocyte sedimentation rate. *J Rheumatol.* 2004;31(9):1723-1726.
46. Hurst NP, Kind P, Ruta D, Hunter M, Stubbings A. Measuring health-related quality of life in rheumatoid arthritis: validity, responsiveness and reliability of EuroQol (EQ-5D). *British journal of rheumatology.* 1997;36(5):551-559.
47. Repping-Wuts H, Hewlett S, van Riel P, van Achterberg T. Fatigue in patients with rheumatoid arthritis: British and Dutch nurses' knowledge, attitudes and management. *J Adv Nurs.* 2009;65(4):901-911.
48. Hewlett S, Cockshott Z, Byron M, Kitchen K, Tipler S, Pope D, et al. Patients' perceptions of fatigue in rheumatoid arthritis: overwhelming, uncontrollable, ignored. *Arthritis Rheum.* 2005;53(5):697-702.
49. Rupp I, Boshuizen HC, Jacobi CE, Dinant HJ, van den Bos GA. Impact of fatigue on health-related quality of life in rheumatoid arthritis. *Arthritis Rheum.* 2004;51(4):578-585.
50. Suurmeijer TP, Waltz M, Moum T, Guillemin F, van Sonderen FL, Briancon S, et al. Quality of life profiles in the first years of rheumatoid arthritis: results from the EURIDISS longitudinal study. *Arthritis Rheum.* 2001;45(2):111-121.
51. Hewlett S, Chalder T, Choy E, Cramp F, Davis B, Dures E, et al. Fatigue in rheumatoid arthritis: time for a conceptual model. *Rheumatology.* 2011;50:1004-1006.
52. Hewlett S, Dures E, Almeida C. Measures of fatigue: Bristol Rheumatoid Arthritis Fatigue Multi-Dimensional Questionnaire (BRAFMQ), Bristol Rheumatoid Arthritis Fatigue Numerical Rating Scales (BRAFNRS) for severity, effect, and coping, Chalder Fatigue Questionnaire (CFQ), Checklist Individual Strength (CIS20R and CIS8R), Fatigue Severity Scale (FSS), Functional Assessment Chronic Illness Therapy (Fatigue) (FACIT-F), Multi-Dimensional Assessment of Fatigue (MAF), Multi-Dimensional Fatigue Inventory (MFI), Pediatric Quality Of Life (PedsQL) Multi-Dimensional Fatigue Scale, Profile of Fatigue (ProF), Short Form 36 Vitality Subscale (SF-36 VT), and Visual Analog Scales (VAS). *Arthritis Care Res (Hoboken).* 2011;63 Suppl 11:S263-286.

53. Nicassio PM, Ormseth SR, Custodio MK, Irwin MR, Olmstead R, Weisman MH. A multidimensional model of fatigue in patients with rheumatoid arthritis. *J Rheumatol.* 2012;39(9):1807-1813.
54. Druce KL, Jones GT, Macfarlane GJ, Basu N. Patients receiving anti-TNF therapies experience clinically important improvements in RA-related fatigue: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. *Rheumatology (Oxford).* 2015;54(6):964-971.
55. Ellert U, Kurth BM. Methodische Betrachtungen zu den Summenscores des SF-36 anhand der erwachsenen bundesdeutschen Bevölkerung. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz.* 2004;47(11):1027-1032.
56. Keystone EC, Kavanaugh AF, Sharp JT, Tannenbaum H, Hua Y, Teoh LS, et al. Radiographic, clinical, and functional outcomes of treatment with adalimumab (a human anti-tumor necrosis factor monoclonal antibody) in patients with active rheumatoid arthritis receiving concomitant methotrexate therapy: a randomized, placebo-controlled, 52-week trial. *Arthritis Rheum.* 2004;50(5):1400-1411.
57. Tuttleman M, Pillemer SR, Tilley BC, Fowler SE, Buckley LM, Alarcon GS, et al. A cross sectional assessment of health status instruments in patients with rheumatoid arthritis participating in a clinical trial. *Minocycline in Rheumatoid Arthritis Trial Group. J Rheumatol.* 1997;24(10):1910-1915.
58. van Vollenhoven RF, Fleischmann R, Cohen S, Lee EB, Garcia Mejjide JA, Wagner S, et al. Tofacitinib or adalimumab versus placebo in rheumatoid arthritis. *N Engl J Med.* 2012;367(6):508-519.
59. Eli Lilly and Company. Post-hoc Analysen RA-BEAM (JADV) - Wirksamkeitsendpunkte. 2017.
60. Eli Lilly and Company. Post-hoc Analysen RA-BEAM (JADV) - Sicherheitsendpunkte. 2017.
61. Europäische Union. RICHTLINIE 2005/28/EG DER KOMMISSION vom 8. April 2005 zur Festlegung von Grundsätzen und ausführlichen Leitlinien der guten klinischen Praxis für zur Anwendung beim Menschen bestimmte Prüfpräparate sowie von Anforderungen für die Erteilung einer Genehmigung zur Herstellung oder Einfuhr solcher Produkte (Text von Bedeutung für den EWR). 2005.
62. Smolen JS, Steiner G. Therapeutic strategies for rheumatoid arthritis. *Nat Rev Drug Discov.* 2003;2(6):473-488.
63. Colmegna I, Ohata BR, Menard HA. Current understanding of rheumatoid arthritis therapy. *Clin Pharmacol Ther.* 2012;91(4):607-620.
64. Nam JL, Ramiro S, Gaujoux-Viala C, Takase K, Leon-Garcia M, Emery P, et al. Efficacy of biological disease-modifying antirheumatic drugs: a systematic literature review informing the 2013 update of the EULAR recommendations for the management of rheumatoid arthritis. *Ann Rheum Dis.* 2014;73(3):516-528.

65. Klareskog L, Catrina AI, Paget S. Rheumatoid arthritis. *Lancet*. 2009;373(9664):659-672.
66. Rubbert-Roth A, Finckh A. Treatment options in patients with rheumatoid arthritis failing initial TNF inhibitor therapy: a critical review. *Arthritis Res Ther*. 2009;11 Suppl 1:S1.
67. Pollard L, Choy EH, Scott DL. The consequences of rheumatoid arthritis: quality of life measures in the individual patient. *Clin Exp Rheumatol*. 2005;23(5 Suppl 39):S43-52.
68. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses. (Stand: 24. Januar 2017). 2017.
69. European Medicines Agency (EMA). Population Exposure: The Extent of Population Exposure to Assess Clinical Safety - CPMP/ICH/375/95 (ICH 152 E1). 1995.
70. Wong SS, 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.
71. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: Updated guidelines for reporting parallel group randomised trials. *J Pharmacol Pharmacother*. 2010;1(2):100-107.
72. Schulz KF, Moher D, Altman DG. CONSORT 2010 comments. *Lancet*. 2010;376(9748):1222-1223.
73. Bundesministerium der Justiz und für Verbraucherschutz. Sozialgesetzbuch (SGB) Fünftes Buch (V) - Gesetzliche Krankenversicherung - (Artikel 1 des Gesetzes v. 20. Dezember 1988, BGBl. I S. 2477). 2016.
74. Aletaha D, Alasti F, Smolen JS. Optimisation of a treat-to-target approach in rheumatoid arthritis: strategies for the 3-month time point. *Ann Rheum Dis*. 2016;75(8):1479-1485.
75. Pincus T, Callahan LF. Quantitative measures to assess, monitor and predict morbidity and mortality in rheumatoid arthritis. *Baillieres Clin Rheumatol*. 1992;6(1):161-191.
76. Ritchie DM, Boyle JA, McInnes JM, Jasani MK, Dalakos TG, Grieveson P, et al. Clinical studies with an articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. *Q J Med*. 1968;37(147):393-406.
77. Fuchs HA, Brooks RH, Callahan LF, Pincus T. A simplified twenty-eight-joint quantitative articular index in rheumatoid arthritis. *Arthritis Rheum*. 1989;32(5):531-537.
78. Scott DL, Choy EH, Greeves A, Isenberg D, Kassiror D, Rankin E, et al. Standardising joint assessment in rheumatoid arthritis. *Clin Rheumatol*. 1996;15(6):579-582.

79. Scott DL, Houssien DA. Joint assessment in rheumatoid arthritis. *British journal of rheumatology*. 1996;35 Suppl 2:14-18.
80. Smolen JS, Breedveld FC, Eberl G, Jones I, Leeming M, Wylie GL, et al. Validity and reliability of the twenty-eight-joint count for the assessment of rheumatoid arthritis activity. *Arthritis Rheum*. 1995;38(1):38-43.
81. Fries J, Spitz P, Kraines R, HR H. Measurement of patient outcome in arthritis. *Arthritis & Rheumatism*. 1980;23(2):137-145.
82. Merkel PA, Herlyn K, Martin RW, Anderson JJ, Mayes MD, Bell P, et al. Measuring disease activity and functional status in patients with scleroderma and Raynaud's phenomenon. *Arthritis Rheum*. 2002;46(9):2410-2420.
83. Merkel PA, Seo P, Aries P, Neogi T, Villa-Forte A, Boers M, et al. Current status of outcome measures in vasculitis: focus on Wegener's granulomatosis and microscopic polyangiitis. Report from OMERACT 7. *J Rheumatol*. 2005;32(12):2488-2495.
84. Sokka T, Pincus T. Quantitative joint assessment in rheumatoid arthritis. *Clin Exp Rheumatol*. 2005;23(5 Suppl 39):S58-62.
85. Keystone E, Emery P, Peterfy CG, Tak PP, Cohen S, Genovese MC, et al. Rituximab inhibits structural joint damage in patients with rheumatoid arthritis with an inadequate response to tumour necrosis factor inhibitor therapies. *Ann Rheum Dis*. 2009;68(2):216-221.
86. Smolen JS, Aletaha D, Machold KP. Therapeutic strategies in early rheumatoid arthritis. *Best Pract Res Clin Rheumatol*. 2005;19(1):163-177.
87. Wollenhaupt J, Albrecht K, Kruger K, Muller-Ladner U. The new 2012 German recommendations for treating rheumatoid arthritis : differences compared to the European standpoint. *Z Rheumatol*. 2013;72(1):6-9.
88. Cheung PP, Mari K, Devauchelle-Pensec V, Jousse-Joulin S, D'Agostino MA, Chales G, et al. Predictive value of tender joints compared to synovitis for structural damage in rheumatoid arthritis. *RMD Open*. 2016;2(1):e000205.
89. Ziegler S, Huscher D, Karberg K, Krause A, Wassenberg S, Zink A. Trends in treatment and outcomes of rheumatoid arthritis in Germany 1997-2007: results from the National Database of the German Collaborative Arthritis Centres. *Ann Rheum Dis*. 2010;69(10):1803-1808.
90. Vastesaeger N, Xu S, Aletaha D, St Clair EW, Smolen JS. A pilot risk model for the prediction of rapid radiographic progression in rheumatoid arthritis. *Rheumatology (Oxford)*. 2009;48(9):1114-1121.
91. Westhovens R, Robles M, Ximenes AC, Nayiager S, Wollenhaupt J, Durez P, et al. Clinical efficacy and safety of abatacept in methotrexate-naive patients with early rheumatoid arthritis and poor prognostic factors. *Ann Rheum Dis*. 2009;68(12):1870-1877.

92. Weinblatt ME, Keystone EC, Cohen MD, Freundlich B, Li J, Chon Y, et al. Factors associated with radiographic progression in patients with rheumatoid arthritis who were treated with methotrexate. *J Rheumatol.* 2011;38(2):242-246.
93. Vander Cruyssen B, Van Looy S, Wyns B, Westhovens R, Durez P, Van den Bosch F, et al. DAS28 best reflects the physician's clinical judgment of response to infliximab therapy in rheumatoid arthritis patients: validation of the DAS28 score in patients under infliximab treatment. *Arthritis Res Ther.* 2005;7(5):R1063-1071.
94. Arya V, Malaviya A, Raja R. CDAI (clinical disease activity index) in rheumatoid arthritis: cut-off values for classification into different grades of disease activity. *Indian Journal of Rheumatology.* 2007;2(3):91-94.
95. Gossec L, Dougados M, Rincheval N, Balanescu A, Boumpas DT, Canadelo S, et al. Elaboration of the preliminary Rheumatoid Arthritis Impact of Disease (RAID) score: a EULAR initiative. *Ann Rheum Dis.* 2009;68:1680-1685.
96. Walsh DA, McWilliams DF. Pain in rheumatoid arthritis. *Curr Pain Headache Rep.* 2012;16(6):509-517.
97. Lee YC. Effect and treatment of chronic pain in inflammatory arthritis. *Curr Rheumatol Rep.* 2013;15(1):300.
98. Heiberg T, Kvien TK. Preferences for improved health examined in 1,024 patients with rheumatoid arthritis: pain has highest priority. *Arthritis Rheum.* 2002;47(4):391-397.
99. Taylor P, Manger B, Alvaro-Gracia J, Johnstone R, Gomez-Reino J, Eberhardt E, et al. Patient perceptions concerning pain management in the treatment of rheumatoid arthritis. *J Int Med Res.* 2010;38(4):1213-1224.
100. Heiberg T, Finset A, Uhlig T, Kvien TK. Seven year changes in health status and priorities for improvement of health in patients with rheumatoid arthritis. *Ann Rheum Dis.* 2005;64(2):191-195.
101. ten Klooster PM, Veehof MM, Taal E, van Riel PL, van de Laar MA. Changes in priorities for improvement in patients with rheumatoid arthritis during 1 year of anti-tumour necrosis factor treatment. *Ann Rheum Dis.* 2007;66(11):1485-1490.
102. Minnock P, FitzGerald O, Bresnihan B. Women with established rheumatoid arthritis perceive pain as the predominant impairment of health status. *Rheumatology (Oxford).* 2003;42(8):995-1000.
103. Kvien TK. Epidemiology and burden of illness of rheumatoid arthritis. *Pharmacoeconomics.* 2004;22(2 Suppl 1):1-12.
104. Sokka T, Toloza S, Cutolo M, Kautiainen H, Makinen H, Gogus F, et al. Women, men, and rheumatoid arthritis: analyses of disease activity, disease characteristics, and treatments in the QUEST-RA study. *Arthritis Res Ther.* 2009;11(1):R7.

105. Courvoisier DS, Agoritsas T, Glauser J, Michaud K, Wolfe F, Cantoni E, et al. Pain as an important predictor of psychosocial health in patients with rheumatoid arthritis. *Arthritis Care Res (Hoboken)*. 2012;64(2):190-196.
106. Wewers ME, Lowe NK. A critical review of visual analogue scales in the measurement of clinical phenomena. *Res Nurs Health*. 1990;13(4):227-236.
107. Dworkin RH, Turk DC, Wyrwich KW, Beaton D, Cleeland CS, Farrar JT, et al. Interpreting the clinical importance of treatment outcomes in chronic pain clinical trials: IMMPACT recommendations. *J Pain*. 2007;9:105-121.
108. Wells G, Li T, Maxwell L, MacLean R, Tugwell P. Determining the minimal clinically important differences in activity, fatigue, and sleep quality in patients with rheumatoid arthritis. *J Rheumatol*. 2007;34:280-289.
109. Wells GA, Tugwell P, Kraag GR, Baker PR, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol*. 1993;20:557-560.
110. Pope JE, Khanna D, Norrie D, Ouimet JM. The minimally important difference for the health assessment questionnaire in rheumatoid arthritis clinical practice is smaller than in randomized controlled trials. *J Rheumatol*. 2009;36:254-259.
111. Kvamme MK, Kristiansen IS, Lie E, Kvien TK. Identification of cutpoints for acceptable health status and important improvement in patient-reported outcomes, in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis. *J Rheumatol*. 2009;37:26-31.
112. Strand V, Boers M, Idzerda L, Kirwan JR, Kvien TK, Tugwell PS, et al. It's Good to Feel Better But It's Better To Feel Good and Even Better to Feel Good as Soon as Possible for as Long as Possible. Response Criteria and the Importance of Change at OMERACT 10. *The Journal of Rheumatology*. 2011;38:1720-1727.
113. Tubach F, Ravaud P, Martin-Mola E, Awada H, Bellamy N, Bombardier C, et al. Minimum clinically important improvement and patient acceptable symptom state in pain and function in rheumatoid arthritis, ankylosing spondylitis, chronic back pain, hand osteoarthritis, and hip and knee osteoarthritis: Results from a prospective multinational study. *Arthritis Care Res (Hoboken)*. 2012;64:1699-1707.
114. Ward MM, Guthrie LC, Alba M. Dependence of the minimal clinically important improvement on the baseline value is a consequence of floor and ceiling effects and not different expectations by patients. *Journal of clinical epidemiology*. 2014;67:689-696.
115. Ramey D, Fries JF, Singh G. The Health Assessment Questionnaire 1995 - status and review. *Quality of Life and Pharmacoeconomics in clinical trials*. 1996;2:227-237.
116. Redelmeier DA, Lorig K. Assessing the clinical importance of symptomatic improvements. An illustration in rheumatology. *Arch Intern Med*. 1993;153:1337-1342.

117. Westhoff G, Buttgereit F, Gromnica-Ihle E, Zink A. Morning stiffness and its influence on early retirement in patients with recent onset rheumatoid arthritis. *Rheumatology (Oxford)*. 2008;47(7):980-984.
118. Sierakowski S, Cutolo M. Morning symptoms in rheumatoid arthritis: a defining characteristic and marker of active disease. *Scand J Rheumatol Suppl*. 2011;125:1-5.
119. Cutolo M, Otsa K, Aakre O, Sulli A. Nocturnal hormones and clinical rhythms in rheumatoid arthritis. *Ann N Y Acad Sci*. 2005;1051:372-381.
120. Hazes JM, Hayton R, Burt J, Silman AJ. Consistency of morning stiffness: an analysis of diary data. *British journal of rheumatology*. 1994;33(6):562-565.
121. Halls S, Dures E, Kirwan J, Pollock J, Baker G, Edmunds A, et al. Stiffness is more than just duration and severity: a qualitative exploration in people with rheumatoid arthritis. *Rheumatology (Oxford)*. 2015;54(4):615-622.
122. Orbai AM, Smith KC, Bartlett SJ, De Leon E, Bingham CO, 3rd. "Stiffness has different meanings, I think, to everyone": examining stiffness from the perspective of people living with rheumatoid arthritis. *Arthritis Care Res (Hoboken)*. 2014;66(11):1662-1672.
123. da Silva JA, Phillips S, Buttgereit F. Impact of impaired morning function on the lives and well-being of patients with rheumatoid arthritis. *Scand J Rheumatol Suppl*. 2011;125:6-11.
124. Khan NA, Yazici Y, Calvo-Alen J, Dadoniene J, Gossec L, Hansen TM, et al. Reevaluation of the role of duration of morning stiffness in the assessment of rheumatoid arthritis activity. *J Rheumatol*. 2009;36(11):2435-2442.
125. Gignac MA, Cao X, McAlpine J, Badley EM. Measures of disability: Arthritis Impact Measurement Scales 2 (AIMS2), Arthritis Impact Measurement Scales 2-Short Form (AIMS2-SF), The Organization for Economic Cooperation and Development (OECD) Long-Term Disability (LTD) Questionnaire, EQ-5D, World Health Organization Disability Assessment Schedule II (WHODASII), Late-Life Function and Disability Instrument (LLFDI), and Late-Life Function and Disability Instrument-Abbreviated Version (LLFDI-Abbreviated). *Arthritis Care Res (Hoboken)*. 2011;63 Suppl 11:S308-324.
126. Fransen M, Edmonds J. Reliability and validity of the EuroQol in patients with osteoarthritis of the knee. *Rheumatology (Oxford)*. 1999;38(9):807-813.
127. Gemeinsamer Bundesausschuss. Bosutinib - Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V. 2013.
128. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Ibrutinib - Nutzenbewertung gemäß § 35a SGB V (A16-04). 2016.
129. Pickard AS, Neary MP, Cella D. Estimation of minimally important differences in EQ-5D utility and VAS scores in cancer. *Health Qual Life Outcomes*. 2007;5:70.

130. Brazier JE, Harper R, Jones NM, O'Cathain A, Thomas KJ, Usherwood T, et al. Validating the SF-36 health survey questionnaire: new outcome measure for primary care. *BMJ*. 1992;305(6846):160-164.
131. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. *Med Care*. 1992;30(6):473-483.
132. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. *Med Care*. 1996;34(3):220-233.
133. Keystone E, Burmester GR, Furie R, Loveless JE, Emery P, Kremer J, et al. Improvement in patient-reported outcomes in a rituximab trial in patients with severe rheumatoid arthritis refractory to anti-tumor necrosis factor therapy. *Arthritis Rheum*. 2008;59(6):785-793.
134. Strand V, Rentz AM, Cifaldi MA, Chen N, Roy S, Revicki D. Health-related quality of life outcomes of adalimumab for patients with early rheumatoid arthritis: results from a randomized multicenter study. *J Rheumatol*. 2012;39(1):63-72.
135. Genovese MC, Schiff M, Luggen M, Becker JC, Aranda R, Teng J, et al. Efficacy and safety of the selective co-stimulation modulator abatacept following 2 years of treatment in patients with rheumatoid arthritis and an inadequate response to anti-tumour necrosis factor therapy. *Ann Rheum Dis*. 2008;67(4):547-554.
136. Lubeck DP. Patient-reported outcomes and their role in the assessment of rheumatoid arthritis. *Pharmacoeconomics*. 2004;22:27-38.
137. Coombs JH, Bloom BJ, Breedveld FC, Fletcher MP, Gruben D, Kremer JM, et al. Improved pain, physical functioning and health status in patients with rheumatoid arthritis treated with CP-690,550, an orally active Janus kinase (JAK) inhibitor: results from a randomised, double-blind, placebo-controlled trial. *Ann Rheum Dis*. 2010;69(2):413-416.
138. The International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH). *Clinical Safety Data Management: Definitions and Standards for Expedited Reporting EA2*. 1994.
139. Katz PP, Yelin EH. Prevalence and correlates of depressive symptoms among persons with rheumatoid arthritis. *J Rheumatol*. 1993;20(5):790-796.
140. VanDyke MM, Parker JC, Smarr KL, Hewett JE, Johnson GE, Slaughter JR, et al. Anxiety in rheumatoid arthritis. *Arthritis Rheum*. 2004;51(3):408-412.
141. Zautra AJ, Burleson MH, Matt KS, Roth S, Burrows L. Interpersonal stress, depression, and disease activity in rheumatoid arthritis and osteoarthritis patients. *Health Psychol*. 1994;13(2):139-148.
142. Sampson HA. Prevention of atopy and allergic disease: type of infant formula. *Nestle Nutr Workshop Ser Pediatr Program*. 2006;57:109-117; discussion 118-123.

143. Sampson HA. Anaphylaxis: persistent enigma. *Emerg Med Australas.* 2006;18(2):101-102.
144. Bradburn M, Deeks JJ, Berlin JA, Russell Localio A. Much ado about nothing: a comparison of the performance of meta-analytical methods with rare events. *Statistics in medicine.* 2007;26(1):53-77.
145. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. *Introduction to Meta-Analysis. Further Methods for Dichotomous Data: John Wiley & Sons, Ltd.; 2009. p. 331-339.*
146. StatsDirect Limited. StatsDirect. 2016. Verfügbar unter: http://www.statsdirect.com/help/default.htm#meta_analysis/peto.htm.
147. Keystone E, Heijde D, Mason D, Jr., Landewe R, Vollenhoven RV, Combe B, et al. Certolizumab pegol plus methotrexate is significantly more effective than placebo plus methotrexate in active rheumatoid arthritis: findings of a fifty-two-week, phase III, multicenter, randomized, double-blind, placebo-controlled, parallel-group study. *Arthritis Rheum.* 2008;58(11):3319-3329.
148. Cohen SB, Emery P, Greenwald MW, Dougados M, Furie RA, Genovese MC, et al. Rituximab for rheumatoid arthritis refractory to anti-tumor necrosis factor therapy: Results of a multicenter, randomized, double-blind, placebo-controlled, phase III trial evaluating primary efficacy and safety at twenty-four weeks. *Arthritis Rheum.* 2006;54(9):2793-2806.
149. Smolen JS, Beaulieu A, Rubbert-Roth A, Ramos-Remus C, Rovensky J, Alecock E, et al. Effect of interleukin-6 receptor inhibition with tocilizumab in patients with rheumatoid arthritis (OPTION study): a double-blind, placebo-controlled, randomised trial. *Lancet.* 2008;371(9617):987-997.
150. Smolen JS, Wasko MC, Ramos-Remus CR, Mysler EF, Day BM, Wiese C. Rheumatoid arthritis patients treated with tocilizumab show improved haemoglobin levels and facit-fatigue scores. *Rheumatology (Oxford).* 2009;48((Suppl 1):i81).
151. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). *Allgemeine Methoden. Entwurf für Version 5.0 vom 07.12.2016.* 2016.
152. Kraemer HC, Frank E, Kupfer DJ. Moderators of treatment outcomes: clinical, research, and policy importance. *JAMA.* 2006;296(10):1286-1289.
153. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet.* 2005;365(9454):176-186.
154. Eli Lilly and Company. *Tabular Listing of All Clinical Studies. Baricitinib. Rheumatoide Arthritis.* 2015.
155. Eli Lilly and Company. *Registerbericht. ClinicalTrials.gov. RA-BEAM (JADV).* 2017.
156. Eli Lilly and Company. *Registerbericht. International Clinical Trials Registry Platform. RA-BEAM (JADV).* 2017.

157. Eli Lilly and Company. Registerbericht. EU Clinical Trials Register. RA-BEAM (JADV). 2017.
158. Eli Lilly and Company. Registerbericht. PharmNetBund. RA-BEAM (JADV). 2017.
159. Genovese MC, Kremer J, Zamani O, Ludivico C, Krogulec M, Xie L, et al. Baricitinib in Patients with Refractory Rheumatoid Arthritis. *N Engl J Med.* 2016;374(13):1243-1252.
160. Eli Lilly and Company. Post-hoc Analysen RA-BEAM (JADV) - Charakterisierung der Studienpopulation. 2017.
161. Eli Lilly and Company. Abbildungen kontinuierliche Endpunkte (mITT Population) der Studie RA-BEAM (JADV). 2017.
162. Marra CA, Woolcott JC, Kopec JA, Shojania K, Offer R, Brazier JE, et al. A comparison of generic, indirect utility measures (the HUI2, HUI3, SF-6D, and the EQ-5D) and disease-specific instruments (the RAQoL and the HAQ) in rheumatoid arthritis. *Soc Sci Med.* 2005;60(7):1571-1582.
163. Hubert MB, Reitmeier P, Vogelmann M, Leidl R. EQ-5D-5L in the General German Population: Comparison and Evaluation of Three Yearly Cross-Section Surveys. *Int J Environ Res Public Health.* 2016;13(3):343.
164. Szende A, Williams A. Measuring Self-Reported Population Health: An International Perspective based on EQ-5D. 2004.

Anhang 4-A: Suchstrategien – bibliografische Literaturrecherche

Geben Sie nachfolgend die Suchstrategien für die bibliografische(n) Literaturrecherche(n) an, und zwar getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.). Für jede durchsuchte Datenbank ist die verwendete Strategie separat darzustellen. 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. Orientieren Sie sich bei Ihren Angaben an dem nachfolgenden Beispiel (eine umfassende Suche soll Freitextbegriffe und Schlagwörter enthalten):

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 ¹) – 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

¹ 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. Hinweis: Für die Suche in der Cochrane-Datenbank „Cochrane Central Register of Controlled Trials (Clinical Trials)“ sollte kein Studienfilter verwendet werden.

Anhang 4-A1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Es wurde eine Suchstrategie für die bibliografische Literaturrecherche für die Suche nach RCT mit dem zu bewertenden Arzneimittel für die Subpopulationen A1, A2 sowie A3 angewendet. Deshalb wird an dieser Stelle auf eine Mehrfachdarstellung der Suchstrategie für die verschiedenen Populationen verzichtet.

Datenbankname	EBM Reviews - Cochrane Central Register of Controlled Trials November 2016 (Cochrane)	
Suchoberfläche	Ovid	
Datum der Suche	02.02.2017	
Zeitsegment	1991 bis Datum der Suche	
Suchfilter	keiner	
# (a)	Suchbegriffe	Ergebnis
1	Baricitinib.mp.	45
2	Olumiant.mp.	0
3	("INCB 028050" or INCB 28050 or INCB028050 or INCB28050).mp.	14
4	(LY 3009104 or LY3009104).mp.	13
5	1 or 2 or 3 or 4	46
6	exp Arthritis, Rheumatoid/	3.950
7	(Rheumat* adj3 Arthrit*).mp.	7.789
8	6 or 7	7.944
9	5 and 8	38
10	remove duplicates from 9	33

Datenbankname	Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R) (MEDLINE)	
Suchoberfläche	Ovid	
Datum der Suche	02.02.2017	
Zeitsegment	1946 bis Datum der Suche	
Suchfilter	Filter für randomisierte kontrollierte Studien nach Wong 2006 (Quelle ¹) – Strategy minimizing difference between sensitivity and specificity	
# (b)	Suchbegriffe	Ergebnis
1	Baricitinib.mp.	34
2	Olumiant.mp.	0
3	("INCB 028050" or INCB 28050 or INCB028050 or INCB28050).mp.	6
4	(LY 3009104 or LY3009104).mp.	4
5	1 or 2 or 3 or 4	36
6	exp Arthritis, Rheumatoid/	101.649
7	(Rheumat* adj3 Arthrit*).mp.	120.262
8	6 or 7	131.744
9	5 and 8	27
10	randomized controlled trial.pt.	447.100
11	controlled clinical trial.pt.	91.836
12	randomized.ab.	387.340
13	placebo.ab.	183.235
14	clinical trials as topic.sh.	181.018
15	randomly.ab.	270.704
16	trial.ti.	173.473
17	10 or 11 or 12 or 13 or 14 or 15 or 16	1.097.532
18	exp animals/ not humans.sh.	4.313.668
19	17 not 18	1.012.820
20	randomized controlled trial.pt.	447.100
21	randomized.mp.	703.778
22	placebo:.mp.	203.244
23	20 or 21 or 22	770.827
24	9 and 19	12
25	9 and 23	13

¹ 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. Hinweis: Für die Suche in der Cochrane-Datenbank „Cochrane Central Register of Controlled Trials (Clinical Trials)“ sollte kein Studienfilter verwendet werden.

Datenbankname	Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R) (MEDLINE)	
Suchoberfläche	Ovid	
Datum der Suche	02.02.2017	
Zeitsegment	1946 bis Datum der Suche	
Suchfilter	Filter für randomisierte kontrollierte Studien nach Wong 2006 (Quelle ¹) – Strategy minimizing difference between sensitivity and specificity	
# (b)	Suchbegriffe	Ergebnis
26	24 or 25	13
27	limit 26 to (english or german)	12
28	remove duplicates from 27	11

Datenbankname	Embase 1974 to 2017 February 01 (EMBASE)	
Suchoberfläche	OVID	
Datum der Suche	02.02.2017	
Zeitsegment	1974 bis Datum der Suche	
Suchfilter	Filter für randomisierte kontrollierte Studien nach Wong 2006 (Quelle ¹) – Strategy minimizing difference between sensitivity and specificity	
# (c)	Suchbegriffe	Ergebnis
1	baricitinib/	273
2	Baricitinib.mp.	273
3	Olumiant.mp.	0
4	("INCB 028050" or INCB 28050 or INCB028050 or INCB28050).mp.	77
5	(LY 3009104 or LY3009104).mp.	53
6	1 or 2 or 3 or 4 or 5	301
7	exp rheumatoid arthritis/	182.194
8	(Rheumat* adj3 Arthrit*).mp.	193.592
9	7 or 8	195.759
10	6 and 9	212
11	random*.tw.	1.175.996
12	placebo*.mp.	394.749
13	double-blind*.tw.	177.576
14	11 or 12 or 13	1.414.227
15	10 and 14	122
16	15 not Medline.cr.	120
17	limit 16 to (english or german)	120
18	remove duplicates from 17	93

¹ 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. Hinweis: Für die Suche in der Cochrane-Datenbank „Cochrane Central Register of Controlled Trials (Clinical Trials)“ sollte kein Studienfilter verwendet werden.

Aus Gründen der Nachvollziehbarkeit werden in der nachfolgenden Tabelle die Schritte der Duplikateliminierung für die verschiedenen Datenbanken dargestellt.

Datenbankname	Duplikateliminierung - Zusammengefasste Suche in mehreren Datenbanken (a) EBM Reviews - Cochrane Central Register of Controlled Trials November 2016 (Cochrane) (b) Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R) (MEDLINE) (c) Embase 1974 to 2017 February 01 (EMBASE)	
Suchoberfläche	Ovid	
Datum der Suche	02.02.2017	
Zeitsegment	1946 bis Datum der Suche	
Suchfilter	Duplikateliminierung in den Datenbanken Cochrane, Medline und EMBASE	
#	Suchbegriffe	Ergebnis
1	Baricitinib.mp.	352
2	Olumiant.mp.	0
3	("INCB 028050" or INCB 28050 or INCB028050 or INCB28050).mp.	97
4	(LY 3009104 or LY3009104).mp.	70
5	1 or 2 or 3 or 4	383
6	exp Arthritis, Rheumatoid/	287.793
7	(Rheumat* adj3 Arthrit*).mp.	321.643
8	6 or 7	335.447
9	5 and 8	277
	EBM Reviews - Cochrane Central Register of Controlled Trials <November 2016>	38
	Embase <1974 to 2017 February 01>	212
	Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R)	27
10	randomized controlled trial.pt.	854.581
11	controlled clinical trial.pt.	180.348
12	randomized.ab.	1.192.756
13	placebo.ab.	579.739
14	clinical trials as topic.sh.	214.137
15	randomly.ab.	754.347
16	trial.ti.	566.838
17	10 or 11 or 12 or 13 or 14 or 15 or 16	2.825.542
18	exp animals/ not humans.sh.	27.305.944
19	17 not 18	1.785.154
20	randomized controlled trial.pt.	854.581
21	randomized.mp.	1.948.670

Datenbankname	Duplikateliminierung - Zusammengefasste Suche in mehreren Datenbanken (a) EBM Reviews - Cochrane Central Register of Controlled Trials November 2016 (Cochrane) (b) Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R) (MEDLINE) (c) Embase 1974 to 2017 February 01 (EMBASE)	
Suchoberfläche	Ovid	
Datum der Suche	02.02.2017	
Zeitsegment	1946 bis Datum der Suche	
Suchfilter	Duplikateliminierung in den Datenbanken Cochrane, Medline und EMBASE	
22	placebo:.mp.	777.784
23	20 or 21 or 22	2.451.842
24	9 and 19	47
25	9 and 23	184
26	24 or 25	184
27	limit 26 to (english or german) EBM Reviews - Cochrane Central Register of Controlled Trials <November 2016> Embase <1974 to 2017 February 01> Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R) (EMBASE)	181 36 133 12
28	baricitinib/	273
29	Baricitinib.mp.	352
30	Olumiant.mp.	0
31	("INCB 028050" or INCB 28050 or INCB028050 or INCB28050).mp.	97
32	(LY 3009104 or LY3009104).mp.	70
33	29 or 30 or 31 or 32	383
34	exp rheumatoid arthritis/	287.793
35	(Rheumat* adj3 Arthrit*).mp.	321.643
36	34 or 35	335.447
37	33 and 36	277
38	random*.tw.	2.590.354
39	placebo*.mp.	777.784
40	double-blind*.tw.	468.253
41	38 or 39 or 40	3.003.000
42	37 and 41	171
43	42 not Medline.cr.	169

Datenbankname	Duplikateliminierung - Zusammengefasste Suche in mehreren Datenbanken (a) EBM Reviews - Cochrane Central Register of Controlled Trials November 2016 (Cochrane) (b) Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R) (MEDLINE) (c) Embase 1974 to 2017 February 01 (EMBASE)	
Suchoberfläche	Ovid	
Datum der Suche	02.02.2017	
Zeitsegment	1946 bis Datum der Suche	
Suchfilter	Duplikateliminierung in den Datenbanken Cochrane, Medline und EMBASE	
44	limit 43 to (english or german) EBM Reviews - Cochrane Central Register of Controlled Trials <November 2016> Embase <1974 to 2017 February 01> Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R)	168 36 120 12
45	9 use cctr EBM Reviews - Cochrane Central Register of Controlled Trials <November 2016>	38 38
46	27 use ppezv Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R)	12 12
47	44 use oomezd Embase <1974 to 2017 February 01>	120 120
48	45 or 46 or 47 EBM Reviews - Cochrane Central Register of Controlled Trials <November 2016> Embase <1974 to 2017 February 01> Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R)	170 38 120 12
49	remove duplicates from 48 EBM Reviews - Cochrane Central Register of Controlled Trials <November 2016> Embase <1974 to 2017 February 01> Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R)	137 33 93 11
50	45 and 49 EBM Reviews - Cochrane Central Register of Controlled Trials <November 2016>	33 33
51	46 and 49 Ovid MEDLINE(R) Epub Ahead of Print, In-Process & Other Non-Indexed Citations, Ovid MEDLINE(R) Daily, Ovid MEDLINE and Versions(R)	11 11
52	47 and 49 Embase <1974 to 2017 February 01>	93 93

Anhang 4-A2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend.

Anhang 4-A3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend.

Anhang 4-A4: Suche nach weiteren Untersuchungen

Nicht zutreffend.

Anhang 4-B: Suchstrategien – Suche in Studienregistern

Geben Sie nachfolgend die Suchstrategien für die Suche(n) in Studienregistern an. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Für jedes durchsuchte Studienregister ist eine separate Strategie darzustellen. 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. 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

Anhang 4-B1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Es wurde eine Suchstrategie für die Studienregisterrecherche für die Suche nach RCT mit dem zu bewertenden Arzneimittel für die Subpopulationen A1, A2 sowie A3 angewendet. Deshalb wird an dieser Stelle auf eine Mehrfachdarstellung der Suchstrategie für die verschiedenen Populationen verzichtet.

Studienregister	<i>ClinicalTrials.gov</i> (CT.gov)
Internetadresse	https://clinicaltrials.gov
Datum der Suche	02.02.2017
Suchstrategie	(BARICITINIB OR LY3009104 OR INCB028050 OR OLUMIANT) [Interventions] AND ("Phase 2" OR "Phase 3" OR "Phase 4") [Phase]
Treffer	15

Studienregister	<i>EU Clinical Trials Register</i> (EUCTR)
Internetadresse	http://www.clinicaltrialsregister.eu
Datum der Suche	02.02.2017
Suchstrategie	(BARICITINIB OR LY3009104 OR INCB028050 OR OLUMIANT) [<i>Suchschlitz</i>] AND ("Phase Two" OR "Phase Three" OR "Phase Four") [<i>Select Trial Phase</i>]
Treffer	8

Studienregister	<i>ICTRP Search Portal (WHO) (ICTRP)</i>
Internetadresse	http://apps.who.int/trialsearch
Datum der Suche	02.02.2017
Suchstrategie	(BARICITINIB OR LY3009104 OR INCB028050 OR OLUMIANT) [Suchschlitz] AND ("Phase 2" OR "Phase 3" OR "Phase 4") [Phases are]
Treffer	41

Studienregister	<i>PharmNet.Bund</i>
Internetadresse	https://www.pharmnet-bund.de/dynamic/de/klinische-pruefungen/index.html
Datum der Suche	02.02.2017
Suchstrategie	Suche 1: (BARICITINIB OR LY3009104 OR INCB028050 OR OLUMIANT) [Textfelder] AND ("Phase II" OR "Phase III" OR "Phase IV") [Trial Phase] Suche 2: (BARICITINIB OR LY3009104 OR INCB028050 OR OLUMIANT) [Title] AND ("Phase II" OR "Phase III" OR "Phase IV") [Trial Phase] Suche 3: (BARICITINIB OR LY3009104 OR INCB028050 OR OLUMIANT) [Active Substance] AND ("Phase II" OR "Phase III" OR "Phase IV") [Trial Phase] Suche 4: (BARICITINIB OR LY3009104 OR INCB028050 OR OLUMIANT) [Product name/code] AND ("Phase II" OR "Phase III" OR "Phase IV") [Trial Phase]
Treffer	4 (4 gleiche Treffer in allen 4 Suchen)

Anhang 4-B2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend.

Anhang 4-B3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend.

Anhang 4-B4: Suche nach weiteren Untersuchungen

Nicht zutreffend.

Anhang 4-C: Liste der im Volltext gesichteten und ausgeschlossenen Dokumente mit Ausschlussgrund (bibliografische Literaturrecherche)

Listen Sie nachfolgend die im Volltext gesichteten und ausgeschlossenen Dokumente aus der / den bibliografischen Literaturrecherche(n) auf. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard) und nummerieren Sie die Zitate fortlaufend. Geben Sie jeweils einen Ausschlussgrund an und beziehen Sie sich dabei auf die im Abschnitt 4.2.2 genannten Ein- und Ausschlusskriterien.

Anhang 4-C1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Subpopulation A1

Nr.	Titel (Autor) der ausgeschlossenen Dokumente	Ausschlussgrund
	Nicht zutreffend.	

Subpopulation A2

Nr.	Titel (Autor) der ausgeschlossenen Dokumente	Ausschlussgrund
	Nicht zutreffend.	

Subpopulation A3

Nr.	Titel (Autor) der ausgeschlossenen Dokumente	Ausschlussgrund
	Nicht zutreffend.	

Für die Subpopulationen A1, A2 und A3 mussten keine Volltexte gesichtet werden, da bereits bei der Bewertung des Titels bzw. des Abstracts keine relevanten Publikationen identifiziert wurden.

Anhang 4-C2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend.

Anhang 4-C3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend.

Anhang 4-C4: Suche nach weiteren Untersuchungen

Nicht zutreffend.

Anhang 4-D: Liste der ausgeschlossenen Studien mit Ausschlussgrund (Suche in Studienregistern)

Listen Sie nachfolgend die durch die Studienregistersuche(n) identifizierten, aber ausgeschlossenen Studien auf. Machen Sie die Angaben getrennt für die einzelnen Recherchen (Suche nach RCT mit dem zu bewertenden Arzneimittel, Suche nach RCT für indirekte Vergleiche etc.) wie unten angegeben. Verwenden Sie hierzu einen allgemein gebräuchlichen Zitierstil (z. B. Vancouver oder Harvard) und nummerieren Sie die Zitate fortlaufend. Geben Sie jeweils einen Ausschlussgrund an und beziehen Sie sich dabei auf die im Abschnitt 4.2.2 genannten Ein- und Ausschlusskriterien.

Anhang 4-D1: Suche nach RCT mit dem zu bewertenden Arzneimittel

Subpopulation A1

Nr.	Referenz der ausgeschlossenen Studien	Ausschlussgrund
<i>ClinicalTrials.gov^a</i>		
1	Eli Lilly and Company. 29-9-2014. ClinicalTrials.gov: A Moderate to Severe Rheumatoid Arthritis Study. https://ClinicalTrials.gov/show/NCT01721044	Vergleichstherapie
2	Eli Lilly and Company. 24-1-2014. ClinicalTrials.gov: Oral JAK1/JAK2 Selective Inhibitor Treatment in Japanese Participants With Active Rheumatoid Arthritis on Background Methotrexate Therapy. https://ClinicalTrials.gov/show/NCT01469013	Vergleichstherapie
3	Eli Lilly and Company. 19-2-2015. ClinicalTrials.gov: A Study in Moderate to Severe Rheumatoid Arthritis Participants. https://ClinicalTrials.gov/show/NCT01721057	Vergleichstherapie
4	Eli Lilly and Company. 24-9-2015. ClinicalTrials.gov: A Study in Participants With Moderate to Severe Rheumatoid Arthritis. https://ClinicalTrials.gov/show/NCT01711359	Population
5	Eli Lilly and Company. 13-5-2016. ClinicalTrials.gov: A Study in Moderate to Severe Rheumatoid Arthritis. https://ClinicalTrials.gov/show/NCT01710358	Vergleichstherapie
6	Eli Lilly and Company. 12-10-2016. ClinicalTrials.gov: A Study of Baricitinib (LY3009104) in Participants With Moderate-to-Severe Atopic Dermatitis. https://ClinicalTrials.gov/show/NCT02576938	Population
7	Eli Lilly and Company. 1-7-2016. ClinicalTrials.gov: A Study of Baricitinib (LY3009104) in Participants With Rheumatoid Arthritis (RA). https://ClinicalTrials.gov/show/NCT02265705	Vergleichstherapie
8	Eli Lilly and Company. 16-11-2016. ClinicalTrials.gov: A Study of Baricitinib (LY3009104) in Participants With Systemic Lupus Erythematosus (SLE). https://ClinicalTrials.gov/show/NCT02708095	Population
9	Eli Lilly and Company. 16-11-2016. ClinicalTrials.gov: An Extension Study in Participants With Moderate to Severe Rheumatoid Arthritis. https://ClinicalTrials.gov/show/NCT01885078	Vergleichstherapie
10	Eli Lilly and Company Incyte Corporation. 25-11-2014. ClinicalTrials.gov: A Phase 2b Study of LY3009104 in Participants With Moderate to Severe Psoriasis. https://ClinicalTrials.gov/show/NCT01490632	Population

Nr.	Referenz der ausgeschlossenen Studien	Ausschlussgrund
11	Eli Lilly and Company Incyte Corporation. 9-4-2014. ClinicalTrials.gov: A Study in Participants With Rheumatoid Arthritis on Background Methotrexate Therapy. https://ClinicalTrials.gov/show/NCT01185353	Vergleichstherapie
12	Eli Lilly and Company Incyte Corporation. 16-1-2015. ClinicalTrials.gov: A Study to Test Safety and Efficacy of Baricitinib in Participants With Diabetic Kidney Disease. https://ClinicalTrials.gov/show/NCT01683409	Population
13	Incyte Corporation. 20-1-2012. ClinicalTrials.gov: INCB028050 Compared to Background Therapy in Patients With Active Rheumatoid Arthritis (RA) With Inadequate Response to Disease Modifying Anti-Rheumatic Drugs. https://ClinicalTrials.gov/show/NCT00902486	Vergleichstherapie
14	Matthew J Koster Eli Lilly and Company Mayo Clinic. 19-1-2017. ClinicalTrials.gov: Baricitinib in Relapsing Giant Cell Arteritis. https://ClinicalTrials.gov/show/NCT03026504	Population
15	National Cancer Institute (NCI) National Institutes of Health Clinical Center (CC). 18-11-2016. ClinicalTrials.gov: Study of Baricitinib, a JAK1/2 Inhibitor, in Chronic Graft-Versus-Host Disease After Allogeneic Hematopoietic Stem Cell Transplantation. https://ClinicalTrials.gov/show/NCT02759731	Population
<i>EU Clinical Trials Register^a</i>		
-		
<i>ICTRP Search Portal (WHO)^a</i>		
16	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Drug Drug Interaction (DDI) Study of Baricitinib (LY3009104) and Digoxin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01859078	Population
17	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Pharmacokinetic Study of Baricitinib in Participants With Liver Disease. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01870388	Population
18	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Relative Bioavailability and Food Effect Study of New Formulations. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01398475	Population
19	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Birth Control Pills in Healthy Females. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01896726	Populatio
20	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Ciclosporin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01968057	Population
21	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Omeprazole in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01925144	Population
22	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Probenecid in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01937026	Population
23	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Rifampicin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01910311	Population

Nr.	Referenz der ausgeschlossenen Studien	Ausschlussgrund
24	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Simvastatin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01960140	Population
25	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib in Healthy Japanese Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT02263911	Population
26	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib When Administered With Ketoconazole or Fluconazole in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01924299	Population
27	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of LY3009104 for Healthy Subjects. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01247350	Population
28	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of LY3009104 in Healthy Volunteers. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01536951	Population
29	Eli Lilly and Company. 9-3-2015. WHO ICTRP: A Study to Measure Baricitinib (LY3009104) Absorption in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT02340104	Population
30	Eli Lilly and Company. 19-2-2015. WHO ICTRP: Disposition of ¹⁴ C-LY3009104 Following Oral Administration in Healthy Human Subjects. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01299285	Population
31	Eli Lilly and Company. 5-9-2016. WHO ICTRP: A Study of LY3009104 in Healthy Chinese Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT02758613	Population
32	Eli Lilly and Company. 11-4-2016. WHO ICTRP: Compassionate Use Protocol for the Treatment of Autoinflammatory Syndromes. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01724580	Population
33	Eli Lilly Japan, K.K. 3-6-2014. WHO ICTRP: A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Dose-Ranging, Phase 2 Study to Evaluate the Safety and Renal Efficacy of LY3009104 in Patients with Diabetic Kidney Disease. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=JPRN-JapicCTI-122022	Population
34	Eli Lilly Japan, K.K. 20-6-2016. WHO ICTRP: A Randomized, Double-Blind, Placebo-Controlled, Dose Ranging, Parallel-Group, Phase 2 Study of LY3009104 in Japanese Patients with Active Rheumatoid Arthritis on Background Methotrexate Therapy. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=JPRN-JapicCTI-111709	Vergleichstherapie
35	Eli Lilly Japan, K.K. 20-6-2016. WHO ICTRP: A Randomized, Double-Blind, Placebo-Controlled, Dose-Ranging, Phase 2b Study of LY3009104 in Patients with Moderate-to-Severe Plaque Psoriasis. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=JPRN-JapicCTI-121871	Population

Nr.	Referenz der ausgeschlossenen Studien	Ausschlussgrund
	<i>PharmNet.Bund</i> ^a	
	-	
	a: Studien, die in mehr als einem Register identifiziert wurden, werden einmalig aufgeführt. Bedeutung des jeweils aufgelisteten Datums: <i>ClinicalTrials.gov</i> : Last updated <i>EU Clinical Trials Register</i> : Start Date <i>ICTRP Search Portal (WHO)</i> : Last refreshed on <i>PharmNet.Bund</i> : Bescheiddatum Bundesbehörde	

Subpopulation A2

Nr.	Referenz der ausgeschlossenen Studien	Ausschlussgrund
<i>ClinicalTrials.gov^a</i>		
1	Eli Lilly and Company. 29-9-2014. ClinicalTrials.gov: A Moderate to Severe Rheumatoid Arthritis Study. https://ClinicalTrials.gov/show/NCT01721044	Vergleichstherapie
2	Eli Lilly and Company. 24-1-2014. ClinicalTrials.gov: Oral JAK1/JAK2 Selective Inhibitor Treatment in Japanese Participants With Active Rheumatoid Arthritis on Background Methotrexate Therapy. https://ClinicalTrials.gov/show/NCT01469013	Vergleichstherapie
3	Eli Lilly and Company. 19-2-2015. ClinicalTrials.gov: A Study in Moderate to Severe Rheumatoid Arthritis Participants. https://ClinicalTrials.gov/show/NCT01721057	Vergleichstherapie
4	Eli Lilly and Company. 24-9-2015. ClinicalTrials.gov: A Study in Participants With Moderate to Severe Rheumatoid Arthritis. https://ClinicalTrials.gov/show/NCT01711359	Population
5	Eli Lilly and Company. 12-10-2016. ClinicalTrials.gov: A Study of Baricitinib (LY3009104) in Participants With Moderate-to-Severe Atopic Dermatitis. https://ClinicalTrials.gov/show/NCT02576938	Population
6	Eli Lilly and Company. 1-7-2016. ClinicalTrials.gov: A Study of Baricitinib (LY3009104) in Participants With Rheumatoid Arthritis (RA). https://ClinicalTrials.gov/show/NCT02265705	Vergleichstherapie
7	Eli Lilly and Company. 16-11-2016. ClinicalTrials.gov: A Study of Baricitinib (LY3009104) in Participants With Systemic Lupus Erythematosus (SLE). https://ClinicalTrials.gov/show/NCT02708095	Population
8	Eli Lilly and Company. 16-11-2016. ClinicalTrials.gov: An Extension Study in Participants With Moderate to Severe Rheumatoid Arthritis. https://ClinicalTrials.gov/show/NCT01885078	Vergleichstherapie
9	Eli Lilly and Company Incyte Corporation. 25-11-2014. ClinicalTrials.gov: A Phase 2b Study of LY3009104 in Participants With Moderate to Severe Psoriasis. https://ClinicalTrials.gov/show/NCT01490632	Population
10	Eli Lilly and Company Incyte Corporation. 9-4-2014. ClinicalTrials.gov: A Study in Participants With Rheumatoid Arthritis on Background Methotrexate Therapy. https://ClinicalTrials.gov/show/NCT01185353	Vergleichstherapie
11	Eli Lilly and Company Incyte Corporation. 16-1-2015. ClinicalTrials.gov: A Study to Test Safety and Efficacy of Baricitinib in Participants With Diabetic Kidney Disease. https://ClinicalTrials.gov/show/NCT01683409	Population
12	Incyte Corporation. 20-1-2012. ClinicalTrials.gov: INCB028050 Compared to Background Therapy in Patients With Active Rheumatoid Arthritis (RA) With Inadequate Response to Disease Modifying Anti-Rheumatic Drugs. https://ClinicalTrials.gov/show/NCT00902486	Vergleichstherapie
13	Matthew J Koster Eli Lilly and Company Mayo Clinic. 19-1-2017. ClinicalTrials.gov: Baricitinib in Relapsing Giant Cell Arteritis. https://ClinicalTrials.gov/show/NCT03026504	Population

Nr.	Referenz der ausgeschlossenen Studien	Ausschlussgrund
14	National Cancer Institute (NCI) National Institutes of Health Clinical Center (CC). 18-11-2016. ClinicalTrials.gov: Study of Baricitinib, a JAK1/2 Inhibitor, in Chronic Graft-Versus-Host Disease After Allogeneic Hematopoietic Stem Cell Transplantation. https://ClinicalTrials.gov/show/NCT02759731	Population
<i>EU Clinical Trials Register⁴</i>		
-		
<i>ICTRP Search Portal (WHO)</i>		
15	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Drug Drug Interaction (DDI) Study of Baricitinib (LY3009104) and Digoxin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01859078	Population
16	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Pharmacokinetic Study of Baricitinib in Participants With Liver Disease. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01870388	Population
17	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Relative Bioavailability and Food Effect Study of New Formulations. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01398475	Population
18	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Birth Control Pills in Healthy Females. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01896726	Population
19	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Ciclosporin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01968057	Population
20	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Omeprazole in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01925144	Population
21	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Probenecid in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01937026	Population
22	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Rifampicin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01910311	Population
23	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Simvastatin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01960140	Population
24	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib in Healthy Japanese Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT02263911	Population
25	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib When Administered With Ketoconazole or Fluconazole in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01924299	Population
26	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of LY3009104 for Healthy Subjects. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01247350	Population
27	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of LY3009104 in Healthy Volunteers. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01536951	Population

Nr.	Referenz der ausgeschlossenen Studien	Ausschlussgrund
28	Eli Lilly and Company. 9-3-2015. WHO ICTRP: A Study to Measure Baricitinib (LY3009104) Absorption in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT02340104	Population
29	Eli Lilly and Company. 19-2-2015. WHO ICTRP: Disposition of 14C-LY3009104 Following Oral Administration in Healthy Human Subjects. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01299285	Population
30	Eli Lilly and Company. 5-9-2016. WHO ICTRP: A Study of LY3009104 in Healthy Chinese Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT02758613	Population
31	Eli Lilly and Company. 11-4-2016. WHO ICTRP: Compassionate Use Protocol for the Treatment of Autoinflammatory Syndromes. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01724580	Population
32	Eli Lilly Japan, K.K. 3-6-2014. WHO ICTRP: A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Dose-Ranging, Phase 2 Study to Evaluate the Safety and Renal Efficacy of LY3009104 in Patients with Diabetic Kidney Disease. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=JPRN-JapicCTI-122022	Population
33	Eli Lilly Japan, K.K. 20-6-2016. WHO ICTRP: A Randomized, Double-Blind, Placebo-Controlled, Dose Ranging, Parallel-Group, Phase 2 Study of LY3009104 in Japanese Patients with Active Rheumatoid Arthritis on Background Methotrexate Therapy. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=JPRN-JapicCTI-111709	Vergleichstherapie
34	Eli Lilly Japan, K.K. 20-6-2016. WHO ICTRP: A Randomized, Double-Blind, Placebo-Controlled, Dose-Ranging, Phase 2b Study of LY3009104 in Patients with Moderate-to-Severe Plaque Psoriasis. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=JPRN-JapicCTI-121871	Population
<i>PharmNet.Bund^a</i>		
-		
<p>Erläuterungen:</p> <p>a: Studien, die in mehr als einem Register identifiziert wurden, werden einmalig aufgeführt.</p> <p>Bedeutung des jeweils aufgelisteten Datums:</p> <p><i>ClinicalTrials.gov</i>: Last updated</p> <p><i>EU Clinical Trials Register</i>: Start Date</p> <p><i>ICTRP Search Portal (WHO)</i>: Last refreshed on</p> <p><i>PharmNet.Bund</i>: Bescheiddatum Bundesbehörde</p>		

Subpopulation A3

Nr.	Referenz der ausgeschlossenen Studien	Ausschlussgrund
<i>ClinicalTrials.gov^a</i>		
1	Eli Lilly and Company. 29-9-2014. ClinicalTrials.gov: A Moderate to Severe Rheumatoid Arthritis Study. https://ClinicalTrials.gov/show/NCT01721044	Vergleichstherapie
2	Eli Lilly and Company. 24-1-2014. ClinicalTrials.gov: Oral JAK1/JAK2 Selective Inhibitor Treatment in Japanese Participants With Active Rheumatoid Arthritis on Background Methotrexate Therapy. https://ClinicalTrials.gov/show/NCT01469013	Vergleichstherapie
3	Eli Lilly and Company. 19-2-2015. ClinicalTrials.gov: A Study in Moderate to Severe Rheumatoid Arthritis Participants. https://ClinicalTrials.gov/show/NCT01721057	Vergleichstherapie
4	Eli Lilly and Company. 24-9-2015. ClinicalTrials.gov: A Study in Participants With Moderate to Severe Rheumatoid Arthritis. https://ClinicalTrials.gov/show/NCT01711359	Population
5	Eli Lilly and Company. 12-10-2016. ClinicalTrials.gov: A Study of Baricitinib (LY3009104) in Participants With Moderate-to-Severe Atopic Dermatitis. https://ClinicalTrials.gov/show/NCT02576938	Population
6	Eli Lilly and Company. 1-7-2016. ClinicalTrials.gov: A Study of Baricitinib (LY3009104) in Participants With Rheumatoid Arthritis (RA). https://ClinicalTrials.gov/show/NCT02265705	Vergleichstherapie
7	Eli Lilly and Company. 16-11-2016. ClinicalTrials.gov: A Study of Baricitinib (LY3009104) in Participants With Systemic Lupus Erythematosus (SLE). https://ClinicalTrials.gov/show/NCT02708095	Population
8	Eli Lilly and Company. 16-11-2016. ClinicalTrials.gov: An Extension Study in Participants With Moderate to Severe Rheumatoid Arthritis. https://ClinicalTrials.gov/show/NCT01885078	Vergleichstherapie
9	Eli Lilly and Company Incyte Corporation. 25-11-2014. ClinicalTrials.gov: A Phase 2b Study of LY3009104 in Participants With Moderate to Severe Psoriasis. https://ClinicalTrials.gov/show/NCT01490632	Population
10	Eli Lilly and Company Incyte Corporation. 9-4-2014. ClinicalTrials.gov: A Study in Participants With Rheumatoid Arthritis on Background Methotrexate Therapy. https://ClinicalTrials.gov/show/NCT01185353	Vergleichstherapie
11	Eli Lilly and Company Incyte Corporation. 16-1-2015. ClinicalTrials.gov: A Study to Test Safety and Efficacy of Baricitinib in Participants With Diabetic Kidney Disease. https://ClinicalTrials.gov/show/NCT01683409	Population
12	Incyte Corporation. 20-1-2012. ClinicalTrials.gov: INCB028050 Compared to Background Therapy in Patients With Active Rheumatoid Arthritis (RA) With Inadequate Response to Disease Modifying Anti-Rheumatic Drugs. https://ClinicalTrials.gov/show/NCT00902486	Vergleichstherapie
13	Matthew J Koster Eli Lilly and Company Mayo Clinic. 19-1-2017. ClinicalTrials.gov: Baricitinib in Relapsing Giant Cell Arteritis. https://ClinicalTrials.gov/show/NCT03026504	Population

Nr.	Referenz der ausgeschlossenen Studien	Ausschlussgrund
14	National Cancer Institute (NCI) National Institutes of Health Clinical Center (CC). 18-11-2016. ClinicalTrials.gov: Study of Baricitinib, a JAK1/2 Inhibitor, in Chronic Graft-Versus-Host Disease After Allogeneic Hematopoietic Stem Cell Transplantation. https://ClinicalTrials.gov/show/NCT02759731	Population
<i>EU Clinical Trials Register^a</i>		
-		
<i>ICTRP Search Portal (WHO)^a</i>		
15	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Drug Drug Interaction (DDI) Study of Baricitinib (LY3009104) and Digoxin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01859078	Population
16	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Pharmacokinetic Study of Baricitinib in Participants With Liver Disease. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01870388	Population
17	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Relative Bioavailability and Food Effect Study of New Formulations. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01398475	Population
18	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Birth Control Pills in Healthy Females. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01896726	Population
19	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Ciclosporin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01968057	Population
20	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Omeprazole in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01925144	Population
21	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Probenecid in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01937026	Population
22	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Rifampicin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01910311	Population
23	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib and Simvastatin in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01960140	Population
24	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib in Healthy Japanese Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT02263911	Population
25	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of Baricitinib When Administered With Ketoconazole or Fluconazole in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01924299	Population
26	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of LY3009104 for Healthy Subjects. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01247350	Population
27	Eli Lilly and Company. 19-2-2015. WHO ICTRP: A Study of LY3009104 in Healthy Volunteers. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01536951	Population

Nr.	Referenz der ausgeschlossenen Studien	Ausschlussgrund
28	Eli Lilly and Company. 9-3-2015. WHO ICTRP: A Study to Measure Baricitinib (LY3009104) Absorption in Healthy Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT02340104	Population
29	Eli Lilly and Company. 19-2-2015. WHO ICTRP: Disposition of 14C-LY3009104 Following Oral Administration in Healthy Human Subjects. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01299285	Population
30	Eli Lilly and Company. 5-9-2016. WHO ICTRP: A Study of LY3009104 in Healthy Chinese Participants. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT02758613	Population
31	Eli Lilly and Company. 11-4-2016. WHO ICTRP: Compassionate Use Protocol for the Treatment of Autoinflammatory Syndromes. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=NCT01724580	Population
32	Eli Lilly Japan, K.K. 3-6-2014. WHO ICTRP: A Multicenter, Randomized, Double-Blind, Placebo-Controlled, Dose-Ranging, Phase 2 Study to Evaluate the Safety and Renal Efficacy of LY3009104 in Patients with Diabetic Kidney Disease. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=JPRN-JapicCTI-122022	Population
33	Eli Lilly Japan, K.K. 20-6-2016. WHO ICTRP: A Randomized, Double-Blind, Placebo-Controlled, Dose Ranging, Parallel-Group, Phase 2 Study of LY3009104 in Japanese Patients with Active Rheumatoid Arthritis on Background Methotrexate Therapy. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=JPRN-JapicCTI-111709	Vergleichstherapie
34	Eli Lilly Japan, K.K. 20-6-2016. WHO ICTRP: A Randomized, Double-Blind, Placebo-Controlled, Dose-Ranging, Phase 2b Study of LY3009104 in Patients with Moderate-to-Severe Plaque Psoriasis. http://apps.who.int/trialsearch/Trial2.aspx?TrialID=JPRN-JapicCTI-121871	Population
<i>PharmNet.Bund^a</i>		
-		
<p>Erläuterungen:</p> <p>a: Studien, die in mehr als einem Register identifiziert wurden, werden einmalig aufgeführt.</p> <p>Bedeutung des jeweils aufgelisteten Datums:</p> <p><i>ClinicalTrials.gov</i>: Last updated</p> <p><i>EU Clinical Trials Register</i>: Start Date</p> <p><i>ICTRP Search Portal (WHO)</i>: Last refreshed on</p> <p><i>PharmNet.Bund</i>: Bescheiddatum Bundesbehörde</p>		

Anhang 4-D2: Suche nach RCT für indirekte Vergleiche

Nicht zutreffend.

Anhang 4-D3: Suche nach nicht randomisierten vergleichenden Studien

Nicht zutreffend.

Anhang 4-D4: Suche nach weiteren Untersuchungen

Nicht zutreffend.

Anhang 4-E: Methodik der eingeschlossenen Studien – RCT

Beschreiben Sie nachfolgend die Methodik jeder eingeschlossenen, in Abschnitt 4.3.1.1.4 genannten Studie. Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle 4-177 inklusive eines Flow-Charts für dem Patientenfluss.

Sollten Sie im Dossier indirekte Vergleiche präsentieren, beschreiben Sie ebenfalls die Methodik jeder zusätzlich in den indirekten Vergleich eingeschlossenen Studie (Abschnitt 4.3.2.1). Erstellen Sie hierfür je Studie eine separate Version der nachfolgend dargestellten Tabelle 4-177 inklusive eines Flow-Charts für dem Patientenfluss

Tabelle 4-177: (Anhang) Studiendesign und -methodik für Studie RA-BEAM (JADV)

Item ^a	Charakteristikum	Studieninformation
-	Studienziel	
2b	Genaue Ziele, Fragestellung und Hypothesen	Überlegenheit von Baricitinib gegen Placebo bei Patienten mit moderater bis schwerer aktiver RA, die ein unzureichendes Ansprechen auf MTX (MTX-IR) zeigten, gemessen am Erreichen einer mindestens 20%-igen Verbesserung auf der ACR-Skala (ACR20) nach 12 Wo.
-	Methoden	
3	Studiendesign	
3a	Beschreibung des Studiendesigns (z. B. parallel, faktoriell) inklusive Zuteilungsverhältnis	<p>Allokation: randomisiert Zuteilungsverhältnis: 3:3:2 Behandlung 3 parallele Behandlungsarme:</p> <ul style="list-style-type: none"> • Baricitinib: (4 mg einmal täglich), für Patienten mit eingeschränkter Nierenfunktion 2 mg einmal täglich; • Adalimumab: 40 mg s.c. alle zwei Wochen; • Placebo: einmal täglich. <p>Follow-up 4 Wo. nach letzter Dosis (falls kein Übergang in Langzeit-Verlängerungsstudie RA-BEYOND/JADY) Randomisierungseinheit: Studienteilnehmer Verblindung: doppelblind Behandlungsdauer: 52 Wo. Design: 3-armig, parallel Studienorganisation: multizentrisch Phase: III Behandlung: ambulant.</p>
3b	Relevante Änderungen der Methodik nach Studienbeginn (z. B. Ein-/Ausschlusskriterien), mit Begründung	<p>Protokolländerungen Das JADV Protokoll wurde im Juli 2012 durch Lilly genehmigt und wurde dreimal geändert.</p> <ul style="list-style-type: none"> • Die Prüfplanänderung (a) wurde von Lilly initiiert, um Korrekturen und Klarstellungen vorzunehmen (Genehmigung durch Lilly im

Item ^a	Charakteristikum	Studieninformation
		<p>Oktober 2012).</p> <ul style="list-style-type: none"> • Die Prüfplanänderung (b) wurde nach Aufforderung durch Europäische Arzneimittelbehörden vorgenommen, um die geschlechtsspezifischen Unterschiede bezüglich der EKG Abnormalitäten zu korrigieren, die Patienten von der Studie ausgeschlossen hätten (Genehmigung durch Lilly im November 2012). • Die Prüfplanänderung (c) wurde von Lilly initiiert, um weitere Korrekturen und Klarstellungen durchzuführen (Genehmigung durch Lilly im Juli 2013). <p>Veränderungen gegenüber den vorgeplanten Analysen des SAP</p> <p>Version 1 des SAP wurde am 10. August 2013 genehmigt, bevor das externe DMC erstmals entblindet wurde. Version 2 des SAP wurde am 17. April 2015 genehmigt, bevor die Entblindung für die primäre Analyse erfolgte. Die folgenden Änderungen wurden an den geplanten Analysen innerhalb der durchgeführten Analysen vorgenommen. Für Details sei auf den Studienbericht verwiesen.</p>
4	Probanden / Patienten	<p>Voraussetzung für die Studienpopulation</p> <p>Erwachsene Patienten mit moderater bis schwerer aktiver RA, mit unzureichendem Ansprechen auf MTX (MTX-IR), die zuvor keine Therapie mit biologischen anti-rheumatischen Basistherapeutika (bDMARDs) erhalten hatten.</p>
4a	Ein-/Ausschlusskriterien der Probanden / Patienten	<p>Einschlusskriterien</p> <p>Bei Screening eingeschlossene Patienten:</p> <ul style="list-style-type: none"> • Mindestalter 18 Jahre. • mit im Erwachsenenalter erworbener RA (nach ACR/EULAR 2010 Kriterien zur Klassifikation von RA). • mit moderater bis schwerer aktiver RA, basierend auf Diagnose von mind. 6/68 druckschmerzhaften und 6/66 geschwollenen Gelenken (operierte Gelenke wurden nicht in TJC und SJC erfasst). • mit CRP (o. hsCRP) Wert ≥ 6 mg/l. • mit MTX-Behandlung entsprechend der klinischen Praxis für mind. 12 Wo. vor Studieneintritt. • MTX-Dosierung: stabile, unveränderte orale Dosis von 7,5 bis 25 mg/Wo. (oder äquivalente injizierbare Dosis) über mind. 8 Wo vor Studieneintritt. Die MTX Dosierung sollte im Studienverlauf stabil bleiben und nur bei Sicherheitsbedenken angepasst werden. <ul style="list-style-type: none"> ○ für Patienten mit MTX Dosierung < 15 mg/Wo.: klare Dokumentation in der

Item ^a	Charakteristikum	Studieninformation
		<p>Patientenakte, dass höhere MTX Dosen nicht toleriert werden oder dass die MTX Dosierung der höchsten akzeptablen Dosierung laut lokaler klinischer Praxis entsprach.</p> <ul style="list-style-type: none"> ○ Einhaltung lokaler Behandlungsstandards bei der begleitenden Gabe von Folsäure. <ul style="list-style-type: none"> ● konnten die schriftliche Einverständniserklärung lesen, verstehen und unterzeichnen. <p>Ausschlusskriterien</p> <p>Bei Screening ausgeschlossene Patienten:</p> <ul style="list-style-type: none"> ● bekommen derzeit Kortikosteroide (Dosierung > 10 mg/Tag Prednison oder Äquivalent) oder hatten eine Änderung der Dosis für Kortikosteroide in den letzten 2 Wo. vor Studieneintritt oder innerhalb der letzten 6 Wo. vor der geplanten Randomisierung. ● Therapiebeginn mit NSAIDs (zum Zweck der Behandlung von Anzeichen und Symptomen der RA) innerhalb der letzten 2 Wo. vor Studieneintritt oder innerhalb der letzten 6 Wo. vor geplanter Randomisierung begonnen ODER erhielten eine Änderung der Dosis für NSAIDs innerhalb von 2 Wo. vor Studieneintritt oder innerhalb von 6 Wo. vor der geplanten Randomisierung. ● bekommen derzeit begleitende Therapie mit MTX, Hydroxychloroquin und Sulfasalazin oder eine andere Kombination drei beliebiger antirheumatoider Basistherapeutika. ● bekommen derzeit oder erhielten andere DMARDs (z. B. Goldsalze, Cyclosporin, Azathiopin, oder andere Immunsuppressiva) als MTX, Hydroxychloroquin (bis zu 400 mg/täglich), oder Sulfasalazin (bis zu 3.000 mg/täglich) innerhalb der letzten 4 Wo. vor Studieneintritt. <ul style="list-style-type: none"> ○ stabile Dosierung von Hydroxychloroquin und Sulfasalazin sollte für mind. 8 Wo. vor Studienbeginn vorliegen, falls die Therapie abgebrochen wurde, durfte keine neue Therapie innerhalb von 4 Wo. vor Studieneintritt gegeben werden. ○ Immunsuppression aufgrund von einem transplantierten Organ war nicht gestattet. ● bekamen Leflunomid in den 12 Wo. vor Studieneintritt (oder unterzogen sich innerhalb der letzten 4 Wo. vor Studieneintritt einem Auswaschverfahren mit Cholestyramin über einen Zeitraum von 11 Tagen). ● hatten eine neue physiotherapeutische Behandlung der RA in den letzten 2 Wo. vor Studienbeginn begonnen.

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • erhielten zu einem früheren Zeitpunkt bereits eine beliebige bDMARD (TNF, Interleukin-1, IL-6, T-Zell- oder B-Zell-gerichtete Therapie) Therapie. • hatten eine Interferontherapie (Roferon-A, Intron-A, Rebetron, Alferon-N, Peg-Intron, Avonex, Betaseron, Infergen, Actimmun, Pegasys) innerhalb der letzten 4 Wo. vor Studieneintritt oder würden während der Studie voraussichtlich eine Interferon Behandlung benötigen. • hatten ein beliebiges parenterales Kortikosteroid durch i. m. oder i. v. Injektionen innerhalb der letzten 2 Wo. vor Studieneintritt oder innerhalb der letzten 6 Wo. vor geplanter Randomisierung erhalten oder es war abzusehen, dass eine parenterale Injektion von Kortikosteroiden während der Studie nötig würde. • wurden an drei oder mehr Gelenken mit intraartikulärem Kortikosteroid oder Hyaluronsäure innerhalb der letzten 2 Wo. vor Studieneintritt oder innerhalb der letzten 6 Wo. vor geplanter Randomisierung behandelt. <ul style="list-style-type: none"> ○ Gelenke mit injiziertem intraartikulärem Kortikosteroid oder Hyaluronsäure innerhalb der letzten 2 Wo. vor Studieneintritt oder innerhalb der letzten 6 Wo. vor geplanter Randomisierung konnten nicht in TJC und SJC erfasst werden. • hatten eine Diagnose oder Kontraindikation, die laut lokaler klinischer Praxis von Adalimumab den Patienten von der Teilnahme an diesem Protokoll ausschließen. • hatten aktive Fibromyalgie, die nach Meinung des Prüfarztes eine angemessene Beurteilung der RA Aktivität für diese Studie erschweren würde. • hatten eine Diagnose für eine beliebige systemisch entzündliche Erkrankung außer der RA, so wie, aber nicht begrenzt auf, juvenile chronische Arthritis, Spondyloarthropathie, Morbus Crohn, ulzerative Kolitis, Psoriasis-Arthritis, aktive Vaskulitis oder Gicht. <ul style="list-style-type: none"> ○ Patienten mit sekundärem Sjögren Syndrom waren nicht ausgeschlossen. • hatten eine Diagnose für das Felty Syndrom. • hatten während der letzten 8 Wo. vor Studieneintritt eine große Operation oder während der Studie war eine große Operation geplant, wodurch nach Meinung des Prüfarztes und in Rücksprache mit dem pU oder eines Beauftragten ein nicht vertretbares Risiko für den Patienten bestand. • hatten einen Myokardinfarkt, instabile ischämische Herzerkrankung, Schlaganfall oder

Item ^a	Charakteristikum	Studieninformation
		<p>NYHA Stufe IV Herzinsuffizienz innerhalb der letzten 12 Wo. vor Studieneintritt.</p> <ul style="list-style-type: none"> • hatten zu einem früheren Zeitpunkt kardiovaskuläre, respiratorische, hepatische, gastrointestinale, endokrine, hämatologische, neurologische oder neuropsychiatrische Erkrankungen oder andere schwerwiegende und/oder un stabile Krankheiten, so dass nach Meinung des Prüf arztes ein Risiko bei Einnahme des zu untersuchenden Produktes bestand oder eine Dateninterpretation beeinträchtigt würde. • waren zum großen Teil oder komplett un selbstständig, beispielsweise bettlägerig oder an einen Rollstuhl gebunden. • hatten eine geschätzte eGFR (basierend auf der aktuellsten Analyse des Serumkreatinins mit der MDRD-Methode) < 40 ml/min/1,73 m². • hatten zu einem früheren Zeitpunkt eine chronische Lebererkrankung mit AST- oder ALT-Werten 1,5-fach über ULN oder Gesamt-Bilirubin Wert 1,5-fach über ULN. • hatten zu einem früheren Zeitpunkt eine lymphoproliferative Erkrankung oder hatten Anzeichen und Symptome, die eine mögliche lymphoproliferative Erkrankung nahelegen, inklusive Lymphadenopathie oder Splenomegalie oder hatten eine aktive primäre oder wiederkehrende maligne Erkrankung oder waren in Remission für eine klinisch signifikante Krebserkrankung für weniger als 5 Jahre. <ul style="list-style-type: none"> ○ Patienten mit Zervixkarzinom (<i>in situ</i>) konnten nach Resektion und Ausschluss von Rückfall oder Metastasenbildung (für mind. 3 Jahre) teilnehmen ○ Patienten mit Basaliom oder Plattenepithelkarzinom konnten nach vollständiger Resektion und Ausschluss von Rückfall (für mind. 3 Jahre) teilnehmen. • hatten in den letzten 12 Wo vor der geplanten Randomisierung oder hätten voraussichtlich während der Studie einen Lebendimpfstoff verabreicht bekommen (ausgenommen Herpes Zoster-Impfung). <ul style="list-style-type: none"> ○ Bei negativem Impfstatus zu Herpes Zoster: Impfung muss > 30 Tage vor der Randomisierung und dem Einnahmebeginn der Studienmedikation erfolgen. Die Patienten, die eine Herpes Zoster-Impfung innerhalb der letzten 30 Tage vor der geplanten Randomisierung verabreicht bekommen haben, wurden ausgeschlossen. ○ Überprüfung des Impfstatus durch den Prüf arzt und Anpassung an die lokalen Leitlinien

Item ^a	Charakteristikum	Studieninformation
		<p data-bbox="919 282 1401 394">zur Impfung von Erwachsenen mit dem Totimpfstoff, um infektiöse Erkrankungen vor dem Studieneintritt der Patienten zu verhindern.</p> <ul style="list-style-type: none"> <li data-bbox="823 416 1342 506">• hatten (< 30 Tage vor Studieneintritt) eine klinisch schwerwiegende virale, bakterielle, fungale oder parasitäre Infektion. <li data-bbox="823 517 1334 607">• hatten eine symptomatische Herpes Zoster-Infektion innerhalb der letzten 12 Wo. vor Studieneintritt. <li data-bbox="823 618 1374 707">• hatten Herpes Zoster disseminatus oder Herpes Zoster mit Komplikationen in ihrer Krankengeschichte. <li data-bbox="823 719 1358 808">• waren immun-supprimiert und würden nach Meinung des Prüfarztes ein nicht vertretbares Risiko bei einer Studienbeteiligung eingehen. <li data-bbox="823 819 1382 931">• hatten zu einem früheren Zeitpunkt eine aktive Hepatitis B oder C, oder humane Immundefizienz Viruserkrankung (HBV, HCV, HIV). <li data-bbox="823 943 1366 1032">• hatten Haushaltskontakt mit einer Person mit aktiver TB und hatten keine angemessene und dokumentierte TB-Vorsorge erhalten. <li data-bbox="823 1043 1390 1133">• Nachweis einer vorangegangenen oder aktuellen aktiven TB, und Patienten erhielten keine angemessene und dokumentierte Behandlung. <li data-bbox="823 1144 1398 1178">• waren schwanger oder stillten bei Studieneintritt. <li data-bbox="823 1189 1390 1469">• waren weiblich, gebärfähig und stimmten nicht zu, für die Studiendauer bis mind. 28 Tage nach der letzten Dosisverabreichung zwei hoch effektive Verhütungsmethoden anzuwenden: <ul style="list-style-type: none"> <li data-bbox="871 1323 1358 1469">○ nicht gebärfähige Frauen waren definiert als: ≥ 60 Jahre, $\geq 40 - < 60$ Jahre, in der Menopause seit mind. 12 Monaten oder angeboren unfruchtbar/chirurgisch sterilisiert. <li data-bbox="871 1480 1382 1659">○ hoch effektive Verhütungsmethoden waren definiert als hormonelle Verhütungsmittel, Kondom plus Spermizid, zervikale Verschlusskappe plus Spermizid, Intrauterinpepparder System, vasektomierter Mann. <li data-bbox="823 1671 1390 1973">• waren männlich und stimmten nicht zu, für die Studiendauer bis mind. 28 Tage nach der letzten oralen Dosisverabreichung zwei hocheffektive Verhütungsmethoden (s. oben) beim Geschlechtsverkehr mit gebärfähigen weiblichen Partnern anzuwenden bzw. für 28 Tage oder für den Zeitraum, der in der lokalen Adalimumab Fachinformation vorgegeben ist (je nachdem, welcher Zeitraum länger war), nach einer s. c. Injektion.

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • hatten > 500 ml Blut gespendet (\leq 30 Tage vor Studieneintritt) oder wollten während der Studie Blut spenden. • hatten eine Historie von chronischem Alkohol- oder i. v. Drogenmissbrauch oder anderem illegalen Drogenmissbrauch innerhalb der letzten zwei Jahre vor Studieneintritt. • wurden vorher in dieser oder einer anderen Studie zu Baricitinib randomisiert. • konnten oder wollten nicht für die Dauer der Studie zur Verfügung stehen und/oder wollten Studienbestimmungen und –abläufe nicht beachten. • wurden bereits vorher mit einem oralen Januskinase-Inhibitor behandelt. • sie selbst oder ihre direkten Familienmitglieder standen in Verbindung zu dem Personal der Studie. • waren Mitarbeiter bei Lilly oder Incyte oder einem Beauftragten dieser Unternehmen. • waren in einer klinischen Studie eingeschrieben oder hatten innerhalb der letzten 30 Tage vor Studienbeginn eine klinische Studie abgebrochen, die nicht mit der vorliegenden Studie kompatibel war. <p>Einschlusskriterien bei Randomisierung</p> <p>Bei Screening eingeschlossene Patienten mussten zum Zeitpunkt der Randomisierung zusätzlich zu den oben genannten Kriterien folgende Einschlusskriterien erfüllen:</p> <ul style="list-style-type: none"> • moderate bis schwere aktive RA, basierend auf Diagnose von mind. 6/68 empfindlichen und 6/66 geschwollenen Gelenken zu <i>Visit 1</i> und 2. • hsCRP Wert \geq6 mg/l zu <i>Visit 1</i> (der Test wurde in einem zentralen Labor durchgeführt). Die Bestimmung des hsCRP-Werten konnte im Zeitraum von ca. 2 Wo. nach der ersten Messung einmal wiederholt werden, und der Wiederholungswert konnte für die Aufnahme in die Studie verwendet werden, sofern er den Einschlusskriterien entsprach. • mind. eine Gelenkerosion an den Hand-, Handwurzel- oder Fußgelenken, basierend auf der radiographischen Interpretation des Hauptuntersuchenden UND RF- oder ACPA-positiv; ODER mind. drei Gelenkerosionen an den Hand-, Handwurzel- oder Fußgelenken, basierend auf der radiographischen Interpretation des Hauptuntersuchenden unabhängig vom RF- oder ACPA-Status (Röntgenbilder konnten innerhalb von 4 Wo. nach dem Studieneintritt dem Hauptuntersuchenden für die Eignungsbestätigung einge-

Item ^a	Charakteristikum	Studieninformation
		<p>reicht werden). Falls der Patient einem <i>Rescreening</i> unterzogen wurde, konnten die Röntgenbilder des ersten <i>Screening</i> bzw. die innerhalb von 10 Wo. nach der Randomisierung erstellten Röntgenbilder genutzt werden.</p> <p>Ausschlusskriterien bei Randomisierung</p> <p>Bei Screening eingeschlossene Patienten waren für den Studieneinschluss ungeeignet, falls ein oder mehrere der folgenden Kriterien zutrafen:</p> <ul style="list-style-type: none"> • Laborwerte, inklusive TSH, außerhalb des Referenzbereichs, so dass nach Meinung des Prüfarztes ein nicht vertretbares Risiko für die Teilnahme des Patienten an der Studie bestand. <ul style="list-style-type: none"> ○ Patienten, die eine ≥ 12 Wo. dauernde stabile Thyroxin-Substitutionstherapie erhalten haben und TSH Werte innerhalb des Referenzbereiches aufwiesen, konnten an der Studie teilnehmen. • Die Bestimmung des TSH-Wertes konnte im Zeitraum von ca. 2 Wo. nach der ersten Messung einmal wiederholt werden, und der Wiederholungswert konnte für die Aufnahme in die Studie genutzt werden, sofern er den Einschlusskriterien entsprach. • spezifische Abweichungen in den Laborwerten <ul style="list-style-type: none"> ○ AST oder ALT $> 1,5$-fach ULN ○ Gesamtbilirubin $\geq 1,5$-fach ULN ○ Hämoglobin $< 10,0$ g/dl ○ Gesamtzahl an weißen Blutzellen WBC: $< 2,500$ Zellen/μl ○ Neutropenie (Gesamtzahl an Zellen < 1200 Zellen/μl) ○ Lymphopenie (Gesamtzahl an Zellen < 750 Zellen/μl) ○ Thrombozytopenie (Plättchen $< 100,000$/μl) ○ eGFR < 40 ml/min/1,73 m². <p>Im Falle solcher Abweichungen konnten die Tests innerhalb von ca. 2 Wo. nach der ersten Messung wiederholt werden, und die Wiederholungswerte konnten für die Aufnahme genutzt werden, sofern sie den Einschlusskriterien entsprachen.</p> • EKG Abweichungen (z. B. nach Fridericia korrigierte > 500 ms), so dass nach Meinung des Prüfarztes oder Sponsors ein nicht vertretbares Risiko für die Teilnahme des Patienten an der Studie bestand. • symptomatischer <i>Herpes simplex</i>. • Nachweis einer aktiven TB nach einem positiven Tuberkulin PPD-Test (Induration ≥ 5 mm ungefähr nach dem 2. oder 3. Tag nach der Anwen-

Item ^a	Charakteristikum	Studieninformation
		<p>dung, unabhängig vom Impfstatus), Krankengeschichte, Symptome, und abnormaler Röntgenuntersuchung.</p> <ul style="list-style-type: none"> ○ <i>QuantiFeron</i>[®]-<i>TB Gold</i>- oder <i>T-SPOT</i>[®].<i>TB</i>-Test konnten anstelle des PPD-Tests angewendet werden. Patienten werden ausgeschlossen, wenn der Test nicht negativ war. ● Nachweis einer latenten TB (positiver PPD-Test, keine Symptome für eine aktive TB, abnormale Röntgenuntersuchung), außer die Patienten hatten eine mind. vierwöchige Behandlung vor der Randomisierung abgeschlossen und waren einverstanden, die verbleibende Behandlung während der Studie abzuschließen. <ul style="list-style-type: none"> ○ Bei positivem PPD-Test und negativer Krankengeschichte sowie Röntgenuntersuchungsbefunde passend zu aktiver TB, konnte ein <i>QuantiFeron</i>[®]-<i>TB Gold</i>- oder <i>T-SPOT</i>[®].<i>TB</i>-Test angewendet werden. Wenn der Test nicht negativ war, wurde eine latente TB angenommen (für Studienzwecke). ○ Ein <i>QuantiFeron</i>[®]-<i>TB Gold</i>- oder <i>T-SPOT</i>[®].<i>TB</i>-Test konnte anstelle des PPD-Tests angewendet werden. Wenn der Test positiv war, wurde eine latente TB angenommen. Wenn der Test nicht negativ war, konnte er innerhalb von ungefähr 2 Wo. nach dem ersten Test einmal wiederholt werden. Wenn der wiederholte Test wieder nicht negativ war, wurde eine latente TB angenommen (für Studienzwecke). ○ Ausnahmen beinhalten Patienten, die zu einem früheren Zeitpunkt bereits eine aktive TB hatten und die über einen Nachweis über die Durchführung einer geeigneten Behandlung verfügten. ● positive Testergebnisse für HBV definiert als: <ul style="list-style-type: none"> ○ positiv auf HBsAg oder ○ positiv auf Anti-HBcAb aber negativ auf HBsAb oder ○ für Patienten aus Japan oder anderswo falls benötigt, positiv auf Anti-HBsAb und positiv auf HBV-DNA. <p>Wenn einer der HBV Tests unbestimmte Ergebnisse lieferte, wurde ein anderer Kontrolltest durchgeführt.</p> ● positive Testergebnisse für HCV. ● Nachweis einer humanen HIV-Infektion und/oder HIV-Antikörper positiv. ● weibliches Geschlecht, Alter $\geq 40 - < 60$ Jahre, in der Menopause seit mind. 12 Monaten;

Item ^a	Charakteristikum	Studieninformation
		<p>Patientinnen hatten FSH < 40 U/ml und stimmten nicht zu, für die Studiendauer bis mind. 28 Tage nach der letzten oralen Dosisverabreichung zwei hocheffektive Verhütungsmethoden anzuwenden bzw. für 28 Tage oder den Zeitraum, der in der lokalen Adalimumab Fachinformation vorgegeben ist (je nachdem, welcher länger war), nach einer s. c. Injektion oder angeboren unfruchtbar/chirurgisch sterilisiert.</p> <p>Wenn die Patienten in die Studie aufgenommen wurden, während sie eine Medikation einnahmen, die vor der Randomisierung abgesetzt werden musste, sollten diese Patienten erst nach der Bestimmung ihrer Eignung für die Randomisierung um den Abbruch der Medikation gebeten werden.</p> <p>Die Patienten, die in die Studie aufgenommen wurden, die Einschlusskriterien aber nicht erfüllten, sollten von der Studie ausgeschlossen werden. Diese Patienten konnten der Studie wieder beitreten (z. B. eine erneute Einwilligung abgeben), wenn der Prüfarzt der Meinung war, dass der Patient in der Zukunft den Einschlusskriterien entsprechen würde. Der Prüfarzt sollte ungefähr einen Monat warten, bevor er den Patient erneut in die Studie aufnahm.</p>
4b	Studienorganisation und Ort der Studiendurchführung	Multizentrische Studie, 335 Studienzentren in 28 Ländern (Argentinien, Belgien, , China, Deutschland, Frankreich, Griechenland, Großbritannien, Japan, Kanada, Kroatien, Lettland, Litauen, Mexiko, Niederlande, Polen, Portugal, Rumänien, Russland, Schweiz, Slowakei, Slowenien, Spanien, Südafrika, Südkorea, Taiwan, Tschechische Republik, Ungarn, Vereinigte Staaten von Amerika).
5	Interventionen Präzise Angaben zu den geplanten Interventionen in jeder Gruppe und zur Administration etc.	<p>Alle Gruppen</p> <ul style="list-style-type: none"> • <i>Screening</i>: Patientenselektion nach den unter Item^{4a} genannten Kriterien. • Randomisierung, Patienten mit Nierenfunktionsstörung (eGFR < 60 ml/min/1,73 m²), die einer aktiven Behandlung zugeordnet wurden, erhielten Baricitinib statt 4 mg einmal täglich 2 mg einmal täglich. • Behandlungsphase von 52 Wo. • Alle Patienten erhielten begleitend die individuell bestehende cDMARD-Therapie (inkl. MTX), NSAIDs, Analgetika oder Kortikosteroide. • <i>Rescue</i>-Therapie für Patienten mit unzureichender Reaktion ab 16. Wo. mit Baricitinib 4 mg einmal täglich. • Placebo-Gruppe wechselte nach 24 Wo. zu Baricitinib 4 mg einmal täglich. • Nach 52 Wo. Übertritt in Extensionsstudie

Item ^a	Charakteristikum	Studieninformation
		<p>(JADY) oder <i>Follow-up</i> ca. 4 Wo. nach letzter Dosis.</p> <p>Teil A (doppelblind; Placebo- und aktiv-kontrolliert, 0 Wo. bis 24 Wo.)</p> <p>Baricitinib</p> <ul style="list-style-type: none"> • Baricitinib täglich, oral, in Tablettenform <ul style="list-style-type: none"> ○ 4 mg (oder 2 mg, wenn eGFR < 60 ml/min/1,73m²) • Adalimumab-Placebo s.c. Injektion alle zwei Wochen. <p>Adalimumab</p> <ul style="list-style-type: none"> • Baricitinib-Placebo täglich, oral, in Tablettenform • Adalimumab s.c. Injektion alle zwei Wochen. <ul style="list-style-type: none"> ○ 40 mg <p>Placebo</p> <ul style="list-style-type: none"> • Baricitinib-Placebo täglich, oral, in Tablettenform • Adalimumab-Placebo s.c. Injektion alle zwei Wochen. <p>Teil B (doppelblind; aktiv-kontrolliert, Ende 24 Wo. bis 52 Wo.)</p> <p>Baricitinib</p> <ul style="list-style-type: none"> • Baricitinib täglich, oral, in Tablettenform <ul style="list-style-type: none"> ○ 4 mg (oder 2 mg, wenn eGFR < 60 ml/min/1,73m²) • Adalimumab-Placebo s.c. Injektion alle zwei Wochen. <p>Adalimumab</p> <ul style="list-style-type: none"> • Baricitinib-Placebo täglich, oral, in Tablettenform • Adalimumab s.c. Injektion alle zwei Wochen <ul style="list-style-type: none"> ○ 40 mg. <p>Placebo</p> <ul style="list-style-type: none"> • Baricitinib täglich, oral, in Tablettenform <ul style="list-style-type: none"> ○ 4 mg (oder 2 mg, wenn eGFR < 60 ml/min/1,73m²) • Adalimumab-Placebo s.c. Injektion alle zwei Wochen. <p>Teil C (4-wöchige <i>Follow-up</i>-Phase nach 52 Wo.)</p> <p>Patienten, die die Studie nach 52 Wochen beendeten und nicht in die Extensionsstudie JADY wechselten, beendeten die jeweilige Intervention und besuchten nach ca. vier Wochen eine <i>Follow-up</i> Studienvsiste.</p>

Item ^a	Charakteristikum	Studieninformation
6	Zielkriterien	
6a	Klar definierte primäre und sekundäre Zielkriterien, Erhebungszeitpunkte, ggf. alle zur Optimierung der Ergebnisqualität verwendeten Erhebungsmethoden (z. B. Mehrfachbeobachtungen, Training der Prüfer) und ggf. Angaben zur Validierung von Erhebungsinstrumenten	<p>Primärer Endpunkt (12 Wo.) Überlegenheit von Baricitinib (4 mg einmal täglich) gegen Placebo bei Patienten mit moderater bis schwerer aktiver RA, die ein unzureichendes Ansprechen auf MTX (MTX-IR) zeigten, gemessen am Anteil der Patienten, die nach 12 Wo. eine mind. 20%-ige Verbesserung auf der ACR-Skala (ACR20) erreichten.</p> <p>Sekundärer Endpunkt (12, 24 Wo.) Vergleich der Wirksamkeit von Baricitinib gegen Placebo basierend auf:</p> <ul style="list-style-type: none"> • Veränderung des mTSS von Baseline zu Wo. 24 im Vergleich zu Placebo. • Veränderung des HAQ-DI von Baseline zu Wo 12. im Vergleich zu Placebo. • Veränderung des DAS28-hsCRP von Baseline zu Wo. 12. im Vergleich zu Placebo. • Anteil Patienten mit SDAI $\leq 3,3$ zu Wo. 12. • Durchschnittliche Dauer der Morgensteifigkeit in den sieben Tagen vor Wo. 12 im Vergleich zu Placebo. • Durchschnittliche Schwere der Morgensteifigkeit in den sieben Tagen vor Wo. 12 im Vergleich zu Placebo. • Durchschnittliche Müdigkeit (NRS) in den sieben Tagen vor Wo. 12 im Vergleich zu Placebo. • Durchschnittlicher Gelenkschmerz (NRS) in den sieben Tagen vor Wo. 12 im Vergleich zu Placebo. <p>Vergleich der Wirksamkeit von Baricitinib gegen Adalimumab basierend auf:</p> <ul style="list-style-type: none"> • Anteil Patienten mit ACR20 zu Wo. 12 im Vergleich zu Adalimumab. • Veränderung des DAS28-hsCRP von Baseline zu Wo. 12. im Vergleich zu Adalimumab. <p>Weitere sekundäre Endpunkte (12, 16, 24, 52 Wo.) Vergleich der Wirksamkeit von Baricitinib basierend auf weiteren Parametern:</p> <ul style="list-style-type: none"> • Anteil Patienten mit DAS28-hsCRP $\leq 3,2$ und DAS28-hsCRP $< 2,6$ zu Wo. 12, 24 und 52. • Anteil Patienten mit HAQ-DI $\geq 0,22$ und $\geq 0,3$ zu Wo. 12, 24 und 52. • Veränderung in mTSS von Baseline zu Wo. 16 und 52. Anteil Patienten mit mTSS ≤ 0 von Baseline zu Wo. 16, 24, 52. • Veränderung in Verengung des Gelenkabstandes und Knochenerosion von Baseline zu Wo. 16, 24

Item ^a	Charakteristikum	Studieninformation
		<p>und 52.</p> <ul style="list-style-type: none"> • Anteil Patienten mit ACR20 zu Wo. 24 und 52. • Anteil Patienten mit ACR50 zu Wo. 12, 24 und 52. • Anteil Patienten mit ACR70 zu Wo. 12, 24 und 52. • Prozentuelle Veränderung der individuellen Komponenten des ACR <i>Core Set</i> von Baseline zu Wo. 12, 24 und 52. • Veränderung des DAS28-hsCRP von Baseline zu Wo. 52. • Veränderung des DAS28-ESR von Baseline zu Wo. 52. • Anteil Patienten mit DAS28-ESR $\leq 3,2$ und DAS28-ESR $< 2,6$ zu Wo. 12, 24 und 52. • Veränderung des HAQ-DI von Baseline zu Wo. 24 und 52. • Veränderung des CDAI von Baseline zu Wo. 12, 24 und 52. • Veränderung des SDAI von Baseline zu Wo. 12, 24 und 52. • Anteil Patienten mit CDAI $\leq 2,8$ zu Wo. 12, 24 und 52. • Anteil Patienten mit SDAI $\leq 3,3$ zu Wo. 24 und 52. • Anteil Patienten mit ACR/EULAR Remission (Boolean-basiert) zu Wo. 12, 24 und 52. • Evaluierung der Hybrid ACR <i>Response</i> (gebunden) zu Wo.12, 24 und 52. • Veränderung des FACIT-F von Baseline zu Wo. 12, 24 und 52. • Veränderung der Einzel- und Summenskalen des SF-36 v2 akut MCS und des SF-36 v2 akut PCS von Baseline zu Wo. 12, 24 und 52. • Veränderung des EQ-5D-5L von Baseline zu Wo. 12, 24 und 52. • Veränderung des WPAI-RA von Baseline zu Wo. 12, 24 und 52. • Evaluierung der Gesundheitswesen-Ressourcennutzung bis Wo. 52. • Veränderung der Dauer der Morgensteifigkeit von Baseline zu Wo. 12, 24 und 52 in (ePRO Tablet). • Veränderung der Müdigkeit (NRS) von Baseline zu Wo. 52 (ePRO Tablet). • Veränderung der Gelenkschmerzen (NRS) von Baseline zu Wo. 52 (ePRO Tablet). <p>Weitere sekundäre Endpunkte gemäß SAP (12, 24, 52 Wo.)</p>

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • Anteil Patienten mit SDAI \leq 11 zu Wo. 12, 24 und 52. • Anteil Patienten mit CDAI \leq 10 zu Wo. 12, 24 und 52. • Anteil Patienten mit Verbesserung im FACIT-F \geq 3,56 zu Wo. 12, 24 und 52. • Veränderungen des RF und ACPA von Baseline zu Wo.12, 24 und 52. • Charakterisierung des zeitlichen Verlaufs der Dosis-Wirkung Beziehung von Baricitinib für wichtige Wirksamkeits- und Sicherheitsendpunkte, Identifizierung möglicher Faktoren, die einen Einfluss auf diese Endpunkte haben.
6b	Änderungen der Zielkriterien nach Studienbeginn, mit Begründung	Keine Angaben
7	Fallzahl	
7a	Wie wurden die Fallzahlen bestimmt?	<p>Es wurden etwa 1.280 Patienten (480 in der Baricitinib Gruppe, 480 in der Placebo Gruppe und 320 in der Adalimumab Gruppe) benötigt,</p> <p>A) um den Unterschied von Baricitinib zu Placebo mit einer Aussagekraft von $> 95\%$ zu zeigen. <u>Überlegenheitsschwelle</u> Anteil Patienten mit ACR20 nach 12 Wo. (Annahme: 60% versus 35%). <u>Berechnungsgrundlage</u> zwei-seitiger Chi-2-Test mit Signifikanzlevel von 0,05.</p> <p>B) um den Unterschied von Baricitinib zu Placebo mit einer Aussagekraft von ungefähr 94% zu zeigen. <u>Überlegenheitsschwelle</u> mTSS Effektgröße 0,25. <u>Berechnungsgrundlage</u> zwei-seitiger t-Test mit Signifikanzlevel von 0,04.</p> <p>C) um die Nicht-Unterlegenheitsanalyse von Baricitinib zu Adalimumab mit einer Aussagekraft von ungefähr 93% zu zeigen. <u>Nichtunterlegenheitsschwelle</u> Anteil Patienten mit ACR20 nach 12 Wo. (Annahme: 60%). <u>Berechnungsgrundlage</u> Ein-seitiges Signifikanzlevel von 0,025.</p>

Item ^a	Charakteristikum	Studieninformation
7b	Falls notwendig, Beschreibung von Zwischenanalysen und Kriterien für einen vorzeitigen Studienabbruch	<p>Zwischenanalysen Keine Angabe</p> <p>Vorzeitiger Studienabbruch</p> <p><u>Gründe für eine Unterbrechung der Behandlung mit der Studienmedikation</u></p> <ul style="list-style-type: none"> • WBC-Zahl < 2.000 Zellen/µl • Gesamtneutrophilenzahl < 1.000 Zellen/µl • Lymphozytenzahl < 500 Zellen/µl • Thrombozytenzahl < 75.000/µl • Ohne dokumentierte Nierenfunktionsstörung: eGFR < 40 ml/min/1,73 m² (Baricitinib 4 mg einmal täglich) • Mit dokumentierter Nierenfunktionsstörung: eGFR < 30 ml/min/1,73 m² (Baricitinib 2 mg einmal täglich) • ALT oder AST > 5-fach ULN • Hämoglobin < 8 g/dl. <p><u>Abbruchregelungen für Patienten</u></p> <p>Jeder Patient, für den ein dauerhafter Studienabbruch aufgrund von unerwünschten Ereignissen oder abnormen Laborwerten indiziert war, sollte inklusive der Angabe des Abbruchgrundes mit der Bezeichnung „Schwerwiegendes unerwünschtes Ereignis“ protokolliert werden (dokumentiert als „Anderer schwerwiegender Grund“).</p> <p><u>Gründe für einen Abbruch der Behandlung mit Prüfmedikation</u></p> <ul style="list-style-type: none"> • ALT oder AST > 8-fach ULN • ALT oder AST > 5-fach ULN für mehr als 2 Wo. nach temporärer Studienunterbrechung • ALT oder AST > 3-fach ULN und Gesamtbilirubin > 2-fach ULN • ALT oder AST > 3-fach ULN mit Auftreten von Fatigue, Nausea, Übelkeit, Schmerzen oder Schmerzempfindlichkeit im rechten Oberbauch, Fieber, Ausschlag und/oder Eosinophilie (> 5%) • WBC Zahl < 1.000 Zellen/µl • Absolute Neutrophilenzahl < 500 Zellen/µl • Lymphozytenzahl < 200 Zellen/µl • Hämoglobin < 6.5 g/dl • Symptomatischer Herpes Zoster • Schwangerschaft • Malignität • Infektion, die es, nach Meinung des Prüfarztes, nötig macht, die Studie abzurechnen • Anaphylaktische oder andere schwerwiegende allergische Reaktion

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • Symptome, die auf ein Lupus-ähnliches Syndrom hindeuten. <p>Außerdem</p> <ul style="list-style-type: none"> • Nachträgliche Feststellung, dass ein Ausschlusskriterium für die Studie bestand (bei Ausschluss des Patienten aus der Studie gehörte der Patient trotzdem bis Wo. 12 (<i>Follow-up</i> für die Primäranalyse) zur ITT-Population). • Teilnahme des Patienten an einer weiteren, aus wissenschaftlicher oder medizinischer Sicht nicht-kompatiblen klinischen Studie (bzw. anderer medizinischer Forschung). • Auf Entscheidung des Prüfarztes. • Auf Entscheidung des pU (oder dessen Beauftragten). • Auf Entscheidung des Patienten. • Patient verhält sich nicht entsprechend der Protokollanweisungen. • Patient widerruft Einwilligungserklärung. • Prüfarzt oder pU (oder dessen Beauftragter) beenden die Studie. <p><u>Ausschluss von Studienorten</u></p> <p>Der Ausschluss von Studienorten war nach Ermessen des pU, des Prüfarztes oder der begutachtenden Ethikkommission aus medizinischen, sicherheitsrelevanten, regulatorischen oder anderen Gründen, in Übereinstimmung mit anzuwendenden Gesetzen, Regularien und den Prinzipien der guten klinischen Praxis, notwendig.</p>
8	Randomisierung, Erzeugung der Behandlungsfolge	
8a	Methode zur Generierung der zufälligen Zuteilung	Computerrandomisierung über ein <i>Interaktives Voice Response-System</i> (IVRS)
8b	Einzelheiten (z. B. Blockrandomisierung, Stratifizierung)	<p>Randomisierung beim <i>Screening</i> stratifiziert nach Region und Status des Gelenkverschleißes (1-2 Gelenkerosionen und seropositiver Status versus mind. 3 Gelenkerosionen).</p> <p>Patienten mit Nierenfunktionsstörung (eGFR \geq 40 und $<$ 60 ml/min/1,73 m²), denen eine Intervention mit Baricitinib zugeteilt wurde, wurden mit einer Baricitinib Dosierung von 2 mg einmal täglich behandelt.</p>
9	<p>Randomisierung, Geheimhaltung der Behandlungsfolge (<i>allocation concealment</i>)</p> <p>Durchführung der Zuteilung (z. B. nummerierte Behälter; zentrale Randomisierung per Fax / Telefon), Angabe, ob Geheimhaltung bis zur Zuteilung gewährleistet war</p>	<p>Übermittlung der Zuteilung</p> <p>Computer-generierte Zufallssequenz anhand IVRS.</p> <p>Verdeckung der Behandlungsfolge</p> <p>Methodenimmanent durch IVRS.</p>

Item ^a	Charakteristikum	Studieninformation
10	Randomisierung, Durchführung Wer führte die Zuteilung durch, wer nahm die Probanden/Patienten in die Studie auf und wer teilte die Probanden/Patienten den Gruppen zu?	Durchführung der Zuteilung Zuteilung der Studienteilnehmer durch IVRS. Aufnahme in die Studie Durch Prüfarzt.
11	Verblindung	
11a	Waren a) die Probanden / Patienten und / oder b) diejenigen, die die Intervention / Behandlung durchführten, und / oder c) diejenigen, die die Zielgrößen beurteilten, verblindet oder nicht verblindet, wie wurde die Verblindung vorgenommen?	a: Patienten waren verblindet. b: Prüfarzte waren verblindet. c: Alle Personen, die die Zielgrößen bewertet haben, waren verblindet. Verblindung durch IVRS. Zwei Datenbanksperren: nach 24 Wo. (<i>Visit 10</i>) und nach der Beendung der Studie. Eine begrenzte Anzahl an Angestellten des pU und der externen Auftragnehmer wurde nach der ersten Sperrung entblindet. Es wurde sichergestellt, dass diese entblindeten Personen bezüglich patientenbezogener Angelegenheiten keinen direkten Kontakt mit noch verblindeten Personen hatten. Von daher blieben Patienten, Prüfarzte und Mitarbeiter, die an der Studie beteiligt waren sowie alle anderen Mitarbeiter in Bezug auf die Daten von Wo. 24 und die individuellen Behandlungszuteilung bis zur Beendung der Studie verblindet. Ihre Entblindung erfolgte erst nach Validierung und der zweiten Sperrung der Berichtsdatenbank.
11b	Falls relevant, Beschreibung der Ähnlichkeit von Interventionen	Baricitinib- und dessen Placebo-Tabletten waren in Farbe, Form, Geruch und Geschmack identisch. Adalimumab- und dessen Placebo-Injektionen waren ebenfalls identisch im Hinblick auf Form, Geruch und Geschmack.
12	Statistische Methoden	
12a	Statistische Methoden zur Bewertung der primären und sekundären Zielkriterien	Primäres Zielkriterium Unterschied zwischen den Anteilen an Patienten, die ACR20 nach 12 Wo. erreichen (mITT Population). <ul style="list-style-type: none"> Regressionsanalyse (logistisch) parametrischer Test (Wald Test mit zweiseitigem Signifikanz-Level von 0,05). Sekundäre Zielkriterien <ul style="list-style-type: none"> Kovarianzanalysen (ANCOVA) Varianzanalyse Nicht-parametrischer Test (Hodges-Lehmann Schätzer) Lineare Extrapolation <i>Last Observation Carried Forward (LOCF)</i> <i>Modified Last Observation Carried Forward (mLOCF)</i> <i>Baseline Observation Carried Forward (BOCF)</i>

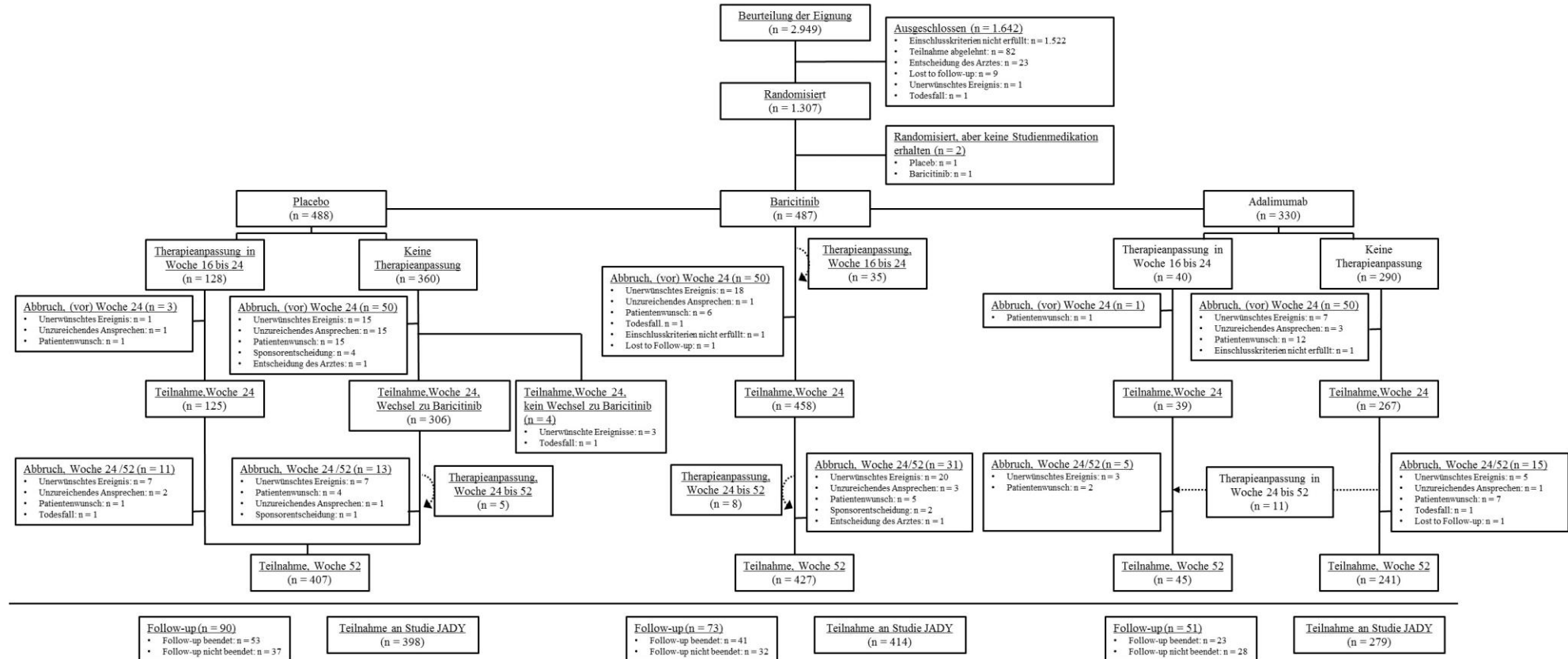
Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • <i>Modified Baseline Observation Carried Forward</i> (mBOCF) • <i>Mixed Effect Model Repeated Measurement</i> (MMRM) • Regressionsanalyse (logistisch) • <i>Nonresponder Imputation</i> (NRI) • Newcombe-Wilson Methode (Newcombe 1998). <p>Studienpopulation</p> <p>Modifizierte <i>Intention-to-Treat</i> Population (mITT) Alle randomisierten Patienten, die mind. eine Dosis der Studienmedikation erhalten hatten.</p> <p><i>Per-Protocol</i> Population (PP) Alle mITT Patienten, die nicht während Teil A als <i>noncompliant</i> eingestuft wurden, die in Teil A keine gravierenden Protokollverletzungen begingen und deren Studienzentren keine wichtigen GCP-Probleme hatten, die an eine regulatorische Behörde übermittelt werden mussten (unabhängig vom Studienteil).</p> <p><i>Safety</i>-Population Alle randomisierten Patienten, die mind. eine Dosis der Studienmedikation erhielten und die nicht aus dem Grund „<i>Lost to Follow-up</i>“ zur ersten Visite Post-Baseline ausgeschieden sind.</p> <p><i>Follow-up</i> Population Alle Patienten, die an der <i>Follow-up</i>-Phase teilnahmen.</p> <p><i>Safety Follow-up</i> Population Alle Patienten der <i>Safety</i>-Population, die an der <i>Follow-up</i>-Phase teilnahmen und die nicht aufgrund von <i>Lost to Follow-up</i> in der <i>Follow-up</i>-Phase ausgeschieden sind.</p>
12b	Weitere Analysen, wie z. B. Subgruppenanalysen und adjustierte Analysen	<p>Subgruppenanalyse (mITT Population) nach 12 und 24 Wo. für ACR20, ACR50, DAS28-hsCRP < 2,6, DAS28-hsCRP ≤ 3,2, SDAI ≤ 3,3 und HAQ-DI ≥ 0,22, sowie für Veränderungen zu Baseline in HAQ-DI und DAS28-hsCRP:</p> <ul style="list-style-type: none"> • Geschlecht (weiblich, männlich) • Alter (< 65 Jahre, ≥ 65 Jahre) • Alter (< 75 Jahre, ≥ 75 Jahre) • Alter (< 85 Jahre, ≥ 85 Jahre) • Altersgruppe (< 65 Jahre, 65 - < 75 Jahre, ≥ 75 - < 85 und ≥ 85 Jahre) • Gewicht zu Baseline (≤ Median, > Median) • Gewichtsklasse zu Baseline (< 60 kg, ≥ 60 - < 100 kg, ≥ 100 kg) • BMI zu Baseline (< 18,5 kg/m², ≥ 18,5 - < 25 kg/m², ≥ 25 - < 30 kg/m², ≥ 30 kg/m²)

Item ^a	Charakteristikum	Studieninformation
		<ul style="list-style-type: none"> • BMI zu Baseline (\leq Median, $>$ Median) • Ethnie (asiatisch, schwarz/afroamerikanisch, kaukasisch, andere [indianische Ureinwohner Amerikas/Ureinwohner Alaskas, Ureinwohner Hawaiis oder anderer Pazifikinseln], multipel) • Region (Vereinigte Staaten und Kanada, Zentral- und Südamerika mit Mexiko, Europa, Asien, Japan, Rest der Welt) • Spezifische Region (Europa, andere) • Spezifisches Land (Vereinigte Staaten von Amerika, andere) • Spezifische Land (Japan, andere) • Nierenfunktion bei <i>Screening</i> (beeinträchtigt (eGFR $<$ 60 ml/min/1,73 m²) oder nicht beeinträchtigt (eGFR \geq 60 ml/min/1,73 m²)) • Erkrankungsdauer seit der Diagnose ($<$ 1 J., $\geq 1 - <$ 5 J., $\geq 5 - <$ 10 J., ≥ 10 J.) • Krankheitsaktivität zu Baseline (DAS28-hsCRP \leq 5,1, DAS28-hsCRP $>$ 5,1) • Gelenkerosion zu Baseline (ja, nein) • Seropositivität (RF positiv oder ACPA positiv, RF negativ und ACPA negativ) • Seropositivität (RF positiv, RF negativ) • Seropositivität (ACPA positiv, ACPA negativ) • Seropositivität und Gelenkerosionsstatus, beides zu Baseline (seropositiv (RF oder ACPA) und mind. 3 Gelenkerosionen; andere) • Art der Begleitmedikation (nur MTX, cDMARDs, MTX und andere cDMARDs) • mTSS zu Baseline (\leq Median, $>$ Median) • Therapie mit Kortikosteroiden zu Baseline (ja, nein)
-	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	a) n = 1307 Zwei Patienten wurden randomisiert, aber nicht behandelt und wurden deshalb nicht in die analysierten Populationen aufgenommen. Interventionsgruppe Baricitinib: n = 488 Interventionsgruppe Adalimumab: n = 330 Kontrollgruppe Placebo: n = 489 b) n = 1305 Interventionsgruppe Baricitinib: n = 487 Interventionsgruppe Adalimumab: n = 330 Kontrollgruppe Placebo: n = 488

Item ^a	Charakteristikum	Studieninformation
		c) n = 1305 Interventionsgruppe Baricitinib: n = 487 Interventionsgruppe Adalimumab: n = 330 Kontrollgruppe Placebo: n = 488
13b	Für jede Gruppe: Beschreibung von verlorenen und ausgeschlossenen Patienten nach Randomisierung mit Angabe von Gründen	Angaben bezogen auf 52 Wo. Studiendauer <u>Interventionsgruppe Baricitinib: n = 60</u> <ul style="list-style-type: none"> • Unerwünschtes Ereignis: n = 38 • Fehlende Wirksamkeit: n = 4 • Ärztliche Entscheidung: n = 1 • Tod: n = 2 • Patientenentscheidung: n = 11 • Einschlusskriterien nicht erfüllt: n = 1 • Entscheidung des Sponsors: n = 2 • <i>Lost-to-Follow-up</i>: n = 1 <u>Interventionsgruppe Adalimumab: n = 44</u> <ul style="list-style-type: none"> • Unerwünschtes Ereignis: n = 15 • Fehlende Wirksamkeit: n = 4 • Tod: n = 1 • Patientenentscheidung: n = 22 • Einschlusskriterien nicht erfüllt: n = 1 • <i>Lost-to-Follow-up</i>: n = 1 <u>Kontrollgruppe Placebo: n = 81</u> <ul style="list-style-type: none"> • Unerwünschtes Ereignis: n = 33 • Fehlende Wirksamkeit: n = 19 • Ärztliche Entscheidung: n = 1 • Tod: n = 2 • Patientenentscheidung: n = 21 • Entscheidung des Sponsors: n = 5
14	Aufnahme / Rekrutierung	
14a	Nähere Angaben über den Zeitraum der Studienaufnahme der Probanden / Patienten und der Nachbeobachtung	<ul style="list-style-type: none"> • Geplanter erster Patient: Oktober 2012 • Geplanter letzter Patient: Januar 2015 • Tatsächlicher erster Patient: 23.10.2012 • Tatsächlicher letzter Patient: 29.09.2015
14b	Informationen, warum die Studie endete oder beendet wurde	Die Studie endete regulär.
a: nach CONSORT 2010. Alle verwendeten Abkürzungen sind im Abkürzungsverzeichnis erläutert.		

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

Stellen Sie für jede Studie den Patientenfluss in einem Flow-Chart gemäß CONSORT dar.



Anhang 4-F: Bewertungsbögen zur Einschätzung von Verzerrungsaspekten

Der nachfolgend dargestellte Bewertungsbogen dient der Dokumentation der Einstufung des Potenzials der Ergebnisse für Verzerrungen (Bias). Für jede Studie soll aus diesem Bogen nachvollziehbar hervorgehen, inwieweit die Ergebnisse für die einzelnen Endpunkte als möglicherweise verzerrt bewertet wurden, was die Gründe für die Bewertung waren und welche Informationen aus den Quellen dafür Berücksichtigung fanden.

Der Bogen gliedert sich in zwei Teile:

- Verzerrungsaspekte auf Studienebene. In diesem Teil sind die endpunktübergreifenden Kriterien aufgelistet.
- Verzerrungsaspekte auf Endpunktebene. In diesem Teil sind die Kriterien aufgelistet, die für jeden Endpunkt separat zu prüfen sind.

Für jedes Kriterium sind unter „Angaben zum Kriterium“ alle relevanten Angaben aus den Quellen zur Bewertung einzutragen (Stichworte reichen ggf., auf sehr umfangreiche Informationen in den Quellen kann verwiesen werden).

Grundsätzlich sollen die Bögen studienbezogen ausgefüllt werden. Wenn mehrere Quellen zu einer Studie vorhanden sind, müssen die herangezogenen Quellen in der folgenden Tabelle genannt und jeweils mit Kürzeln (z. B. A, B, C ...) versehen werden. Quellenspezifische Angaben im weiteren Verlauf sind mit dem jeweiligen Kürzel zu kennzeichnen.

Hinweis: Der nachfolgend dargestellte Bewertungsbogen ist die Blankoversion des Bogens. Dieser Blankobogen ist für jede Studie heranzuziehen. Im Anschluss daran ist ein Bewertungsbogen inklusive Ausfüllhinweisen abgebildet, der als Ausfüllhilfe dient, aber nicht als Vorlage verwendet werden soll.

Beschreiben Sie nachfolgend die Verzerrungsaspekte jeder eingeschlossenen Studie (einschließlich der Beschreibung für jeden berücksichtigten Endpunkt). Erstellen Sie hierfür je Studie eine separate Version des nachfolgend dargestellten Bewertungsbogens.

Tabelle 4-178 (Anhang): Bewertungsbogen zur Beschreibung von Verzerrungsaspekten für Studie RA-BEAM (JADV)

Studie: RA-BEAM (JADV)

Tabelle: Liste der für die Bewertung herangezogenen Quellen

Genauere Benennung der Quelle	Kürzel
Studienbericht	A [9]
Studienprotokoll	B [7]

A Verzerrungsaspekte auf Studienebene:

Einstufung als randomisierte Studie

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein → Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

A: Randomisierte, doppelblinde, aktiv-kontrollierte Phase-III-Studie zur Untersuchung der Wirksamkeit und Sicherheit von Baricitinib (LY3009104) bei Patienten mit mittelschwerer bis schwerer aktiver RA, die ein unzureichendes Ansprechen, oder eine Unverträglichkeit gegenüber Basistherapeutika (cDMARDs) gezeigt haben.

1.

für randomisierte Studien: Adäquate Erzeugung der Randomisierungssequenz

ja **unklar** **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

für nicht randomisierte Studien: Zeitliche Parallelität der Gruppen

ja **unklar** **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien: Verdeckung der Gruppenzuteilung („allocation concealment“)

ja **unklar** **nein**

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

für nicht randomisierte Studien: Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren

ja unklar nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und behandelnden Personen**Patient:**

ja unklar nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verblindung war zu jeder Zeit der Studie gewährleistet.

behandelnde bzw. weiterbehandelnde Personen:

ja unklar nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verblindung war zu jeder Zeit der Studie gewährleistet.

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

ja unklar nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

5. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrungen führen können

ja nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B

Einstufung des Verzerrungspotenzials der Ergebnisse auf Studienebene (ausschließlich für randomisierte Studien durchzuführen):

niedrig hoch

Begründung für die Einstufung:

A, B: Für die Studie R-BEAM (JADV) liegen keine Aspekte vor, die zu einem erhöhten Verzerrungspotenzial führen. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt.

Begründung für die Einstufung:

Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

B Verzerrungsaspekte auf Endpunktebene pro Endpunkt:**Endpunkt: Anzahl druckschmerzhafter Gelenke und geschwollener Gelenke****1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppenzuteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A; B

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt Anzahl der druckschmerzhaften und geschwollenen Gelenke wurde als niedrig eingeschätzt. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor. Die Feststellung von geschwollenen und druckschmerzhaften Gelenken ist objektiv möglich.

Endpunkt: Disease Activity Score 28 High Sensitivity C-Reactive Protein (DAS28-hsCRP)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppenzuteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A; B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt DAS28-hsCRP wurde als niedrig eingeschätzt. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor. Die Feststellung von geschwollenen und druckschmerzhaften Gelenken, als Einzelkomponenten dieses Konstruktes, mit dem das komplexe Krankheitsgeschehen der rheumatoiden Arthritis abgebildet werden soll, ist objektiv möglich.

Endpunkt: *Simplified Disease Activity Index (SDAI)***1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppeneinteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A; B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt SDAI war niedrig. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor. Die Feststellung von geschwollenen und druckschmerzhaften Gelenken, als Einzelkomponenten dieses Konstruktes, mit dem das komplexe Krankheitsgeschehen der rheumatoiden Arthritis abgebildet werden soll, ist objektiv möglich.

Endpunkt: Clinical Disease Activity Index (CDAI)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppeneinteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt CDAI war niedrig. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor. Die Feststellung von geschwollenen und druckschmerzhaften Gelenken, als Einzelkomponenten dieses Konstruktes, mit dem das komplexe Krankheitsgeschehen der rheumatoiden Arthritis abgebildet werden soll, ist objektiv möglich.

Endpunkt: EULAR Responsekriterien („good Response“, DAS28-basiert)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppenzuteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A; B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt EULAR Responsekriterien („good Response“, DAS28-basiert) war niedrig. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Endpunkt: Patient Assessment of Pain (PAP) Visual Analog Scale (VAS)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppenzuteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A; B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt PAP war niedrig. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Endpunkt: Patient's Global Assessment of Disease Activity (PtGADA) (VAS)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und MTX-Medikation waren in Farbe, Form, Geruch und Geschmack identisch.
Die Gruppenzuteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A; B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt PtGADA war niedrig. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Endpunkt: Health Assessment Questionnaire-Disability Index (HAQ-DI)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppenzuteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A; B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt HAQ-DI war niedrig. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Endpunkt: Dauer der Morgensteifigkeit (min)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppenzuteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A; B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B: Die zur Messung verwendeten ePRO-Tablets wurden an die teilnehmenden Studienzentren ausgeteilt, nachdem bereits ein Teil der Studienpopulation eingeschlossen war. Daher existieren für ein Subset Patienten (560/1305) keine Baseline-Daten zu diesem Endpunkt und diese Patienten flossen auch nicht in die Auswertung dieses Endpunkts mit ein.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt Dauer der Morgensteifigkeit war niedrig. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung. Die zur Messung verwendeten ePRO-Tablets wurden an die teilnehmenden Studienzentren ausgeteilt, nachdem bereits ein Teil der Studienpopulation eingeschlossen war. Daher existieren für ein Subset Patienten (560/1305) keine Baseline-Daten zu diesem Endpunkt und diese Patienten flossen auch nicht in die Auswertung dieses Endpunkts mit ein.

Begründung für die Einstufung:

Da für den Großteil der eingeschlossenen Patienten Baseline-Daten vorlagen und das Nichtvorliegen von Baseline-Daten nicht mit bestimmten Patientencharakteristika in Verbindung steht (*Missing Completely at Random*), wird nicht von einer Verzerrung des Endpunkts ausgegangen.

Endpunkt: European Quality of Life-5 Dimensions (EQ-5D) VAS**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppenzuteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A; B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt EQ-5D VAS wurde als niedrig eingeschätzt. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Endpunkt: Schwerste Müdigkeit (Worst Tiredness)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppenzuteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A; B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B: Die zur Messung verwendeten ePRO-Tablets wurden an die teilnehmenden Studienzentren ausgeteilt, nachdem bereits ein Teil der Studienpopulation eingeschlossen war. Daher existieren für ein Subset Patienten (560/1305) keine Baseline-Daten zu diesem Endpunkt und diese Patienten flossen auch nicht in die Auswertung dieses Endpunkts mit ein.

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt Dauer der Morgensteifigkeit war niedrig. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Darüber hinaus erfolgte eine ergebnisunabhängige Berichterstattung. Die zur Messung verwendeten ePRO-Tablets wurden an die teilnehmenden Studienzentren ausgeteilt, nachdem bereits ein Teil der Studienpopulation eingeschlossen war. Daher existieren für ein Subset Patienten (560/1305) keine Baseline-Daten zu diesem Endpunkt und diese Patienten flossen auch nicht in die Auswertung dieses Endpunkts mit ein.

Begründung für die Einstufung:

Da für den Großteil der eingeschlossenen Patienten Baseline-Daten vorlagen und das Nichtvorliegen von Baseline-Daten nicht mit bestimmten Patientencharakteristika in Verbindung steht (*Missing Completely at Random*), wird nicht von einer Verzerrung des Endpunkts ausgegangen.

Endpunkt: Short Form (36 item) Health Survey Version 2 akut (SF-36 v2 akut)**1. Verblindung der Endpunkterheber** ja unklar neinAngaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppenzuteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A; B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine ja unklar neinAngaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können ja neinAngaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen): niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für den Endpunkt SF-36 v2 akut wurde als niedrig eingeschätzt. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor.

Endpunkt: Sicherheit (Todesfälle, SUE, UE, Schwere UE (ohne ausschließliche Laborereignisse), Studien-/Therapieabbruch aufgrund von UE, UE von besonderem Interesse)

1. Verblindung der Endpunkterheber

ja unklar nein

Angaben zum Kriterium; obligate Begründung für die Einstufung:

A, B: Die Verum- und Vergleichsmedikation waren in Farbe, Form, Geruch und Geschmack identisch. Die Gruppenzuteilung erfolgte mittels IVRS.

2. Adäquate Umsetzung des ITT-Prinzips

ja unklar nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

3. Ergebnisunabhängige Berichterstattung dieses Endpunkts alleine

ja unklar nein

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

A, B

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

ja nein

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

A, B

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

niedrig hoch

Begründung für die Einstufung:

A, B: Das Verzerrungspotenzial für die Endpunkte zur Sicherheit wurde als niedrig eingeschätzt. Sowohl die Verblindung der Patienten und der Endpunkterheber, das ITT-Prinzip sowie die Erzeugung der Randomisierungssequenz wurden adäquat umgesetzt. Es erfolgte eine ergebnisunabhängige Berichterstattung. Darüber hinaus liegen keine weiteren, das Verzerrungspotenzial beeinflussenden Aspekte vor. Die Feststellung von Tod ist objektiv möglich.

Hinweis: Der nachfolgend dargestellte Bewertungsbogen mit Ausfüllhinweisen dient nur als Ausfüllhilfe für den Blankobogen. Er soll nicht als Vorlage verwendet werden.

Bewertungsbogen zur Beschreibung von Verzerrungsaspekten (Ausfüllhilfe)

Anhand der Bewertung der folgenden Kriterien soll das Ausmaß möglicher Ergebnisverzerrungen eingeschätzt werden (A: endpunktübergreifend; B: endpunktspezifisch).

A Verzerrungsaspekte auf Studienebene:**Einstufung als randomisierte Studie**

ja → Bewertung der Punkte 1 und 2 für randomisierte Studien

nein: Aus den Angaben geht klar hervor, dass es keine randomisierte Zuteilung gab, oder die Studie ist zwar als randomisiert beschrieben, es liegen jedoch Anzeichen vor, die dem widersprechen (z. B. wenn eine alternierende Zuteilung erfolgte). Eine zusammenfassende Bewertung der Verzerrungsaspekte soll für nicht randomisierte Studien nicht vorgenommen werden.

→ Bewertung der Punkte 1 und 2 für nicht randomisierte Studien

Angaben zum Kriterium:

1.

für randomisierte Studien:**Adäquate Erzeugung der Randomisierungssequenz**

ja: Die Gruppenzuteilung erfolgte rein zufällig, und die Erzeugung der Zuteilungssequenz ist beschrieben und geeignet (z. B. computergenerierte Liste).

unklar: Die Studie ist zwar als randomisiert beschrieben, die Angaben zur Erzeugung der Zuteilungssequenz fehlen jedoch oder sind ungenügend genau.

nein: Die Erzeugung der Zuteilungssequenz war nicht adäquat.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:**Zeitliche Parallelität der Gruppen**

ja: Die Gruppen wurden zeitlich parallel verfolgt.

unklar: Es finden sich keine oder ungenügend genaue diesbezügliche Angaben.

nein: Die Gruppen wurden nicht zeitlich parallel verfolgt.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

2.

für randomisierte Studien:**Verdeckung der Gruppenzuteilung („allocation concealment“)** **ja:** Eines der folgenden Merkmale trifft zu:

- Zuteilung durch zentrale unabhängige Einheit (z. B. per Telefon oder Computer)
- Verwendung von für die Patienten und das medizinische Personal identisch aussehenden, nummerierten oder kodierten Arzneimitteln/Arzneimittelbehältern
- Verwendung eines seriennummerierten, versiegelten und undurchsichtigen Briefumschlags, der die Gruppenzuteilung beinhaltet

 unklar: Die Angaben der Methoden zur Verdeckung der Gruppenzuteilung fehlen oder sind ungenügend genau. **nein:** Die Gruppenzuteilung erfolgte nicht verdeckt.Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

für nicht randomisierte Studien:**Vergleichbarkeit der Gruppen bzw. adäquate Berücksichtigung von prognostisch relevanten Faktoren** **ja:** Eines der folgenden Merkmale trifft zu:

- Es erfolgte ein Matching bzgl. der wichtigen Einflussgrößen und es gibt keine Anzeichen dafür, dass die Ergebnisse durch weitere Einflussgrößen verzerrt sind.
- Die Gruppen sind entweder im Hinblick auf wichtige Einflussgrößen vergleichbar (siehe Baseline-Charakteristika), oder bestehende größere Unterschiede sind adäquat berücksichtigt worden (z. B. durch adjustierte Auswertung oder Sensitivitätsanalyse).

 unklar: Die Angaben zur Vergleichbarkeit der Gruppen bzw. zur Berücksichtigung von Einflussgrößen fehlen oder sind ungenügend genau. **nein:** Die Vergleichbarkeit ist nicht gegeben und diese Unterschiede werden in den Auswertungen nicht adäquat berücksichtigt.Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

3. Verblindung von Patienten und behandelnden Personen**Patient:** **ja:** Die Patienten waren verblindet. **unklar:** Es finden sich keine diesbezüglichen Angaben. **nein:** Aus den Angaben geht hervor, dass die Patienten nicht verblindet waren.Angaben zum Kriterium; obligate Begründung für die Einstufung:

behandelnde bzw. weiterbehandelnde Personen:

ja: Das behandelnde Personal war bzgl. der Behandlung verblindet. Wenn es, beispielsweise bei chirurgischen Eingriffen, offensichtlich nicht möglich ist, die primär behandelnde Person (z. B. Chirurg) zu verblinden, wird hier beurteilt, ob eine angemessene Verblindung der weiteren an der Behandlung beteiligten Personen (z. B. Pflegekräfte) stattgefunden hat.

unklar: Es finden sich keine diesbezüglichen Angaben.

nein: Aus den Angaben geht hervor, dass die behandelnden Personen nicht verblindet waren.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

4. Ergebnisunabhängige Berichterstattung aller relevanten Endpunkte

Falls die Darstellung des Ergebnisses eines Endpunkts von seiner Ausprägung (d. h. vom Resultat) abhängt, können erhebliche Verzerrungen auftreten. Je nach Ergebnis kann die Darstellung unterlassen worden sein (a), mehr oder weniger detailliert (b) oder auch in einer von der Planung abweichenden Weise erfolgt sein (c).

Beispiele zu a und b:

- *Der in der Fallzahlplanung genannte primäre Endpunkt ist nicht / unzureichend im Ergebnisteil aufgeführt.*
- *Es werden (signifikante) Ergebnisse von vorab nicht definierten Endpunkten berichtet.*
- *Nur statistisch signifikante Ergebnisse werden mit Schätzern und Konfidenzintervallen dargestellt.*
- *Lediglich einzelne Items eines im Methodenteil genannten Scores werden berichtet.*

Beispiele zu c: Ergebnisgesteuerte Auswahl in der Auswertung verwendeter

- *Subgruppen*
- *Zeitpunkte/-räume*
- *Operationalisierungen von Zielkriterien (z. B. Wert zum Studienende anstelle der Veränderung zum Baseline-Wert; Kategorisierung anstelle Verwendung stetiger Werte)*
- *Distanzmaße (z. B. Odds Ratio anstelle der Risikodifferenz)*
- *Cut-off-points bei Dichotomisierung*
- *statistischer Verfahren*

Zur Einschätzung einer potenziell vorhandenen ergebnisgesteuerten Berichterstattung sollten folgende Punkte – sofern möglich – berücksichtigt werden:

- *Abgleich der Angaben der Quellen zur Studie (Studienprotokoll, Studienbericht, Registerbericht, Publikationen).*
- *Abgleich der Angaben im Methodenteil mit denen im Ergebnisteil. Insbesondere eine stark von der Fallzahlplanung abweichende tatsächliche Fallzahl ohne plausible und ergebnisunabhängige Begründung deutet auf eine selektive Beendigung der Studie hin.*

Zulässige Gründe sind:

- *erkennbar nicht ergebnisgesteuert, z. B. zu langsame Patientenrekrutierung*
- *Fallzahladjustierung aufgrund einer verblindeten Zwischenauswertung anhand der Streuung der Stichprobe*
- *geplante Interimanalysen, die zu einem vorzeitigen Studienabbruch geführt haben*
- *Prüfen, ob statistisch nicht signifikante Ergebnisse weniger ausführlich dargestellt sind.*
- *Ggf. prüfen, ob „übliche“ Endpunkte nicht berichtet sind.*

Anzumerken ist, dass Anzeichen für eine ergebnisgesteuerte Darstellung eines Endpunkts zu Verzerrungen der Ergebnisse der übrigen Endpunkte führen kann, da dort ggf. auch mit einer selektiven Darstellung gerechnet werden muss. Insbesondere bei Anzeichen dafür, dass die Ergebnisse einzelner Endpunkte selektiv nicht berichtet werden, sind Verzerrungen für die anderen Endpunkte möglich. Eine von der Planung abweichende selektive Darstellung des Ergebnisses eines Endpunkts führt jedoch nicht zwangsläufig zu einer Verzerrung der anderen Endpunkte; in diesem Fall ist die ergebnisgesteuerte Berichterstattung endpunktspezifisch unter Punkt B.3 (siehe unten) einzutragen. Des Weiteren ist anzumerken, dass die Berichterstattung von unerwünschten Ereignissen üblicherweise ergebnisabhängig erfolgt (es werden nur Häufungen / Auffälligkeiten berichtet) und dies nicht zur Verzerrung anderer Endpunkte führt.

- ja:** Eine ergebnisgesteuerte Berichterstattung ist unwahrscheinlich.
- unklar:** Die verfügbaren Angaben lassen eine Einschätzung nicht zu.
- nein:** Es liegen Anzeichen für eine ergebnisgesteuerte Berichterstattung vor, die das Verzerrungspotenzial aller relevanten Endpunkte beeinflusst.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

5. Keine sonstigen (endpunktübergreifenden) Aspekte, die zu Verzerrung führen können

z. B.

- zwischen den Gruppen unterschiedliche Begleitbehandlungen außerhalb der zu evaluierenden Strategien
- intransparenter Patientenfluss
- Falls geplante Interimanalysen durchgeführt wurden, so sind folgende Punkte zu beachten:
 - Die Methodik muss exakt beschrieben sein (z. B. alpha spending approach nach O'Brien Fleming, maximale Stichprobengröße, geplante Anzahl und Zeitpunkte der Interimanalysen).
 - Die Resultate (p-Wert, Punkt- und Intervallschätzung) des Endpunktes, dessentwegen die Studie abgebrochen wurde, sollten adjustiert worden sein (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.

Begründung für die Einstufung:

B Verzerrungsaspekte auf Endpunktebene pro Endpunkt:

Die folgenden Punkte B.1 bis B.4 dienen der Einschätzung der endpunktspezifischen Aspekte für das Ausmaß möglicher Ergebnisverzerrungen. Diese Punkte sollten i. d. R. für jeden relevanten Endpunkt separat eingeschätzt werden (ggf. lassen sich mehrere Endpunkte gemeinsam bewerten, z. B. Endpunkte zu unerwünschten Ereignissen).

Endpunkt: _____

1. Verblindung der Endpunkterheber

Für den Endpunkt ist zu bestimmen, ob das Personal, welches die Zielkriterien erhoben hat, bzgl. der Behandlung verblindet war.

In manchen Fällen kann eine Verblindung auch gegenüber den Ergebnissen zu anderen Endpunkten (z. B. typischen unerwünschten Ereignissen) gefordert werden, wenn die Kenntnis dieser Ergebnisse Hinweise auf die verabreichte Therapie gibt und damit zu einer Entblindung führen kann.

- ja:** Der Endpunkt wurde verblindet erhoben.
- unklar:** Es finden sich keine diesbezüglichen Angaben.
- nein:** Aus den Angaben geht hervor, dass keine verblindete Erhebung erfolgte.

Angaben zum Kriterium; obligate Begründung für die Einstufung:

2. Adäquate Umsetzung des ITT-Prinzips

Kommen in einer Studie Patienten vor, die die Studie entweder vorzeitig abgebrochen haben oder wegen Protokollverletzung ganz oder teilweise aus der Analyse ausgeschlossen wurden, so sind diese ausreichend genau zu beschreiben (Abbruchgründe, Häufigkeit und Patientencharakteristika pro Gruppe) oder in der statistischen Auswertung angemessen zu berücksichtigen (i. d. R. ITT-Analyse, siehe Äquivalenzstudien). Bei einer ITT („intention to treat“)-Analyse werden alle randomisierten Patienten entsprechend ihrer Gruppenzuteilung ausgewertet (ggf. müssen fehlende Werte für die Zielkriterien in geeigneter Weise ersetzt werden). Zu beachten ist, dass in Publikationen der Begriff ITT nicht immer in diesem strengen Sinne Verwendung findet. Es werden häufig nur die randomisierten Patienten ausgewertet, die die Therapie zumindest begonnen haben und für die mindestens ein Post-Baseline-Wert erhoben worden ist („full analysis set“). Dieses Vorgehen ist in begründeten Fällen Guideline-konform, eine mögliche Verzerrung sollte jedoch, insbesondere in nicht verblindeten Studien, überprüft werden. Bei Äquivalenz- und Nichtunterlegenheitsstudien ist es besonders wichtig, dass solche Patienten sehr genau beschrieben werden und die Methode zur Berücksichtigung dieser Patienten transparent dargestellt wird.

- ja:** Eines der folgenden Merkmale trifft zu:
- Laut Studienunterlagen sind keine Protokollverletzer und Lost-to-follow-up-Patienten in relevanter Anzahl (z. B. Nichtberücksichtigungsanteil in der Auswertung < 5%) aufgetreten, und es gibt keine Hinweise (z. B. diskrepante 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 Anzeichen für eine ergebnisgesteuerte Berichterstattung vor.

Angaben zum Kriterium; falls unklar oder nein, obligate Begründung für die Einstufung:

4. Keine sonstigen (endpunktspezifischen) Aspekte, die zu Verzerrungen führen können

z. B.

- *relevante Dateninkonsistenzen innerhalb der oder zwischen Studienunterlagen*
- *unplausible Angaben*
- *Anwendung inadäquater statistischer Verfahren*

- ja**

- nein**

Angaben zum Kriterium; falls nein, obligate Begründung für die Einstufung:

Einstufung des Verzerrungspotenzials der Ergebnisse des Endpunkts (ausschließlich für randomisierte Studien durchzuführen):

Die Einstufung des Verzerrungspotenzials erfolgt unter Berücksichtigung der einzelnen Bewertungen der vorangegangenen endpunktspezifischen Punkte B.1 bis B.4 sowie der Einstufung des Verzerrungspotenzials auf Studienebene. Falls die endpunktübergreifende Einstufung mit „hoch“ erfolgte, ist das Verzerrungspotenzial für den Endpunkt i. d. R. auch mit „hoch“ einzuschätzen. Eine relevante Verzerrung bedeutet hier, dass sich die Ergebnisse bei Behebung der verzerrenden Aspekte in ihrer Grundaussage verändern würden.

- niedrig:** Es kann mit großer Wahrscheinlichkeit ausgeschlossen werden, dass die Ergebnisse für diesen Endpunkt durch die endpunktspezifischen sowie endpunktübergreifenden Aspekte relevant verzerrt sind.

- hoch:** Die Ergebnisse sind möglicherweise relevant verzerrt.

Begründung für die Einstufung:
