

**Biotechnologisch hergestellte
Arzneimittel in der
Zweitlinientherapie bei der
rheumatoiden Arthritis**

**Dokumentation und Würdigung der
Anhörung zum Berichtsplan**

Auftrag: A10-01
Version: 1.0
Stand: 09.05.2011

Impressum

Herausgeber:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Thema:

Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis

Auftraggeber:

Gemeinsamer Bundesausschuss

Datum des Auftrags:

20.05.2010

Interne Auftragsnummer:

A10-01

Anschrift des Herausgebers:

Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen

Dillenburger Str. 27

51105 Köln

Tel.: +49 221 35685-0

Fax: +49 221 35685-1

Berichte@iqwig.de

www.iqwig.de

Inhaltsverzeichnis

	Seite
Abkürzungsverzeichnis.....	v
1 Dokumentation der Anhörung.....	1
2 Würdigung der Anhörung.....	2
2.1 Hintergrund.....	2
2.1.1 Ziele der Behandlung	2
2.1.2 Therapie der rheumatoiden Arthritis und Leitlinienempfehlungen.....	2
2.1.3 Begriffe „Erstlinien- und Zweitlinientherapie“	3
2.2 Ziele der Untersuchung	3
2.2.1 Definition Zweitlinientherapie	3
2.3 Kriterien für den Einschluss von Studien in die Untersuchung.....	5
2.3.1 MTX im Rahmen einer Kombinationstherapie oder als Vergleichsintervention	5
2.3.2 Patientenrelevante Endpunkte	6
2.3.2.1 Festlegung von Skalen und kombinierten Endpunkten	6
2.3.2.2 Krankheitsaktivität und Messung der Remission	7
2.3.2.3 Symptomatik der Entzündung	10
2.3.2.4 Strukturelle Gelenkveränderungen – Radiografische Progression	11
2.3.2.5 Soziales Funktionsniveau	13
2.3.2.6 Gesundheitsbezogene Lebensqualität	14
2.3.2.7 Unerwünschte Arzneimittelwirkungen.....	14
2.3.2.8 Interventions- und erkrankungsbedingter Aufwand	15
2.3.3 Studientypen	16
2.3.4 Studiendauer	17
2.3.5 Vorgehen im Falle einer Zulassungsänderung im Projektverlauf	19
2.3.6 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen.....	20
2.4 Informationsbeschaffung.....	20
2.4.1 Bibliografische Recherche.....	21
2.4.2 Selektion relevanter Studien.....	21
2.5 Informationsbewertung.....	22
2.6 Informationssynthese und -analyse.....	23

2.6.1	Patientenrelevante Endpunkte	23
2.6.2	Zusammenfassen von Ergebnissen.....	23
2.6.3	Relevanzschwelle	23
2.6.4	Indirekte Vergleiche	24
2.6.5	Gegenüberstellung der Ergebnisse der Einzelstudien	24
2.6.6	Meta-Analysen	26
2.6.7	Subgruppenmerkmale und andere Effektmodifikatoren	27
2.7	Literaturverzeichnis.....	29
3	Offenlegung potenzieller Interessenkonflikte.....	35
3.1	Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen	35
3.2	Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (externe Sachverständige)	36
4	Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll	38
4.1	Teilnehmerliste der wissenschaftlichen Erörterung.....	38
4.2	Tagesordnung der wissenschaftlichen Erörterung.....	39
4.3	Protokoll der wissenschaftlichen Erörterung.....	40
4.3.1	Begrüßung und Einleitung.....	40
4.3.2	Tagesordnungspunkt 1: Definition der Zweitlinientherapie	42
4.3.3	Tagesordnungspunkt 2: Endpunkte, a) Teilminderungen der Krankheitsaktivität (wie ACR 20, ACR 50, ACR 70)	49
4.3.4	Tagesordnungspunkt 2: Endpunkte, b) Radiologischer Verlauf	60
4.3.5	Tagesordnungspunkt 3: Effektmodifikatoren.....	63
4.3.6	Tagesordnungspunkt 4: Verschiedenes	66
5	Anhang: Dokumentation der Stellungnahmen.....	75

Abkürzungsverzeichnis

Abkürzung	Bedeutung
ACR	American College of Rheumatology
bDMARD	Biotechnologisch hergestellte Disease-Modifying Antirheumatic Drug
BIOBADASER	Spanish Registry of Adverse Events of Biological Therapies in Rheumatic Diseases
BSRBR	British Society for Rheumatology Biologics Register
DAS	Disease Activity Score
DGRh	Deutsche Gesellschaft für Rheumatologie
DMARD	Disease-Modifying Antirheumatic Drug
EMA	European Medicines Agency
FACIT-F	Functional Assessment of Chronic Illness Therapy-Fatigue
FDA	Food and Drug Administration
FEM	Fixed Effects Model (Modell mit festen Effekten)
G-BA	Gemeinsamer Bundesausschuss
GREES	Group for the Respect of Ethics and Excellence in Science
HAQ	Health Assessment Questionnaire
HTA	Health Technology Assessment
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
ITT	Intention-to-Treat
MCID	Minimal Clinical Important Differences
MTX	Methotrexat
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
RA	Rheumatoide Arthritis
RCT	Randomised Controlled Trial (randomisierte kontrollierte Studie)
REM	Random Effects Model (Modell mit zufälligen Effekten)
sDMARD	Synthetisch hergestellte Disease-Modifying Antirheumatic Drug
SF-36	Short Form 36
SGB V	Sozialgesetzbuch V
SMD	Mittelwertdifferenz
SSATG	South Swedish Arthritis Treatment Group
TNF	Tumornekrosefaktor
UAW	Unerwünschte Arzneimittelwirkung

1 Dokumentation der Anhörung

Am 28.09.2010 wurde der (vorläufige) Berichtsplan in der Version Nr. 1.0 vom 20.09.2010 veröffentlicht und zur Anhörung gestellt. Bis zum 26.10.2010 konnten schriftliche Stellungnahmen eingereicht werden. Insgesamt wurden 9 Stellungnahmen form- und fristgerecht abgegeben. Diese Stellungnahmen sind im Anhang abgebildet.

Unklare Aspekte in den schriftlichen Stellungnahmen wurden in einer wissenschaftlichen Erörterung am 17.02.2011 im IQWiG diskutiert. Das Wortprotokoll der Erörterung befindet sich in Kapitel 4.

Eine Würdigung der in der Anhörung vorgebrachten Aspekte befindet sich in Kapitel 2. Im überarbeiteten Berichtsplan sind darüber hinaus Änderungen, die sich durch die Anhörung ergeben haben, zusammenfassend dargestellt. Der überarbeitete Berichtsplan ist auf der Website des IQWiG unter www.iqwig.de veröffentlicht.

2 Würdigung der Anhörung

Die im Rahmen der Anhörung vorgebrachten Aspekte wurden hinsichtlich valider wissenschaftlicher Argumente für eine Änderung des Berichtsplans überprüft. Die wesentlichen Argumente werden im Folgenden diskutiert. Neben projektspezifischen wissenschaftlichen Aspekten wurden auch übergeordnete Punkte, z. B. zu rechtlichen Vorgaben für das Institut, angesprochen. Auf solche Punkte wird im Rahmen dieser projektspezifischen Würdigung der Anhörung nicht weiter eingegangen.

2.1 Hintergrund

2.1.1 Ziele der Behandlung

In einer Stellungnahme wurde angemerkt, dass im vorläufigen Berichtsplan beschrieben worden sei, die Gelenkzerstörung könne nur während einer Remission der Krankheitsaktivität gehemmt werden. Damit verbunden war die Anregung, an dieser Stelle zu beschreiben, dass die „klinische Remission bzw. niedrige Krankheitsaktivität und die Hemmung der Gelenkzerstörung zwei unabhängige Therapieziele“ seien. Es wurde des Weiteren vermerkt, dass diese sicherlich zusammenhängen.

Der Absatz, den dieses Argument adressierte, bezog sich ausschließlich auf den Einfluss der Krankheitsaktivität, nicht auf alle denkbaren Einflüsse auf die Gelenkzerstörung. Die Formulierung, dass „nur eine Remission (...) der Krankheitsaktivität das Fortschreiten der Gelenkzerstörung aufhalten kann“, bezog sich darauf, dass durch „Teilminderungen der Krankheitsaktivität“ dies unklar ist, wie im weiteren Verlauf des Abschnitts beschrieben und begründet ist.

Im Berichtsplan wurde der Absatz gestrichen, in dem beschrieben worden war, dass Untersuchungen gezeigt hätten, dass nur eine Remission der Krankheitsaktivität das Fortschreiten der Gelenkzerstörung aufhalten kann. Es soll vermieden werden, dass der Eindruck erweckt werden könnte, die in diesem Zusammenhang zitierten Arbeiten könnten ausreichen, um einen derartigen kausalen Zusammenhang zu begründen.

2.1.2 Therapie der rheumatoiden Arthritis und Leitlinienempfehlungen

In 2 Stellungnahmen wurde darauf hingewiesen, dass im vorläufigen Berichtsplan nicht die aktuellsten Fachinformationen zitiert worden seien. Da die Fachinformationen sich in relativ kurzen Zeiträumen ändern, ist es nicht möglich, den Hintergrund diesbezüglich auf dem aktuellsten Stand zu halten. Um dies aber zu kennzeichnen, wurde jeweils das entsprechende Datum der zitierten Fachinformationen hinzugefügt. Im Rahmen der Vor- und Abschlussberichtserstellung werden die Fachinformationen regelmäßig auf Änderungen überprüft, sodass jeweils die aktuellste Version als Grundlage dient.

Dies wurde im Berichtsplan angepasst und für alle weiteren Fachinformationen ebenfalls überprüft und bei Bedarf überarbeitet.

2.1.3 Begriffe „Erstlinien- und Zweitlinientherapie“

Eine Stellungnahme wies darauf hin, dass der vorläufige Berichtsplan die Begriffe „Erstlinientherapie“ und „Zweitlinientherapie“ zwar diskutiere, aber für den vorliegenden Auftrag nicht definiere. Dies ist nicht richtig, da sich die Definition im 2. Kapitel „Ziele der Untersuchung“ findet. Die im Rahmen der Stellungnahme vorgeschlagene Definition wurde an anderer Stelle diskutiert (siehe hierzu Abschnitt 2.2.1).

Hieraus ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.2 Ziele der Untersuchung

2.2.1 Definition Zweitlinientherapie

Der Begriff der Zweitlinientherapie wird im Bereich der rheumatoiden Arthritis nicht einheitlich verwendet, was in Stellungnahmen bestätigt wurde. Der vorläufige Berichtsplan enthält die Definition, dass darunter der Einsatz von biotechnologisch hergestellten Arzneimitteln bei Personen zu verstehen ist, die mit einem krankheitsmodifizierenden Antirheumatikum vorbehandelt sind und dass jeweils der erste Einsatz des zu untersuchenden Arzneimittels gemäß Zulassungsstatus zu betrachten ist.

Mehrere Stellungnahmen nahmen Bezug auf diese Definition – zumeist hinsichtlich des Kapitels „Ziele der Untersuchung“, da sich dort die Definition des Begriffs der Zweitlinientherapie findet. Eine Stellungnahme bezog ihr Argument auf den Abschnitt „Prüf- und Vergleichsintervention“, adressierte aber ebenfalls die Definition der Zweitlinientherapie, weshalb dies hier zusammen mit den anderen Argumenten gewürdigt wird.

Die Argumente der Stellungnahmen waren recht heterogen, sodass diese im Folgenden zunächst aufgeführt werden:

- Unter Zweitlinientherapie sei der Einsatz der biotechnologisch hergestellten krankheitsmodifizierenden antirheumatischen Arzneimittel (bDMARDs) nach einem oder mehreren sogenannten konventionellen Disease-Modifying Antirheumatic Drugs (krankheitsmodifizierenden antirheumatischen Arzneimittel, DMARDs) zu verstehen.
- Es bleibe unklar, was genau mit der Definition des vorläufigen Berichtsplans gemeint sei – der Einsatz eines bDMARDs als zweiten Schritt nach einer Methotrexat(MTX)-Monotherapie oder dessen Einsatz, nachdem bereits sowohl eine MTX-Monotherapie als auch eine Kombinationstherapie aus MTX und einem synthetisch hergestellten krankheitsmodifizierenden antirheumatischen Arzneimittel (sDMARD) erfolgt sei, oder ob vielleicht beide Möglichkeiten gemeint seien.

- Der Begriff der Zweitlinientherapie bedeute, dass dem Einsatz eines bDMARDs, eine Therapie mit mindestens einem krankheitsmodifizierenden Antirheumatikum vorausgegangen sein müsse.
- Die getroffene Definition sei unzureichend, da die relevante Population und Indikation des Arzneimittels exakt zu definieren seien und darüber hinaus für verschiedene Arzneimittel der Zulassungsstatus unterschiedlich sei. Diese Unterschiede im Zulassungsstatus führten zu einem Verzerrungspotenzial, das sich bis in die Informationssynthese durchsetzen könne.
- Eine Behandlung mit bDMARDs, die auf eine bereits durchgeführte Therapie mit einem anderen bDMARD folge, sei als Drittlinientherapie zu bezeichnen.
- Rituximab (MabThera®) stelle kein Präparat der Zweitlinientherapie, sondern der Drittlinientherapie dar. Begründet wurde dies damit, dass die Zweitlinientherapie im Berichtsplan als „erster Einsatz des zu untersuchenden Arzneimittels einer Behandlung mit einem sDMARD“ definiert sei. Die Zulassung von Rituximab (MabThera®) beinhalte aber das Versagen einer Tumornekrosefaktor(TNF)-Therapie. Zur Stützung dieser Anmerkung wurde die Arbeit von Wollenhaupt et al. [1] zitiert, welche laut Stellungnahme Therapierichtlinien beschreibe und das in der Stellungnahme beschriebene Vorgehen bestätige.
- Für den vorliegenden Auftrag ist die Zweitlinientherapie als Therapie nach Versagen einer Vorbehandlung von der Erstlinientherapie mit einem bDMARD als Therapie von behandlungsnaiven Patienten abzugrenzen. Die im vorläufigen Berichtsplan getroffene Formulierung „mit einem krankheitsmodifizierenden Antirheumatikum vorbehandelt“ ist dahin gehend missverständlich, weil ihr entnommen werden könnte, dass die Vorbehandlung auf ausschließlich ein einziges krankheitsmodifizierendes Antirheumatikum beschränkt ist. Tatsächlich umfasst die verwendete Definition die Vorbehandlung mit mehreren DMARDs, darunter sowohl sDMARDs als auch bDMARDs.
- Die geäußerte Befürchtung, der unterschiedliche Zulassungsstatus verschiedener Arzneimittel führe in der Informationssynthese zu einem Verzerrungspotenzial, ist nicht gerechtfertigt. Aufgrund eines unterschiedlichen Zulassungsstatus ist eine Verzerrung, eine systematische Abweichung eines Ergebnisses von der „Wahrheit“ [2], nicht zu erwarten. Der unterschiedliche Zulassungsstatus könnte vielmehr zu heterogenen Ergebnissen der Einzelstudien führen, sodass auch eine mögliche metaanalytische Zusammenfassung dieser Ergebnisse davon beeinflusst wäre. Im Rahmen von IQWiG-Bewertungen wird daher die Heterogenität der Studienergebnisse regelhaft untersucht: Zunächst wird vor einer meta-analytischen Zusammenfassung über deren Sinnhaftigkeit entschieden. Darüber hinaus würde beispielsweise, wie im Berichtsplan beschrieben, bei hoher Heterogenität (p -Wert des Interaktionstests $< 0,2$) auf die Berechnung eines gemeinsamen Schätzers verzichtet.

- Eine weitere Spezifizierung der Indikation und relevanten Population ist aus Sicht des IQWiG nicht notwendig, da beides durch den Zulassungsstatus für die vorliegende Fragestellung festgelegt ist.
- Gerade weil der Zulassungsstatus bei verschiedenen Arzneimitteln unterschiedlich ist, wurde für den vorliegenden Auftrag eine Formulierung gewählt, die für alle Arzneimittel gültig ist. Der Zulassungsstatus ist für jede Intervention dann zu prüfen.
- Zum Zitat Wollenhaupt et al. ist zu ergänzen, dass die Autoren in ihrem Artikel explizit beschreiben, dass es sich bei ihrer vorgeschlagenen Therapiestrategie nicht um eine Empfehlung handle – damit also auch nicht um „Therapierichtlinien“ –, sondern um die „Darstellung einer Konsensusentscheidung einer erweiterten Gruppe erfahrener Rheumatologen“.
- Auch aus der wissenschaftlichen Erörterung am 17.02.2011 gingen keine weiteren Argumente hervor, die das Vorgehen des IQWiG als nicht sinnvoll erscheinen lassen. Es bestand Einvernehmen darüber, dass unterschiedliche Populationen auch getrennt voneinander betrachtet werden müssen.
- Die Formulierung zur Definition der Zweitlinientherapie wurde angepasst, um das Vorgehen beim vorliegenden Auftrag klarer zu machen.

2.3 Kriterien für den Einschluss von Studien in die Untersuchung

2.3.1 MTX im Rahmen einer Kombinationstherapie oder als Vergleichsintervention

Es wurde darauf hingewiesen, dass aufgrund des Zulassungsstatus für verschiedene bDMARDs auch verschiedene Kombinationen dieser Arzneimittel mit sDMARD möglich seien. So sei zum Beispiel die Kombination nicht nur mit MTX möglich, auch wenn dies in den Fachinformationen nicht explizit beschrieben sei. Studien, in denen Kombinationen aus bDMARDs und anderen sDMARDs als MTX eingesetzt worden seien, dürften daher nicht ausgeschlossen werden.

Einige Stellungnahmen beschrieben außerdem, welche Dosierungen von MTX bei der Bearbeitung des Auftrags berücksichtigt werden sollten und welche nicht. Ein Grund für die differenzierte Berücksichtigung sei, Verzerrungen zu vermeiden.

Für beide Argumente gilt, dass der Einschluss von Studien gemäß dem Zulassungsstatus erfolgen wird – dies betrifft auch zugelassene Kombinationen und Dosierungen. Eine Möglichkeit zur Stellungnahme hinsichtlich des vorgenommenen Einschlusses von Studien besteht nach der Veröffentlichung des Vorberichts.

Verzerrungen in der Informationssynthese aufgrund unterschiedlicher Dosierungen sind nicht zu befürchten, da auch hierdurch allenfalls Heterogenität entstehen kann (siehe hierzu die

Ausführungen in Abschnitt 2.2.1). Unterschiedliche Dosierungen sind in der Informationssynthese inhaltlich sinnvoll zu berücksichtigen.

Zusammenfassend ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.3.2 Patientenrelevante Endpunkte

Mehrere Stellungnahmen bezogen sich auf die Definition der patientenrelevanten Endpunkte.

2.3.2.1 Festlegung von Skalen und kombinierten Endpunkten

In mehreren Stellungnahmen wurde angemerkt, dass vorab definiert werden müsse, welche Instrumente im Rahmen der Bewertung berücksichtigt werden sollen. Diese Forderung betraf Skalen zur Lebensqualitäts- und Funktionsmessung sowie kombinierte Endpunkte zur Messung der Remission der Krankheitsaktivität. So wurde beispielsweise gefordert, die Definition der Remissionskriterien wie für den Disease Activity Score (DAS)28, der bei $< 2,6$ liegt, oder Minimal Clinical Important Differences (MCID) für den Health Assessment Questionnaire (HAQ) explizit zu benennen.

In einer Stellungnahme wurde angemerkt, dass wissenschaftliches Arbeiten bedeutet, die Methoden vor Kenntnis der Daten festzulegen.

Im Gegensatz zu einer Studie mit prospektiv geplanter Datenerhebung ist zum Zeitpunkt der Planung einer systematischen Übersicht die Datenstruktur der verfügbaren Daten im Sinne einer Zielgrößendefinition (z. B. hinsichtlich der verwendeten Messinstrumente beziehungsweise Operationalisierungen) für alle einzuschließenden Studien nicht bekannt. Der Versuch einer Aufzählung aller möglichen oder wahrscheinlich verwendeten Messinstrumente und Operationalisierungen von Endpunkten in den zu bewertenden Studien ist deshalb nicht zielführend. Es ergibt sich vielmehr die Notwendigkeit, mit den in der Recherche identifizierten Datenstrukturen umzugehen.

Die Entscheidung über den Einschluss von Messinstrumenten für die im Berichtsplan präspezifizierten Endpunkte erfolgt auf Basis der Methodik des IQWiG [2]. So müssen die Messinstrumente z. B. patientenrelevante Konstrukte abbilden. Das Methodenpapier beschreibt auch Anforderungen an patientenberichtete Endpunkte, die z. B. bei der Untersuchung von Nutzendimensionen wie der gesundheitsbezogenen Lebensqualität oder bei der Messung von Symptomen zum Einsatz kommen können. Die Definition solcher Kriterien für den Einschluss von Messinstrumenten wird der Durchführung einer systematischen Übersicht eher gerecht als der Versuch, alle möglicherweise zu bewertenden Instrumente bereits im Vorfeld zu definieren.

Hinsichtlich der zu verwendenden Definitionen der Remission wurde bereits im Hintergrund des Berichtsplans darauf hingewiesen, dass für den Auftrag A10-01 der Begriff „Remission“

im „Sinne der Definitionen verschiedener Messinstrumente verwendet“ wird. In einer Stellungnahme wurde das Instrument American College of Rheumatology(ACR) 70 in diesem Zusammenhang genannt. Dieser ist allerdings nicht als Remissionskriterium definiert und somit nicht Gegenstand der Untersuchung der Remission der Krankheitsaktivität. Hinsichtlich der MCIDs gilt, dass Responderanalysen auf Basis von validierten MCIDs, die entsprechende methodische Anforderungen erfüllen, berücksichtigt werden.

Zusammenfassend ergab sich hieraus keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.3.2.2 Krankheitsaktivität und Messung der Remission

Mehrere Stellungnahmen bezogen sich auf die Krankheitsaktivität. Dabei wurde angemerkt, dass Teilminderungen der Krankheitsaktivität, wie sie durch den ACR 20, ACR 50 oder ACR 70 beschrieben werden, als patientenrelevante Endpunkte zu berücksichtigen seien. Es wurde auch auf die Patientenrelevanz von Messinstrumenten zur Erhebung der Remission eingegangen.

Teilminderungen der Krankheitsaktivität

Eine Stellungnahme kritisierte, dass die Krankheitsaktivität sich nicht in der Definition der patientenrelevanten Endpunkte wiederfinde.

Das ist nicht richtig, da dort „Remission“ aufgeführt ist.

In Stellungnahmen wurde beschrieben, dass Teilminderungen der Krankheitsaktivität in der Nutzenbewertung berücksichtigt werden sollten, da sie ein guter „Prädiktor für die zu erwartende Gelenkschädigung und die Funktionseinschränkung“ seien.

Keine der in diesem Zusammenhang zitierten Publikationen [3-11] beschreibt eine hierzu durchgeführte randomisiert-kontrollierte Studie (RCT) oder eine systematische Übersichtsarbeit, die geeignet wäre, diese Frage zu klären. Die zitierten Arbeiten liefern allenfalls Anzeichen dafür, dass die Veränderung der Krankheitsaktivität und die der Gelenkzerstörung miteinander einhergehen, d. h. dass gewisse Korrelationen bestehen. Wie auch an anderer Stelle des vorliegenden Dokuments (siehe unten) zur radiologischen Progression beschrieben, scheinen diese Studien einem besseren Verständnis des Krankheitsbildes und einem Erkenntnisgewinn zum gegenseitigen Einfluss dieser Prozesse zu dienen. Dies ist insbesondere für die Entwicklung von Medikamenten und Therapiestrategien wichtig. Dies wurde von einer Stellungnahme insofern bestätigt, als sie darauf hinwies, dass die Messung der Teilminderungen der Krankheitsaktivität, wie in Leitlinien beschrieben, für die Verlaufskontrolle empfohlen sei oder wie von der European Medicines Agency (EMA) zur Medikamentenentwicklung heranzuziehen sei. Diese beiden Aspekte sind nicht Gegenstand der vorliegenden Fragestellung.

Eine dem IQWiG genannte Veröffentlichung des ACR Committee [12] führt in dieser Frage nicht weiter, weil hier nicht untersucht wurde, welchen Stellenwert eine Teilminderung der Krankheitsaktivität (hier handelt es sich konkret um den ACR 20) im Hinblick auf patientenrelevante Endpunkte hat.

Es wurde in Stellungnahmen angemerkt, dass aufgrund regulatorischer Vorgaben der ACR 20 in Studien als primärer Endpunkt definiert sei und dies nicht zum Ausschluss dieser Studien führen dürfe. Zudem seien aufgrund dieser regulatorischen Vorgaben der ACR 20, ACR 50 und ACR 70 in der Nutzenbewertung des IQWiG zu berücksichtigen. Eine Stellungnahme beschrieb aber auch, dass der ACR 20 hinsichtlich seiner Patientenrelevanz kontrovers diskutiert werden könne. Es wurde in einer Stellungnahme eine stärkere Konsistenz zwischen den Vorgehensweisen regulatorischer Behörden und dem IQWiG gefordert.

Es ist nicht zu befürchten, dass eine Studie aufgrund der Untersuchung des ACR 20 als primären Endpunkt aus der Bewertung des IQWiG ausgeschlossen wird. Ergebnisse zu im Berichtsplan definierten Endpunkten werden selbstverständlich in der Bewertung berücksichtigt werden, unabhängig davon, welche primären Endpunkte die jeweiligen Studien haben. In die Bewertungen des IQWiG fließen auch sekundäre Endpunkte ein. Eine Inkonsistenz der Vorgehensweisen zwischen regulatorischen Behörden und dem IQWiG wird nicht gesehen, da die jeweiligen Vorgehensweisen von den gesetzlich festgeschriebenen Aufgaben und den zu bearbeitenden Fragestellungen abhängen.

Als weiteres Argument für die Berücksichtigung von Teilminderungen der Krankheitsaktivität wiesen Stellungnahmen darauf hin, dass die Remissionskriterien des ACR oder DAS sehr harte Kriterien seien und die Ansprechraten daher eher gering ausfielen. Da Remissionskriterien in der Fallzahlplanung nicht berücksichtigt worden seien, sei die Güte für die Analysen gegebenenfalls zu klein, um Effekte zu entdecken.

Das Argument der unzureichenden Power aufgrund des zugrunde liegenden primären Endpunkts mag auf eine Einzelstudie zutreffen. Gerade darin liegt der Vorteil von systematischen Reviews und Meta-Analysen, wie sie vom IQWiG erstellt werden: Es können Daten mehrerer Studien meta-analytisch zusammengefasst werden, sodass das Problem der fehlenden Power minimiert wird.

Ein weiteres Argument der Stellungnahmen für die Berücksichtigung von Teilminderungen, die als relative Änderungen dargestellt werden, war, dass Ergebnisse des DAS stark von den Ausgangswerten zu Baseline abhingen, weil hier absolute Werte wiedergegeben würden. Hierdurch könne es zu „Verzerrungen“ in den Ergebnissen kommen. Dies sei bei den ACR-Responsekriterien anders, weil es sich hierbei um relative Werte handele.

Es wird kein Problem darin gesehen, Ergebnisse von Scores zu berücksichtigen, die absolute Werte liefern. Innerhalb einer Studie stellt dies kein Problem dar, da im Rahmen eines RCT diesbezüglich eine Strukturgleichheit in den Gruppen herrscht. In der metaanalytischen Zu-

sammenfassung und der Interpretation der Daten ist dies selbstverständlich ein Aspekt, der zu berücksichtigen ist. Allerdings sind hierbei keine Verzerrungen der Ergebnisse zu befürchten. Vielmehr könnten heterogene Ergebnisse das Resultat von solchen Zusammenfassungen sein.

Auch in der wissenschaftlichen Erörterung am 17.02.2011 konnte die Frage nicht geklärt werden, warum Teilminderungen der Krankheitsaktivität als patientenrelevant angesehen werden. Es wurden über die in den Stellungnahmen beschriebenen Argumente hinaus keine weiteren ausgetauscht.

Die Krankheitsaktivität ist, wie auch weiter unten beschrieben, ein Konstrukt, mit dem das komplexe Krankheitsgeschehen der rheumatoiden Arthritis abgebildet wird (siehe Absatz „Messung der Remission von Krankheitsaktivität“). Aus diesem Grund ist die Remission, also die Reduktion der Krankheitsaktivität auf ein Niveau, auf dem die Patienten weitgehend frei von Krankheitssymptomen sind und der Krankheitsverlauf maßgebend beeinflusst wird, ein patientenrelevanter Endpunkt. Hingegen ist der Stellenwert von Teilminderungen der Krankheitsaktivität für den Verlauf der rheumatoiden Arthritis bzw. den Behandlungserfolg über die Verbesserung der Symptomatik hinaus, unklar. Das IQWiG verneint nicht, dass Teilminderungen der Krankheitsaktivität von den Patienten als positive Veränderungen der Symptomatik wahrgenommen werden können. Diese Änderungen können über weitere, im Berichtsplan definierte patientenrelevante Endpunkte, wie Schmerz oder Lebensqualität, erfasst werden. So ist nicht zu befürchten, dass positive Auswirkungen für Patienten durch Teilminderungen aus der Bewertung ausgeschlossen werden. Aus diesem Grund wurden Teilminderungen der Krankheitsaktivität nicht als patientenrelevanter Endpunkt definiert.

Zusammenfassend ergab sich hieraus keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

Messung der Remission von Krankheitsaktivität

Im vorläufigen Berichtsplan wurde beschrieben, dass kombinierte Endpunkte mit Surrogatparametern betrachtet werden, falls diese anhand der gegebenen Daten validiert werden können. Zum Zeitpunkt der Erstellung des Berichtsplans wurden die Maße für die Krankheitsaktivität als kombinierte Endpunkte angesehen. In den Stellungnahmen wurde danach gefragt, wie eine solche Validierung durchgeführt werden soll.

Nach Sichtung der Stellungnahmen und der beigefügten Literatur sowie nach einer Diskussion zur Definition und zum Umgang mit kombinierten Endpunkten wurde deutlich, dass es sich bei der Krankheitsaktivität nicht um einen kombinierten Endpunkt handelt. Die Krankheitsaktivität ist vielmehr ein Konstrukt, mit dem das komplexe Krankheitsgeschehen der rheumatoiden Arthritis abgebildet werden soll. Zur Messung der Krankheitsaktivität wurden verschiedene Messinstrumente (z. B. der DAS 28) entwickelt, die die verschiedenen Teilaspekte der Erkrankung abdecken.

Hieraus ergaben sich für den Berichtsplan die folgenden Anpassungen: Im Abschnitt „Hintergrund“ wird die Krankheitsaktivität nicht als kombinierter Endpunkt bezeichnet. Im Abschnitt zu den patientenrelevanten Endpunkten wurde der Satz „Kombinierte Endpunkte mit Surrogatparametern werden betrachtet, falls diese anhand der gegebenen Daten validiert werden können“ gestrichen.

Zur Auflösung eines Missverständnisses sei an dieser Stelle explizit darauf hingewiesen, dass sich der entfernte Satz ausschließlich auf die Erhebung der Remission der Krankheitsaktivität bezog. Diese Information fand sich nicht explizit in der Formulierung. Dies führte zu dem Missverständnis, auch Messinstrumente, wie der HAQ, Short Form 36 (SF-36) oder Functional Assessment of Chronic Illness Therapy-Fatigue (FACIT-F) seien darunter subsumiert.

2.3.2.3 Symptomatik der Entzündung

2 Stellungnahmen bezogen sich – ohne Literaturbelege für ihre Aussagen – auf den Endpunkt „Symptomatik der Entzündung“. Es wurde angemerkt, dass die Gelenkschwellung / -überwärmung und der Gelenkschmerz als patientenrelevante Endpunkte aufzunehmen seien. Diese wurden in der Stellungnahme als Hauptsymptomatik einer aktiven Entzündung beschrieben. Darüber hinaus beschrieb eine der Stellungnahmen, dass der Schmerz nicht nur an die Entzündung gekoppelt sei.

Der im vorläufigen Berichtsplan verwendete Begriff „Fatigue“ sei ein unscharfer Begriff – Abgeschlagenheit, Erschöpfung, Müdigkeit und Schlaflosigkeit seien zu ergänzen. Oftmals seien dies auch Symptome der immunologischen Aktivität ohne „wesentliche Ausprägung der entzündlichen Kardinalsymptome“.

Das IQWiG sieht die genannten Aspekte in der Definition der Endpunkte im vorläufigen Berichtsplan abgebildet. So heißt der Endpunkt dort „Symptomatik der Entzündung (insbesondere Schmerz, Fatigue, Morgensteifigkeit)“. Das Wort „insbesondere“ weist darauf hin, dass die 3 in der Klammer genannten Symptome nicht ausschließlich betrachtet werden, sondern verdeutlichen sollen, was hierunter zu verstehen ist.

Die Aspekte, die laut Stellungnahme zur Fatigue zu ergänzen sind, sieht das IQWiG unter dem Begriff „Fatigue“ subsumiert. Dieser ist in der internationalen Literatur üblich, um den in der Stellungnahme genannten Komplex zu beschreiben [13-15]. Fragebögen, mit denen die Fatigue erfasst wird, wie zum Beispiel die Lee Fatigue Scale [16] oder der FACIT-F [17], umfassen die in der Stellungnahme genannten Aspekte.

Inhaltlich ergab sich hieraus keine Änderung für den Berichtsplan. Um aber mögliche Missverständnisse zu vermeiden, wurde aufgrund der Argumente die Formulierung in „Symptomatik der rheumatoiden Arthritis (insbesondere Schmerz, Fatigue, Morgen-

steifigkeit)“ geändert. Damit soll verdeutlicht werden, dass die Symptomatik sowohl von der akuten Inflammation als auch von dem morphologischen Schaden beeinflusst wird.

2.3.2.4 Strukturelle Gelenkveränderungen – Radiografische Progression

Mehrere Stellungnahmen bezogen sich auf die Berücksichtigung der radiologischen Progression in der Bewertung des IQWiG.

Eine Stellungnahme hinterfragte zunächst, ob Gelenkversteifungen tatsächlich ein patientenrelevanter Endpunkt bei an rheumatoider Arthritis (RA) Erkrankten seien. Dies treffe eher auf die Erkrankung Morbus Bechterew als auf die RA zu.

Im Berichtsplan sind nicht Versteifungen der Wirbelsäule gemeint, wie sie beim Morbus Bechterew auftreten. Bei der RA kommt es zu einer Versteifung der Gelenke. Daher ist dies ein patientenrelevanter Endpunkt für die vorliegende Fragestellung.

In mehreren Stellungnahmen wurde beschrieben, dass die radiografische Progression bzw. die radiologische Erfassung von Gelenkveränderungen als patientenrelevanter Endpunkt zu betrachten sei und daher vom IQWiG als solcher aufgenommen werden solle.

So sei die radiologische Erfassung von Gelenkveränderungen in fast jeder Studie als primärer oder sekundärer Endpunkt untersucht worden, was für den im Berichtsplan definierten Endpunkt der „strukturellen Gelenkveränderungen (Deformitäten, Versteifungen, Kontrakturen)“ nicht der Fall sei. Außerdem sei durch verschiedene Arbeiten gezeigt worden, dass zwischen radiologisch nachgewiesenem Gelenkschaden und der Verschlechterung der Gelenkfunktion ein Zusammenhang bestehe bzw. diese beiden Ereignisse miteinander einhergingen (van der Heijde et al. 2008 [18], Breedveld et al. 2005 [19]). Ebenso bestehe ein solcher Zusammenhang mit der Lebensqualität und der verringerten Teilhabe am beruflichen Leben (Kavanaugh et al. 2004 [20]). Der Sharp Score, eine der Methoden zur Erfassung der radiologischen Veränderung, sei eine unabhängige „Determinante der Funktionalität“ (Drossaers-Bakker et al. 1999 [6], Scott et al. 2000 [21], Welsing et al. 2001 [22], Aletaha et al. 2006 [23], Odegard et al. 2006 [24]). Es wurde auch darauf hingewiesen, dass es Definitionen für klinisch relevante Veränderungen gebe [25-27].

Für die Bewertung des IQWiG ist nicht ausschlaggebend, welche Endpunkte in Studien als primär oder sekundär definiert sind, sondern ob sie patientenrelevant sind.

Eine Validität des Surrogats „radiologische Veränderungen“ für die Funktionsverschlechterung kann aus den zitierten Arbeiten nicht abgeleitet werden. Vielmehr ergeben sich aus den Untersuchungen Anzeichen dafür, dass die Gelenkzerstörung mit der Funktionsverschlechterung „einhergeht“. So liefern bspw. Drossaers-Bakker et al. [6] keine Anzeichen dafür, dass die frühzeitige radiologische Erfassung Aussagen zur irreversiblen Gelenkschäden und damit zu Einbußen in der Funktionalität erlaubt. Der Zusammenhang, gemessen über

Spearman's Korrelationskoeffizienten, zeigt mit dem Wert 0,6 zwar ein mittleres Ausmaß der Korrelation, aber keinen (ausreichend) starken Zusammenhang. Ein Zusammenhang zwischen Behandlungseffekten wurde darüber hinaus nicht gezeigt. Das heißt, hier ist allenfalls eine Korrelation zwischen der Gelenkzerstörung und der Funktionsverschlechterung abzuleiten. Die Methoden 3.0 des IQWiG [2] äußern sich zu den Voraussetzungen, die erfüllt sein müssen, um einen Surrogatendpunkt im Rahmen der Nutzenbewertung zu berücksichtigen. Mit geeigneten statistischen Methoden muss gezeigt worden sein, „dass der Effekt einer Intervention mit vergleichbarem Wirkmechanismus auf den zu ersetzenden patientenrelevanten Endpunkt durch den Effekt auf den Surrogatendpunkt in einem ausreichenden Maß erklärt“ wurde [28,29]. Benötigt werden hierfür in der Regel Belege aus Interventionsstudien für einen „plausiblen, starken, konsistenten und gleichgerichteten Zusammenhang zwischen der Änderung des Surrogates und der Änderung des patientenrelevanten Endpunkts“. Ein solcher Zusammenhang muss durch die Korrelation von Effektschätzern von Surrogat und klinischem Endpunkt nachgewiesen werden, wozu grundsätzlich Daten mehrerer Studien einzubeziehen sind. Eine Validierung gemäß den Anforderungen, wie sie in der biometrischen Literatur [29-31] diskutiert werden, erfolgt damit nicht. Die in den Stellungnahmen zitierten Arbeiten, die gezeigt haben sollen, dass radiologische Veränderungen eine unabhängige Determinante der körperlichen Funktionalität sind, haben auch den Einfluss der Krankheitsaktivität auf die körperliche Funktion untersucht [6,21-24]. Diese Untersuchungen hatten zum Ziel herauszufinden, welchen Einfluss jeweils radiografische Veränderungen und die Krankheitsaktivität auf den Funktionsstatus haben. Wie bereits im Rahmen der Würdigung der Argumente zur Krankheitsaktivität beschrieben, ist dieses Wissen für das Verständnis der RA und die Medikamentenentwicklung relevant. Es geht dabei unter anderem darum, den Einfluss beider Prozesse auf den körperlichen Funktionsstatus zu untersuchen und die Zeitpunkte des Krankheitsverlaufs, zu denen sie jeweils eine dominierende Rolle übernehmen [22].

Hinsichtlich des Arguments, es gebe Definitionen zu klinisch relevanten radiologischen Gelenkveränderungen, wurde auf Odegard et al. [24] verwiesen. Die Autoren folgern aus den Ergebnissen eines statistischen Modells anhand von Daten aus einer Kohortenstudie mit 238 Patienten, dass eine Veränderung von 10 Einheiten auf dem Sharp Score zu einer Veränderung von 0,03 Einheiten des HAQ führt. Allerdings gilt für den HAQ eine Veränderung von $\geq 0,22$ Einheiten als patientenrelevant [32]. Bruynesten et al. 2005 [25] treffen keine Aussagen über klinisch relevante Veränderungen, sondern vielmehr darüber, welche kleinste Veränderung der radiologischen Veränderung gemessen werden kann, die nicht einem Messfehler entspricht. Die Arbeit von Combe et al. 2007 [27] wurde in Stellungnahmen herangezogen, um zu belegen, dass eine klinisch relevante Veränderung des Sharp Score bei 4 Einheiten liege: hierbei beziehen sie sich vermutlich auf intraindividuelle Veränderungen und nicht auf einen Gruppenunterschied. Eine Schwierigkeit ergibt sich hinsichtlich der Aussage von Odegard et al., dass eine Veränderung 10 Einheiten auf dem Sharp Score zu einer Veränderung von 0,03 Einheiten führe, die aber nicht patientenrelevant ist (vergleiche oben). Zudem treffen Combe et al. eine zur Food and Drug Administration

(FDA) divergierende Aussage. Während Combe et al. beschreiben, dass die Untersuchung der radiologischen Progression gegebenenfalls um die Untersuchung der körperlichen Funktion ergänzt werden könne, formuliert die FDA die Anforderung gerade umgekehrt [33].

Insgesamt gilt für die radiografische Erfassung von Gelenkveränderungen, dass unklar ist, welche Veränderungen der radiografischen Progression patientenrelevant sind.

Eine Stellungnahme forderte darüber hinaus, die Instabilität als wichtiges Symptom im Berichtsplan aufzunehmen.

Das Symptom der Instabilität sieht das IQWiG durch bereits im Berichtsplan definierte patientenrelevante Endpunkte abgedeckt. Instabilität hat einen Einfluss auf den körperlichen Funktionsstatus und die Aktivitäten des täglichen Lebens.

Zusammenfassend ergab sich keine Notwendigkeit zur inhaltlichen Änderung oder Ergänzung des Berichtsplans. Für den im Berichtsplan definierten Endpunkt „strukturelle Gelenkveränderungen“ wurde die Aufzählung in der nachfolgenden Klammer durch das Wort „wie“ ergänzt, um das Missverständnis zu vermeiden, hier seien ausschließlich die 3 erwähnten Ausprägungen gemeint. So heißt es dort nun: „(wie Deformitäten, Versteifungen, Kontrakturen)“.

2.3.2.5 Soziales Funktionsniveau

Eine Stellungnahme befand die Beschreibung „Teilhabe am beruflichen und sozialen Leben“ für zu unscharf und merkte an, es seien hier auch „Arbeitsunfähigkeitstage bzw. Krankenhausaufenthalte, Minderung der Erwerbsfähigkeit, Pflegebedürftigkeit“ zu berücksichtigen.

Die als unscharf wahrgenommene Formulierung ist der Tatsache geschuldet, dass zum Zeitpunkt der Planung einer systematischen Übersichtsarbeit die Datenstruktur der verfügbaren Daten unbekannt ist (siehe hierzu Abschnitt 2.6.2.1).

Die in der Stellungnahme genannten Operationalisierungen wie Arbeitsunfähigkeitstage, Minderung der Erwerbsfähigkeit und Pflegebedürftigkeit werden in der Bewertung berücksichtigt, sofern Daten hierfür vorliegen.

Es ist darüber hinaus auch vorstellbar, dass Untersuchungen über Patientenbefragungen durchgeführt wurden, wie über die Social Adjustment Scale – Self Report (SAS-SR), einem generischen Fragebogen (vergleiche www.proquolid.org). Krankenhausaufenthalte selbst sind nicht zweifelsfrei unter das soziale Funktionsniveau zu fassen. Sie können ein Grund dafür sein, dass Betroffene sich eingeschränkt fühlen, was allerdings entsprechend zu untersuchen ist. Sie können aber auch eine Folge von z. B. Komplikationen der Erkrankung, einer Symptomverschlechterung oder einer unerwünschten Arzneimittelwirkung (UAW) sein und wären dann als solche zu bewerten.

Hieraus ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.3.2.6 Gesundheitsbezogene Lebensqualität

Gemäß einer Stellungnahme solle unter dem Endpunkt gesundheitsbezogene Lebensqualität neben dem Wohlbefinden auch die Depression bzw. sollten depressive Verstimmungen aufgeführt werden.

Die gesundheitsbezogene Lebensqualität ist ein komplexes Konstrukt, das per definitionem psychische, physische und soziale Teilaspekte umfasst [34]. Dabei bilden sich bei Verwendung geeigneter Fragebögen auch die Depression oder depressive Verstimmungen ab. Eine gesonderte Nennung dieser Teilaspekte erscheint deshalb nicht notwendig.

Hieraus ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.3.2.7 Unerwünschte Arzneimittelwirkungen

2 Stellungnahmen bezogen sich auf unerwünschte Ereignisse bzw. eine „Schadensanalyse“. Eine der Stellungnahmen merkte an, dass die wichtigsten unerwünschten Ereignisse explizit definiert werden sollten.

Die andere Stellungnahme führte eine ähnliche, aber weitergefasste Kritik an. Sie beschrieb, dass der Berichtsplan ausführlich auf die Nutzenbewertung eingehe, aber nicht auf die Bewertung des Schadens. So liege es im besonderen Interesse der Patienten, den Schaden vorab zu definieren und „das Schadenspotenzial angebracht zu analysieren“. Dabei wurde unter Nennung von Singh et al. 2009 [35] auf Studienabbrüche aufgrund von Nebenwirkungen verwiesen.

Das IQWiG stimmt darin überein, dass Erkenntnisse zum Schaden im besonderen Interesse der Patienten liegen. Das IQWiG stimmt nicht darin überein, dass eine vorab getroffene Einschränkung auf bestimmte unerwünschte Ereignisse auch im besonderen Interesse der Patienten liegt. Daher sollen in die Bewertung der unerwünschten Arzneimittelwirkungen alle Ereignisse einbezogen werden und nicht nur die sogenannten wichtigsten – was darüber hinaus eine subjektive Auswahl zur Folge haben könnte. Das IQWiG schließt keine unerwünschten Arzneimittelwirkungen vorab aus. In der Regel berücksichtigt das IQWiG mindestens die Gesamtrate der unerwünschten Ereignisse, die Gesamtrate der schwerwiegenden unerwünschten Ereignisse und die Gesamtrate der Studienabbrüche wegen unerwünschter Ereignisse, sofern hierzu Daten in eingeschlossenen Studien vorliegen. In der systematischen Übersichtsarbeit von Singh et al. 2009, die in einer der Stellungnahmen zitiert wurde, definierten die Autoren den Schaden vorab als Studienabbrüche aufgrund von Nebenwirkungen. Die Autoren beschreiben aber auch, dass sie spezielle unerwünschte Ereignisse in ihrer Analyse betrachten wollten. Daher trafen sie eine entsprechende Auswahl.

Die Kritik, dass sich der Berichtsplan ausschließlich auf die Bewertung des Nutzens und nicht des Schadens beziehe, sieht das IQWiG als nicht begründet. Es ist richtig, dass im Berichtsplan der Begriff „Nutzenbewertung“ wiederholt verwendet wird, aber nicht der Begriff „Schadensbewertung“. Allerdings erläutert das IQWiG in seinen Allgemeinen Methoden (Version 3.0) [2], in welchem Sinne der Begriff „Nutzenbewertung“ verwendet wird. Damit wird der gesamte Prozess bezeichnet, der die „Evaluation medizinischer Interventionen hinsichtlich ihrer kausal begründeten positiven und negativen Effekte im Vergleich mit einer klar definierten anderen Therapie, einem Placebo (oder einer andersartigen Scheinbehandlung) oder keiner Behandlung“ umfasst. „Negative Effekte“ beschreiben hierbei den Schaden.

Zusammenfassend ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.3.2.8 Interventions- und erkrankungsbedingter Aufwand

Eine Stellungnahme fragte, was unter der Formulierung „interventions- und erkrankungsbedingter Aufwand“ zu verstehen sei, und bat um eine präzisere Beschreibung. Eine andere kritisierte, dass dieser Endpunkt ergänzend betrachtet werde und ein Nutzen oder Zusatznutzen der Behandlung allein basierend auf diesen Zielgrößen sich nicht ergeben könne.

Zur ersten Anmerkung ist zu sagen, dass hierunter der Aufwand zu verstehen ist, der für Patienten aufgrund von bestimmten Interventionen oder aufgrund der Erkrankung entsteht. Dies können die in der Stellungnahme genannten Punkte sein wie Zahl der Arztbesuche oder Zahl der Arzneimittelwendungen aber durchaus auch andere Aspekte, die für Patienten einen Aufwand bedeuten. Die Zahl der Klinikeinweisungen müsste sicherlich dahin gehend differenziert werden, wieso die Klinikeinweisung notwendig ist – aufgrund von unerwünschten Arzneimittelwirkungen oder beispielsweise, weil eine Intervention nur in einer Klinik appliziert werden kann.

Eine Präzisierung dieses Endpunktes wird nicht vorgenommen, weil – wie auch schon an anderen Stellen des vorliegenden Dokuments beschrieben – vorab keine Einschränkung vorgenommen werden soll.

Bezugnehmend auf den Kritikpunkt, dass dieser Endpunkt als zweitrangig betrachtet werde, ist anzumerken, dass zunächst ein (Zusatz-)Nutzen in patientenrelevanten Endpunkten gezeigt sein muss, die die Mortalität, Morbidität oder Lebensqualität abbilden, bzw. dass der Schaden den (Zusatz-)Nutzen nicht überwiegt. Erst dann können Vorteile in interventions- und erkrankungsbedingten Aufwänden in Betracht gezogen werden. Darüber hinaus sollte sich eine relevante Beeinträchtigung von Patienten durch den interventions- und erkrankungsbedingten Aufwand in der gesundheitsbezogenen Lebensqualität niederschlagen.

Zusammenfassend ergab sich hieraus keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.3.3 Studientypen

Einige Stellungnahmen kritisierten, dass ausschließlich RCTs in die Nutzenbewertung eingeschlossen werden sollen. Die Forderung nach dem Einschluss weiterer Studiendesigns bezog sich auf die Erhebung von Daten zur Sicherheit und zur Gesamtmortalität sowie von Langzeitdaten bezüglich der Wirksamkeit. So wurde die Berücksichtigung von Registerstudien, Beobachtungsstudien und Extensionsstudien von RCTs gefordert. Eine Stellungnahme schlug vor, zunächst gar keine Einschränkung bezüglich einzuschließender Studiendesigns vorzunehmen.

Zur Beurteilung von Schadensmerkmalen ist anzumerken, dass unerwünschte Ereignisse und damit Aspekte der Sicherheit und Verträglichkeit gemäß Good Clinical Practice in jeder Interventionsstudie und damit auch in jedem RCT erhoben werden müssen [36]. Es ist allerdings richtig, dass sehr seltene unerwünschte Ereignisse in RCTs erst detektiert werden, wenn ausreichend große Patientenzahlen eingeschlossen werden. Bezüglich der Bewertung von Schäden ist aber festzuhalten, dass eine Signaldetektion hinsichtlich seltener unerwünschter Ereignisse nicht das Ziel der Bearbeitung des Auftrags des Gemeinsamen Bundesausschusses (G-BA) ist.

Das Institut stimmt mit den Stellungnehmenden überein, dass häufig zu wenige geeignete Langzeitstudien vorliegen, um den langfristigen Nutzen von Arzneimitteln bei progressiven Erkrankungen zu beschreiben.

Trotzdem erscheint der vorgeschlagene Einschluss ergebnisunsicherer Studien keine geeignete Lösung für diese Probleme. Der G-BA benötigt für seine Entscheidungsfindung Schlussfolgerungen aus ergebnissicheren Studien. Liegen solche Studien nicht vor, ist es aus Sicht des Instituts sinnvoller, die Wissenslücken zu beschreiben, als auf ergebnisunsichere Studien zurückzugreifen.

Zunächst gar keine Einschränkung in Bezug auf einzuschließende Designs vorzunehmen und die Auswahl der Studien dann basierend auf den vorliegenden Informationen zu treffen, löst ebenfalls das Problem ergebnisunsicherer Studien nicht und wäre darüber hinaus problematisch, da der spätere Studieneinschluss in Kenntnis der Datenlage geschehen würde.

Wie in den Allgemeinen Methoden 3.0 des IQWiG [2] beschrieben, können im Rahmen der Stellungnahmeverfahren nicht randomisierte Studien eingereicht werden, wenn sie nach Meinung der Stellungnehmenden für die Bewertung relevant sind. Es ist allerdings in solchen Fällen „eine adäquate Begründung für die Validität der kausalen Interpretation der in solchen Studien beschriebenen Effekte erforderlich“ [2]. Von den Stellungnehmenden wurden verschiedene Registerstudien vorgelegt, allerdings ohne eine ausreichende Begründung der

Validität der in diesen Studien beschriebenen Effekte. In einer von Sander im Jahr 2010 veröffentlichten Arbeit [38] wurden Publikationen zu Mortalitätsdaten aus verschiedenen Registern – insbesondere mit Fokus auf die Datengewinnung und -auswertung – gesichtet. Die Publikationen berichteten Daten aus dem britischen Register (British Society for Rheumatology Biologics Register [BSRBR]), dem schwedischen (South Swedish Arthritis Treatment Group [SSATG]) und dem spanischen Register (Spanish Registry of Adverse Events of Biological Therapies in Rheumatic Diseases [BIOBADASER]). Als problematisch wurden insbesondere die unterschiedlichen Vorgehensweisen bei der Rekrutierung von Kontrollgruppen beschrieben. So wurden entweder historische Kontrollgruppen herangezogen (z. B. das schwedische Register) oder bei der Rekrutierung von Kontrollgruppen die Daten mit Unterstützung von Sterberegistern gesammelt (z. B. dem britischen Register). Auch die unterschiedliche Berücksichtigung von Störgrößen führe dazu, dass Daten zwischen den Registern nicht vergleichbar seien. Eine Schlussfolgerung von Sander 2010 ist, dass die Publikationen nicht die Wirklichkeit widerspiegeln, sondern das „vom Leser interpretierte Abbild der Wahrnehmung der Autoren“ dieser Publikationen. Wie auch in der wissenschaftlichen Erörterung am 17.02.2011 dargestellt (siehe Wortprotokoll, Abschnitt 4.3.6), ist selbst für die Mortalität, ein gut zu messendes Ereignis, nicht festzustellen, wie groß der Nutzen oder Schaden ist. Explizit wurde dabei am Beispiel des schwedischen Registers hinsichtlich der Rekrutierung der Kontrollgruppe dargestellt, welche Schwierigkeiten bei der Interpretation der Daten zur Mortalität bestehen.

Ein weiteres Argument in einigen Stellungnahmen bezog sich darauf, dass die in RCTs eingeschlossenen Populationen sich von der Population, die in der Praxis mit bDMARDs behandelt werde, unterscheiden.

Wie im Berichtsplan [39] beschrieben, sind RCTs, sofern sie „methodisch adäquat und der jeweiligen Fragestellung angemessen durchgeführt wurden, mit der geringsten Ergebnisunsicherheit behaftet“. Es ist darauf hinzuweisen, dass die in einem RCT untersuchte Population viel weniger durch die Methode der Gruppenallokation der Patienten bedingt als z. B. durch die Ein- und Ausschlusskriterien der Studien [2].

Zusammenfassend ergab sich hieraus keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.3.4 Studiendauer

Mehrere Stellungnahmen enthielten die Forderung, die im Berichtsplan definierte Mindeststudiendauer von 6 Monaten (24 Wochen) auf 3 Monate (12 Wochen) anzupassen. Eine Stellungnahme bewertet die Studiendauer von 6 Monaten dagegen explizit als positiv.

Sowohl die EMA [37] als auch die FDA [33] fordern in ihren Guidelines für die Untersuchung der „efficacy“ und die im Berichtsplan definierten patientenrelevanten Endpunkten ausdrücklich eine Studiendauer von mindestens 6 Monate.

Als Begründung für eine Studiendauer von 3 Monaten wurden in den Stellungnahmen andere systematische Übersichten zitiert und wurde beschrieben, dass sich das IQWiG nach Anpassung der Studiendauer im Einklang mit dem Großteil internationaler Health Technology Assessments (HTAs) und der Cochrane Reviews befinde. Außerdem besagten Therapieempfehlungen, dass die erste Überprüfung des Therapieerfolgs nach den ersten 3 bis 6 Monaten durchzuführen sei. Hierzu wurden dem IQWiG eine Reihe von Zitaten beigelegt [33,37,40-47]. Auch der Gemeinsame Bundesausschuss (G-BA) erkenne die Beurteilung des Therapieerfolgs nach 3 Monaten an [48].

Die implizite Aussage, das IQWiG befinde sich mit der Definition einer Studiendauer von 6 Monaten nicht im Einklang mit internationalen systematischen Übersichten, kann nicht nachvollzogen werden, da mindestens 2 Cochrane Reviews [49,50] ebenfalls eine Mindeststudiendauer von 6 Monaten definiert haben. Außerdem steht das Vorgehen des IQWiG in Übereinstimmung mit den Empfehlungen der EMA und der FDA.

Aus der in den Stellungnahmen zitierten Literatur lässt sich ablesen, dass die Untersuchung nach 3 Monaten der Kontrolle eines ersten Therapieerfolgs dient und damit der Therapiesteuerung. So geht es in den Arbeiten [40-42] um die Frage nach der Verlaufskontrolle und der Therapiesteuerung. Eine Stellungnahme spricht selbst von einer „engmaschigen Kontrolle“ und zitiert dabei 3 Arbeiten, deren Ziel es war, die Ausgestaltung der Therapiesteuerung zu untersuchen [45-47]. Auch in dem vom G-BA genannten Therapiehinweis zu Adalimumab handelt es sich um Therapiesteuerung – nämlich darum Adalimumab abzusetzen, falls nach 3 Monaten kein deutliches klinisches Ansprechen vorliegt. Die Leitlinie der Deutschen Gesellschaft für Rheumatologie (DGRh, Schneider et al. [40]) stützt das Vorgehen des IQWiG, da sie für die Erhebung der Remission eine Studiendauer von 6 Monaten empfiehlt.

Im Rahmen der Stellungnahmen wurde beschrieben, dass die Vorgaben der Zulassungsbehörden ein sogenanntes „Early-Escape“ zwischen 3 und 6 Monaten für Personen zuließen, die in die Placebogruppe randomisiert wurden. Die Einhaltung dieser Vorgaben dürfe nicht zum Ausschluss von Studien aus der IQWiG-Bewertung führen.

Es ist nicht zu befürchten, dass Studien ausgeschlossen werden, weil für sie eine sogenannte „Early-Escape“-Option für Patienten der Kontrollgruppe zwischen 3 und 6 Monaten geplant wurde. Dies ist ein Aspekt, der in der Datenanalyse und -synthese berücksichtigt und adäquat adressiert werden muss. Die EMA [37] geht auch darauf ein, welche Studiendesigns es ermöglichen können, trotzdem eine ausreichend lange Studiendauer zu erzielen (siehe auch Abschnitt 2.3.3).

Eine Stellungnahme wies darauf hin, dass Etanercept bereits zugelassen worden sei, bevor die EMA-Richtlinien im Jahr 2003 veröffentlicht wurden. Es könne nicht zum Ausschluss von Studien führen, wenn in ihnen diese Richtlinien noch nicht berücksichtigt werden konnten. Außerdem würde der Therapieerfolg nicht über-, sondern eher unterschätzt werden.

Zunächst ist anzumerken, dass die Nutzenbewertung auf Basis der aktuellen Kenntnisse der Wissenschaft erfolgen wird. Das Argument, der Therapieerfolg werde nach 3 Monaten eher unterschätzt, erscheint nicht ausreichend. Aus Sicht des Instituts ist bei Patienten mit einer chronischen Erkrankung der tatsächlich erreichte langfristige Therapieerfolg von Bedeutung.

Eine Arbeitsgruppe, die sich Group for the Respect of Ethics and Excellence in Science (GREES) nennt, hat für die angekündigte Aktualisierung [51] der EMA-Richtlinien [37] eine konsentiertere Empfehlung im Rahmen der Zulassung für die Studiendauer bezüglich der Endpunkte Symptomkontrolle („signs and symptoms“), strukturelle Schäden sowie körperliche Funktion abgegeben [52]. Hinsichtlich der Studiendauer empfiehlt sie für Phase-3-Studien 3 Monate für Symptomkontrolle, um eine Überlegenheit der Prüflintervention gegenüber Placebo zu zeigen. Idealerweise würde auch die Überlegenheit für strukturelle Schäden oder die körperliche Funktion gezeigt werden. Die Arbeit enthält aber auch die Beschreibung, dass die Erhebung der strukturellen Schäden nach 3 Monaten verhindern soll, dass Patienten einer inadäquaten Therapie ausgesetzt sind. Hier handelt es sich wiederum um Therapiesteuerung. Zur körperlichen Funktion wird beschrieben, dass Veränderungen zum Zeitpunkt von 3 Monaten innerhalb des ersten Jahres erhalten oder sogar verbessert werden müssen.

Abschließend lässt sich auch aus diesem Papier nicht ableiten, dass die Erhebung der radiologischen Progression und der körperlichen Funktion nach 3 Monaten ausreicht. Das IQWiG hält für die Bewertung des Nutzens und damit der langfristigen Wirkung auch für die Symptomkontrolle eine Studiendauer von mindestens 6 Monaten für notwendig.

Zusammenfassend ergab sich hieraus keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.3.5 Vorgehen im Falle einer Zulassungsänderung im Projektverlauf

Die Autoren von 2 Stellungnahmen merkten an, dass ggf. entstehende Anpassungen der Kriterien für den Studieneinschluss im Falle eine Zulassungsänderung im Projektverlauf der zu bewertenden Interventionen, im Vorbericht bzw. im Abschlussbericht explizit vermerkt werden sollen.

Eine dieser Stellungnahmen bemerkte, dass dies als eine wesentliche methodische Änderung einzustufen sei und dass daher mit Verweis auf die Methoden 3.0 des IQWiG in Form eines Amendments Gelegenheit zur Stellungnahme gegeben werden müsse. Eine andere Stellungnahme schlug vor, Änderungen, die mit dem Zulassungsstatus zusammenhängen, zeitnah auf der Webseite zur veröffentlichen. Eine Anhörung wurde hier nicht gefordert.

Von einer Änderung des Zulassungsstatus ist die Methodik der Projektbearbeitung nicht betroffen bzw. sie wird dadurch nicht geändert Die Methodik besagt, dass der aktuelle Zulassungsstatus Grundlage der Bearbeitung ist. Dies wird eingehalten, indem die Kriterien

an Zulassungsänderungen angepasst werden. Somit läge in einem solchen Fall keine wesentliche methodische Änderung vor. Das IQWiG hält es für ausreichend diese Information im Vor- und Abschlussbericht darzustellen. Eine Information über Änderungen des Zulassungsstatus auf den Webseiten des IQWiG wird nicht für nötig gehalten, da die Verbreitung dieser Information nicht zu den Aufgaben des IQWiG gehört. Die Transparenz der Bewertung ist durch die Beschreibung der Zulassungsänderung im Vor- und Abschlussbericht ausreichend gewährleistet.

Hieraus ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.3.6 Einschluss von Studien, die die vorgenannten Kriterien nicht vollständig erfüllen

Eine Stellungnahme bezog sich auf das im Berichtsplan beschriebene Vorgehen für den Fall, dass die Einschlusskriterien Population (E1) und Prüfintervention (E2) oder Vergleichsintervention (E3) nicht in Gänze erfüllt sind.

In der Stellungnahme wurde auf die missverständliche Beschreibung hingewiesen, dass für den Fall, dass weniger als 80 % der Patienten das Einschlusskriterium E1 erfüllen, die Studien dennoch eingeschlossen werden, wenn entsprechende Subgruppenanalysen vorliegen, aber dieser Zusatz für die Einschlusskriterien E2 und E3 fehle. Darüber hinaus wurde gefordert, Subgruppenanalysen auch für das Einschlusskriterium der patientenrelevanten Endpunkte (E4) zu berücksichtigen.

Es ist richtig, dass das Vorgehen hinsichtlich der Kriterien E2 und E3 missverständlich beschrieben ist. In der Praxis wird sich für E2 und E3 seltener zeigen, dass diese von weniger als 80 % der Personen erfüllt sind, als für E1. Daher wurde dieser Satz bezüglich der Subgruppenanalysen im vorläufigen Berichtsplan zunächst nicht explizit aufgenommen, obwohl für E2 und E3 dieselbe Vorgehensweise wie für E1 gilt.

Für das Einschlusskriterium E4 gilt, dass, wenn der Anteil der in die Auswertung einzuschließenden Personen weniger als 70 % beträgt, die Daten nicht in die Nutzenbewertung einbezogen werden (vergleiche hierzu auch Abschnitt 2.6.5 des vorliegenden Dokuments und Abschnitt 4.4.1 des Berichtsplans). Das generelle Vorgehen zu Subgruppenanalysen wird in Abschnitt 4.4.4 des Berichtsplans beschrieben.

Zusammenfassend ergab sich hieraus keine Notwendigkeit zur inhaltlichen Änderung des Berichtsplans. In Abschnitt 4.1.8 wurde ein Satz aufgenommen, aus dem deutlich hervorgeht, dass für die Einschlusskriterien E2 und E3 das gleiche Vorgehen wie für E1 gilt.

2.4 Informationsbeschaffung

Eine Stellungnahme wies auf Studien des Unternehmens hin und bat darum, diese Studien in der Nutzenbewertung zu berücksichtigen.

Das IQWiG wird alle Studien in der Nutzenbewertung berücksichtigen, deren Ergebnisse ausgewertet sind, die vom IQWiG identifiziert werden konnten und die die Einschlusskriterien des Berichtsplans erfüllen.

Hieraus ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.4.1 Bibliografische Recherche

3 Stellungnahmen kritisierten, dass im Berichtsplan keine Termini der Suchstrategie angegeben seien, wie es auch im Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement [53] empfohlen sei. Hierdurch gebe es keine Möglichkeit, vor Abschluss des Vorberichts Stellung zu nehmen.

Wie bisher in allen IQWiG-Berichten sind die Suchstrategie und das damit verbundene Ergebnis Gegenstand der Anhörung zum Vorbericht. Zu diesem Zeitpunkt kann das IQWiG sowohl auf Fehler in der Recherchestrategie als auch auf fehlende oder nicht identifizierte Studien aufmerksam gemacht werden. PRISMA bezieht sich auf das Berichten und die Darstellung der Ergebnisse eines systematischen Reviews. Diese liegen zum Zeitpunkt des Vorberichts vor.

Im Übrigen geht das IQWiG über die Forderung des PRISMA-Statements zur Darstellung der Suchstrategien hinaus. Dort wird beschrieben, dass die Suchstrategie für mindestens eine elektronische Datenbank dokumentiert werden sollte. Das IQWiG hingegen stellt in seinen Vor- und Abschlussberichten immer die Suchstrategien für alle elektronischen Datenbanken dar.

Hieraus ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.4.2 Selektion relevanter Studien

Eine Stellungnahme bemängelte die Beschreibung der Selektion relevanter Studien. Dabei wurde mit Verweis auf das Cochrane Handbook for Systematic Reviews of Interventions [54] gefordert, dass die Auswahl der 2 Reviewer zum Beispiel hinsichtlich des Erfahrungslevels beschrieben sein müsse. Es finde sich im Berichtsplan auch keine Beschreibung des Vorgehens, wenn eine Diskrepanz durch die beiden Reviewer nicht gelöst werden könne. Darüber hinaus wurde in der Stellungnahme auf das Abschnitt 6.1.5 der Allgemeinen Methoden 3.0 des IQWiG [2] verwiesen, in dem die Dokumentation der bibliografischen Literaturrecherche beschrieben ist. Abschließend wurde angemerkt, dass im Selektionsprozess Werturteile einfließen, die so gering wie möglich zu halten seien.

Es ist richtig, dass das Cochrane Handbook for Systematic Reviews of Interventions in Abschnitt 7.2.4 aufführt, dass die Angabe empfohlen ist, welche Entscheidungen von wem getroffen wurden: allerdings geht es hierbei nicht um das Erfahrungslevel generell, sondern um den Hintergrund – Methodiker, Arzt oder Experte in der Fragestellung. Das IQWiG sieht

in seinen Bewertungen diese Anforderungen als erfüllt und die entsprechende Transparenz als gewährleistet an. So werden in die Projektbearbeitung des vorliegenden Auftrags externe Sachverständige einbezogen, was explizit im Berichtsplan dargestellt ist. Diese werden im Abschlussbericht namentlich benannt, inklusive der Auflistung von potenziellen Interessenkonflikten. Darüber hinaus bietet das IQWiG zu mehreren Zeitpunkten Gelegenheit zur Stellungnahme, sodass weitere Expertise zur Klinik, Methodik, Studienauswahl oder Informationssynthese eingebracht werden kann.

Zum Argument, in den Selektionsprozess gingen Werturteile ein, ist darauf hinzuweisen, dass die Kriterien für den Studieneinschluss im Berichtsplan festgelegt sind und überprüfbar eingesetzt werden können. Der Einfluss ggf. auftretender unterschiedlicher, individueller Einschätzungen von Studienmerkmalen bezüglich der Einschlusskriterien wird dadurch reduziert, dass 2 Reviewer die Studieneinschlussentscheidung zunächst unabhängig voneinander durchführen und im Anschluss abweichende Entscheidungen konsentiert werden. Durch die klare Definition der Kriterien für den Studieneinschluss ist es in der Regel möglich, Diskrepanzen zu zweit zu lösen. Dieses Vorgehen entspricht den internationalen Standards der evidenzbasierten Medizin.

Zusammenfassend ergab sich hieraus keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.5 Informationsbewertung

Eine Stellungnahme bezog sich auf die Bewertung des Intention-to-Treat(ITT)-Prinzips, eine andere auf die Bewertung der ergebnisgesteuerten Berichterstattung.

Das die Bewertung des ITT-Prinzips adressierende Argument bezog sich darauf, dass es unterschiedliche Imputationsmethoden gebe, die ebenfalls zu extrahieren und zu bewerten seien. Sie könnten zu unterschiedlichen Ergebnissen führen.

Im Rahmen der ergebnisgesteuerten Berichterstattung sei der Zeitpunkt der Endpunkterhebung hinsichtlich der „klinischen Krankheitsaktivität (z. B. Schmerzen und Morgensteifigkeit)“, insbesondere bei längeren Dosierungsintervallen, zu berücksichtigen. So könne je nach Zeitpunkt der Endpunkterhebung beeinflusst werden, ob ein Ergebnis positiver oder negativer ausfalle. Es wurde daher beantragt, bei der Auswahl von RCTs zu Arzneimitteln mit langen Dosierungsintervallen hierauf zu achten.

Hinsichtlich des ITT-Prinzips ist zu sagen, dass grundsätzlich bei seiner Bewertung auch die Imputationsmethode berücksichtigt wird. Dosierungsintervalle sind bei der Informationssynthese, -analyse und -bewertung selbstverständlich zu berücksichtigen.

Zusammenfassend ergab sich hieraus keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.6 Informationssynthese und -analyse

2.6.1 Patientenrelevante Endpunkte

Eine Stellungnahme stellte fest, dass unklar bleibe, wie die patientenrelevanten Endpunkte gewichtet würden.

Es findet keine gewichtete Aggregation der patientenrelevanten Endpunkte statt, weshalb sich hierzu auch keine Beschreibung im Berichtsplan findet. Die Ergebnisse werden für jeden einzelnen Endpunkt betrachtet und interpretiert.

Hieraus ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.6.2 Zusammenfassen von Ergebnissen

Eine Stellungnahme verwies darauf, dass Ergebnisse verschiedener Scores zur Erhebung der Krankheitsaktivität nicht zusammengefasst werden sollten. In den Studien könnten verschiedene Scores eingesetzt worden sein: dies sei zu berücksichtigen.

Die Art der Zusammenfassung von Ergebnissen, also auch Scores, ist Gegenstand der Anhörung des Vorberichts. Wie bei allen Endpunkten ist vor der Durchführung einer meta-analytischen Zusammenfassung zu überprüfen, was sinnvoll ist.

Hieraus ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.6.3 Relevanzschwelle

6 Stellungnahmen äußerten sich zum Vorgehen zur Einschätzung der Relevanz von Effekten, die mithilfe von (komplexen) Skalen gemessen wurden. Zur Klarstellung wird das geplante hierarchische Vorgehen hier noch einmal kurz dargestellt:

Es ist grundsätzlich geplant, zur Beurteilung der Relevanz vorrangig validierte bzw. etablierte Irrelevanzgrenzen einzusetzen. Sofern diese für das jeweilige Messinstrument und den jeweiligen Endpunkt nicht existieren und auch keine validierten bzw. etablierten Relevanzgrenzen bekannt sind, aus denen Irrelevanzgrenzen abgeleitet werden könnten, werden Responderanalysen verwendet. Sind auch keine Responderanalysen verfügbar, können ersatzweise andere geeignete Informationen zu Relevanzgrenzen herangezogen werden, um Irrelevanzgrenzen zu approximieren. Das Ableiten von Irrelevanzgrenzen aus z. B. akzeptierten Relevanzgrenzen kann als entsprechende Fortsetzung des von Kieser und Hauschke [55] vorgeschlagenen Ansatzes gesehen werden. Damit steht die in einer Stellungnahme vorgeschlagene Bewertungsalternative nicht im Gegensatz zu dem vom IQWiG geplanten Vorgehen.

Nur in dem Fall, dass keine entsprechenden Angaben zur Beurteilung der Relevanz vorliegen, wird eine Irrelevanzschwelle der standardisierten Mittelwertdifferenz (SMD) von 0,2 angenommen.

Die in 2 Stellungnahmen vorgebrachte Forderung, dass die Relevanzbeurteilung in der vorliegenden Bewertung durch eine übergreifende Konsensbildung ggf. unter Hinzuziehung externer Experten erfolgen soll, ist aus Sicht des Instituts nicht primär geeignet, adäquate Relevanzschwellen zu ermitteln. Primär sollte die Relevanzbeurteilung auf mit geeigneten Verfahren empirisch erhobenen Daten beruhen [56].

Aus den Stellungnahmen ergab sich damit keine Notwendigkeit für eine Änderung des Berichtsplans.

2.6.4 Indirekte Vergleiche

Eine Stellungnahme kritisierte, dass sich der Berichtsplan nicht zu indirekten Vergleichen äußere – insbesondere, da das IQWiG in seinem Papier „Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten“ (Version 1.0) [57] beschreibe, dass die Betrachtung von indirekten Vergleichen unter gewissen Voraussetzungen möglich sei.

Wie in der Stellungnahme richtig vermerkt wurde, stammt diese Aussage aus den Allgemeinen Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten und nicht aus den Allgemeinen Methoden (Version 3.0) [2], die für die reine Nutzenbewertung gelten. Im vorliegenden Auftrag handelt es sich um eine solche reine Nutzenbewertung. In den Allgemeinen Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten findet sich außerdem folgendes Zitat auf der Seite 16, auf das sich auch die Stellungnahme bezieht: „Um eine Kosten-Nutzen-Bewertung multipler Interventionen zu ermöglichen, kann das Institut unter Inkaufnahme einer – im Vergleich zum Ansatz der reinen Nutzenbewertung – geringeren Ergebnissicherheit auch indirekte Vergleiche zur Bewertung von Kosten-Nutzen-Verhältnissen heranziehen.“ Aus diesem Satz geht deutlich hervor, dass das IQWiG für die reine Nutzenbewertung in der Regel wegen der geringen Ergebnissicherheit keine indirekten Vergleiche durchführt. Im Bereich der rheumatoiden Arthritis ist zu vermuten, dass aufgrund der in den Stellungnahmen zahlreich beschriebenen Confounder, Ergebnisse aus indirekten Vergleichen zusätzliche Unsicherheiten mit sich bringen.

Hieraus ergab sich keine Notwendigkeit zur Änderung oder Ergänzung des Berichtsplans.

2.6.5 Gegenüberstellung der Ergebnisse der Einzelstudien

3 Stellungnahmen bezogen sich – in zum Teil gegensätzlicher Weise – auf das geplante Vorgehen, dass in bestimmten Fällen einzelne Ergebnisse aus den Studien zu einem Endpunkt nicht in die Nutzenbewertung einbezogen werden. Dies ist in der Regel dann der Fall, wenn

diese auf weniger als 70 % der in die Auswertung einzubeziehenden Patienten basieren (= Nichtberücksichtigungsanteil liegt bei über 30 %) oder wenn der Unterschied der Nichtberücksichtigungsanteile zwischen den Gruppen mehr als 15 Prozentpunkte beträgt.

Während 2 Stellungnahmen dieses Vorgehen grundsätzlich als sinnvoll erachteten, forderte eine andere Stellungnahme, diesen Abschnitt aus dem Berichtsplan zu entfernen.

Eine dieser Stellungnahmen schlug eine neue Formulierung vor, denn es wurde davon ausgegangen, dass es sich um den Umgang mit „Drop-out-Raten“ innerhalb einer Studie handelt. Im Langzeitverlauf einer Studie steige die Drop-out-Rate naturgemäß an. Aus diesem Grund müssten Messzeitpunkte definiert werden, sodass Ergebnisse von Langzeitstudien nicht ausgeschlossen werden würden.

Hier handelt es sich um ein Missverständnis. Es handelt sich nicht um Drop-out-Raten, sondern um den Nichtberücksichtigungsanteil, der – wie ebenfalls im Berichtsplan beschrieben – den Anteil der Patienten beschreibt, der ohne jegliche Berücksichtigung in der Auswertung ist. Durch Studienabbrüche bedingt fehlenden Werten kann mit Ersetzungsverfahren wie beispielsweise Last Observation Carried Forward (LOCF) und einer adäquaten ITT-Analyse begegnet werden. Somit kann auch auf progressive Krankheitsverläufe und dadurch bedingte Abbrüche reagiert werden sowie auf die Situation in Langzeitstudien. Vielmehr betrifft das Kriterium Situationen, in denen mehr als 30 % der eingeschlossenen (in der Regel randomisierten) Patienten keinerlei Berücksichtigung in der Analyse eines Endpunkts finden. Des Weiteren muss an dieser Stelle klargestellt werden, dass es sich hierbei nicht um ein Kriterium der Studienauswahl handelt. Lediglich der betreffende Endpunkt wird dann nicht betrachtet.

Die andere der Stellungnahmen, die dieses Vorgehen als „schlüssig“ einschätzt, bat um eine Referenz, die für den Gruppenunterschied von 15 % angeführt werden könne, ohne die gewählte Größe des Gruppenunterschieds selbst zu kritisieren.

Aus der Literatur sind derzeit keine expliziten und allgemein anerkannten Grenzen für maximal akzeptable Unterschiede in den Anteilen nicht berücksichtigter Patienten bekannt. Ein Unterschied in Nichtberücksichtigungsanteilen ist ein Indiz für ein Selektionsbias, der dadurch entsteht, dass die durch die Randomisierung erzielte Vergleichbarkeit der Behandlungsgruppen eingeschränkt wird (siehe hierzu auch die Würdigung des nachfolgenden Arguments).

Die dritte Stellungnahme befand die Größe der Nichtberücksichtigungsanteile für willkürlich gewählt und forderte, den Abschnitt zu streichen und stattdessen diesen Aspekt im Rahmen der Einschätzung des Verzerrungspotenzials einzubeziehen.

Der Grund, diese Ergebnisse nicht zu berücksichtigen und nicht lediglich im Verzerrungspotenzial zu berücksichtigen, ist, dass die Daten nicht nur potenziell hochverzerrt sind,

sondern aus Sicht des IQWiG grundsätzlich keine Schlussfolgerungen zulassen. Wie im Berichtsplan bereits beschrieben, werden in der internationalen Literatur bereits Nichtberücksichtigungsanteile von 20 Prozentpunkten als nicht mehr aussagekräftig betrachtet [58]. Daher kann das Vorgehen des IQWiG als vergleichsweise liberal betrachtet werden. Ein Gruppenunterschied von mehr als 15 Prozentpunkten stellt die Strukturgleichheit, die einst durch die Randomisierung hergestellt wurde, infrage.

2.6.6 Meta-Analysen

In einer Stellungnahme wurde das im Berichtsplan beschriebene Vorgehen bei der Durchführung von Meta-Analysen kritisiert. So seien Modelle mit zufälligen Effekten (REMs) nicht generell als Modell der Wahl einzusetzen. Bei einer kleinen Anzahl von Studien seien diese nicht geeignet und falls die Heterogenität nicht zu groß sei, sollten Modelle mit festen Effekten (FEMs) eingesetzt werden.

Nach Einschätzung des Instituts gibt es zu der Frage nach dem grundsätzlich geeigneten Modell, FEM oder REM, zur Berechnung eines gepoolten Schätzers im Rahmen einer Meta-Analyse keine allgemeingültige, eindeutige Empfehlung [59-61]. Das Institut verwendet zur Durchführung von Meta-Analysen primär das REM, da dieses Modell – im Gegensatz zum FEM – die Berücksichtigung von möglicherweise heterogenen Studienergebnissen ermöglicht. Das REM ist ein allgemeineres Modell, welches das FEM als Spezialfall enthält. Mit sinkender Heterogenität der Einzelstudienresultate konvergiert das REM gegen das FEM und bei Homogenität der Studien ($I^2 = 0$) fallen beide Modelle zusammen. Bei homogener Studienlage liefern beide Modelle also (fast) identische Ergebnisse, sodass in der Praxis keine abweichenden Resultate zu erwarten sind. Bei heterogener Studienlage sind die Voraussetzungen des FEM dagegen nicht erfüllt, daher ist in solchen Fällen das REM das geeignetere Modell. Laut Borenstein et al. 2009 ist im Falle der Zusammenfassung von publizierten Studienergebnissen das REM im Allgemeinen der plausiblere Ansatz [59]. Auch Whitehead 2002 hat (bei einer größeren Anzahl von Studien) eine Präferenz für das REM [61]. Die Simulationsuntersuchung von Brockwell und Gordon 2001 hat gezeigt, dass schon bei geringer Heterogenität (bei einer geringen Anzahl von Studien) das FEM das vorgegebene Niveau (α -Fehler) weniger gut einhält als das REM [62].

In speziellen Situationen kann ein REM zu unsicheren Ergebnissen führen, z. B. beim Vorliegen weniger Studien mit sehr unterschiedlichen Fallzahlen. Eine Verwendung des FEM in derartigen Situationen ist unter Umständen durchaus sinnvoll und im Berichtsplan auch vorgesehen („In begründeten Ausnahmefällen werden Modelle mit festen Effekten eingesetzt“). Im Berichtsplan lassen sich nur Methodenkriterien für vorhersehbare Konstellationen nennen; eine erschöpfende Aufzählung aller Eventualitäten, die eine Spezifizierung der Methodik erfordern, ist nicht möglich. Sollten in Einzelfällen FEMs zum Einsatz kommen, wird dies im Vorbericht dargelegt und steht daraufhin in der Anhörung zum

Vorbericht zur Diskussion, sodass eine Stellungnahmemöglichkeit auch in einem solchen Fall gewährleistet ist.

In einer Stellungnahme wurde darauf hingewiesen, dass neben frequentistischen Meta-Analysen zunehmend bayesianische Analysen verwendet würden. Zur Argumentation wurde auf eine bislang unveröffentlichte Studie verwiesen, in der die Unterschiede von frequentistischen und bayesianischen Ansätzen dargestellt würden. Außerdem wurde darauf hingewiesen, dass Bayes-Ergebnisse für Kliniker leichter zu interpretieren seien.

Hierzu ist anzumerken, dass bei bayesianischen Verfahren grundsätzlich subjektive Wahrscheinlichkeitsbewertungen über sogenannte A-priori-Verteilungen in die Berechnung eingehen und diese daher vor dem Hintergrund einer objektiven, datengesteuerten Auswertung kritisch zu sehen sind. Im speziellen Fall, in dem gleichverteilte A-priori-Verteilungen (nichtinformativ A-priori-Verteilungen) verwendet werden, sind approximativ gleiche Ergebnisse nach einem bayesianischen und frequentistischen Vorgehen zu erwarten [63]. In Bezug auf die Interpretation von Ergebnissen durch Kliniker können klassische Verfahren von Vorteil sein, da ihr Einsatz regelhaft und etabliert ist.

Eine andere Stellungnahme bezog sich auf das im Berichtsplan beschriebene Vorgehen, Schätzer für die Lage und Streuung zur Durchführung von Meta-Analysen selbst zu berechnen, falls diese nicht vorlägen. Hier wurde eine Spezifizierung des Vorgehens gefordert, mit welcher Methodik diese Berechnungen durchgeführt werden sollen, inklusive der Nennung von Beispielen.

Nach den Methoden des IQWiG (Version 3.0) werden Daten, die für die Durchführung einer Meta-Analyse notwendig sind und die in den Studienpublikationen nicht vorliegen, nach Möglichkeit aus anderen vorhandenen Angaben berechnet oder näherungsweise bestimmt. Welche spezielle Ersetzungsmethodik gewählt wird, hängt dabei von der jeweiligen Datensituation ab. Das Institut orientiert sich am grundlegenden Vorgehen, wie es in [54] dargestellt ist; als Beispiele für spezielle Verfahren können [64-66] genannt werden.

2.6.7 Subgruppenmerkmale und andere Effektmodifikatoren

Mehrere Stellungnahmen merkten an, dass neben den im Berichtsplan definierten potenziellen Effektmodifikatoren (Geschlecht, Alter, Art der Vorbehandlung) weitere zu betrachten seien. Es wurden insgesamt 9 Faktoren in den Stellungnahmen genannt.

Es ist generell zu beachten, dass die Anzahl der Analysen umso größer ist, je höher die Anzahl der geplanten potenziellen Effektmodifikatoren ist. Dies wiederum brächte die Probleme des multiplen Testens mit sich [67]. Die Zahl der geplanten potenziellen Effektmodifikatoren soll daher so gering wie möglich und so groß wie nötig sein. Aus diesem Grund müssen gute Gründe vorliegen, zusätzliche potenzielle Effektmodifikatoren in den Berichtsplan aufzunehmen.

Aus den Vorschlägen in den Stellungnahmen ging kein Effektmodifikator hervor, der vorab zwingend definiert werden müsste. Insbesondere bei den Effektmodifikatoren der Krankheitsdauer und Krankheitsschwere stellte sich die Frage, wie diese operationalisiert werden könnten. Die Diskussion in der wissenschaftlichen Erörterung zeigte, dass dies unklar ist und dass Aspekte der Krankheitsdauer und Krankheitsschwere bereits in die als Effektmodifikator geplante Art der Vorbehandlung eingehen. In der Diskussion wurde klargestellt, dass bei der Betrachtung von heterogenen Ergebnissen die Basisdaten dahingehend überprüft werden müssten, ob die Studien unterschiedliche Populationen eingeschlossen hätten. Dies entspricht dem üblichen Vorgehen des IQWiG, bei der Untersuchung von Heterogenität die Basisdaten sowie die Ein- und Ausschlusskriterien von Studien als mögliche Ursache für die Heterogenität zu prüfen. Zudem enthält der Berichtsplan in Abschnitt 4.4.4 auch die Möglichkeit über die im Berichtsplan definierten potenziellen Effektmodifikatoren hinausgehende zu untersuchen, sofern sich aus den verfügbaren Informationen dafür Anhaltspunkte ergeben.

- Eine Stellungnahme merkte an, dass es bisher keine Anhaltspunkte dafür gebe, dass das Geschlecht ein Effektmodifikator sei.
- Das IQWiG ist gesetzlich dazu angehalten regelmäßig die potenziellen Effektmodifikatoren Alter und Geschlecht zu untersuchen (siehe Sozialgesetzbuch [SGB] V § 139a [3]).

2.7 Literaturverzeichnis

1. Wollenhaupt J, Alten R, Backhaus M, Baerwald C, Braun J, Burkhard H et al. Aktualisiertes Therapieschema der Rheumatoiden Arthritis: Ergebnisse eines Konsensprozesses deutscher Rheumatologen 2009. *Aktuelle Rheumatologie* 2009; 34(4): 234-239.
2. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden: Version 3.0 [online]. 25.06.2010 [Zugriff: 08.10.2008]. URL: http://www.iqwig.de/download/IQWiG_Methoden_Version_3_0.pdf.
3. 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-S108.
4. Combe B, Cantagrel A, Goupille P, Bozonnat MC, Sibilia J, Eliaou JF et al. Predictive factors of 5-year health assessment questionnaire disability in early rheumatoid arthritis. *J Rheumatol* 2003; 30(11): 2344-2349.
5. Dawes PT, Fowler PD, Clarke S, Fisher J, Lawton A, Shadforth MF. Rheumatoid arthritis: treatment which controls the C-reactive protein and erythrocyte sedimentation rate reduces radiological progression. *Br J Rheumatol* 1986; 25(1): 44-49.
6. Drossaers-Bakker KW, De Buck M, Van Zeben D, Zwinderman AH, Breedveld FC, Hazes JM. Long-term course and outcome of functional capacity in rheumatoid arthritis: the effect of disease activity and radiologic damage over time. *Arthritis Rheum* 1999; 42(9): 1854-1860.
7. Smolen JS, Han C, Van der Heijde DMFM, Emery P, Bathon JM, Keystone E et al. Radiographic changes in rheumatoid arthritis patients attaining different disease activity states with methotrexate monotherapy and infliximab plus methotrexate: the impacts of remission and tumour necrosis factor blockade. *Ann Rheum Dis* 2009; 68(6): 823-827.
8. Smolen JS, Van der Heijde DM, St Clair EW, Emery P, Bathon JM, Keystone E et al. Predictors of joint damage in patients with early rheumatoid arthritis treated with high-dose methotrexate with or without concomitant infliximab: results from the ASPIRE trial. *Arthritis Rheum* 2006; 54(3): 702-710.
9. Van der Heijde DM, Van Riel PL, Van Leeuwen MA, Van 't Hof MA, Van Rijswijk MH, Van de Putte LB. Prognostic factors for radiographic damage and physical disability in early rheumatoid arthritis: a prospective follow-up study of 147 patients. *Br J Rheumatol* 1992; 31(8): 519-525.
10. Van Leeuwen MA, Van der Heijde DM, Van Rijswijk MH, Houtman PM, Van Riel PL, Van de Putte LB et al. Interrelationship of outcome measures and process variables in early rheumatoid arthritis: a comparison of radiologic damage, physical disability, joint counts, and acute phase reactants. *J Rheumatol* 1994; 21(3): 425-429.

11. Welsing PM, Landewe RB, Van Riel PL, Boers M, Van Gestel AM, Van der Linden S et al. The relationship between disease activity and radiologic progression in patients with rheumatoid arthritis: a longitudinal analysis. *Arthritis Rheum* 2004; 50(7): 2082-2093.
12. American College of Rheumatology Committee to Reevaluate Improvement Criteria. A proposed revision to the ACR20: the hybrid measure of American College of Rheumatology response. *Arthritis Rheum* 2007; 57(2): 193-202.
13. Kirwan J, Minnock P, Adebajo A, Bresnihan B, Choy E, De Wit M et al. Patient perspective: fatigue as a recommended patient-centered outcome measure in rheumatoid arthritis. *J Rheumatol* 2007; 34(5): 1174-1177.
14. Riemsma RP, Rasker JJ, Taal E, Griep EN, Wouters JM, Wiegman O. Fatigue in rheumatoid arthritis: the role of self-efficacy and problematic social support. *Rheumatology* 1998; 37(10): 1042-1046.
15. Meads K, Barton P, Tsourapas A, Ashfaq K, Connock M, Fry-Smith A et al. Tocilizumab for the treatment of rheumatoid arthritis: a single technology appraisal [online]. 04.2009 [Zugriff: 09.12.2010]. URL: <http://www.nice.org.uk/nicemedia/live/12033/45575/45575.pdf>.
16. Lee KA. Lee Fatigue Scale (LFS or VAS-F) [online]. In: PROQOLID Patient-Reported Outcome and Quality of Life Instruments Database. 05.2008 [Zugriff: 29.06.2010]. URL: http://www.proqolid.org/proqolid/instruments/lee_fatigue_scale_lfs_or_vas_f.
17. FACIT.org. FACIT-F: version 4 [online]. 16.11.2007 [Zugriff: 02.12.2010]. URL: <http://www.facit.org/LiteratureRetrieve.aspx?ID=42363>.
18. Van der Heijde D, Landewe R, Van Vollenhoven R, Fatenejad S, Klareskog L. Level of radiographic damage and radiographic progression are determinants of physical function: a longitudinal analysis of the TEMPO trial. *Ann Rheum Dis* 2008; 67(9): 1267-1270.
19. Breedveld FC, Han C, Bala M, Van der Heijde D, Baker D, Kavanaugh AF et al. Association between baseline radiographic damage and improvement in physical function after treatment of patients with rheumatoid arthritis. *Ann Rheum Dis* 2005; 64(1): 52-55.
20. Kavanaugh A, Han C, Bala M. Functional status and radiographic joint damage are associated with health economic outcomes in patients with rheumatoid arthritis. *J Rheumatol* 2004; 31(5): 849-855.
21. Scott DL, Pugner K, Kaarela K, Doyle DV, Woolf A, Holmes J et al. The links between joint damage and disability in rheumatoid arthritis. *Rheumatology* 2000; 39(2): 122-132.
22. Welsing PM, Van Gestel AM, Swinkels HL, Kiemeny LA, Van Riel PL. The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. *Arthritis Rheum* 2001; 44(9): 2009-2017.
23. Aletaha D, Smolen J, Ward MM. Measuring function in rheumatoid arthritis: identifying reversible and irreversible components. *Arthritis Rheum* 2006; 54(9): 2784-2792.

24. Odegard S, Landewe R, Van der Heijde D, Kvien TK, Mowinckel P, Uhlig T. Association of early radiographic damage with impaired physical function in rheumatoid arthritis: a ten-year, longitudinal observational study in 238 patients. *Arthritis Rheum* 2006; 54(1): 68-75.
25. Bruynesteyn K, Boers M, Kostense P, Van der Linden S, Van der Heijde D. Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. *Ann Rheum Dis* 2005; 64(2): 179-182.
26. Bruynesteyn K, Van der Heijde D, Boers M, Saudan A, Peloso P, Paulus H et al. Determination of the minimal clinically important difference in rheumatoid arthritis joint damage of the Sharp/van der Heijde and Larsen/Scott scoring methods by clinical experts and comparison with the smallest detectable difference. *Arthritis Rheum* 2002; 46(4): 913-920.
27. Combe B, Landewe R, Lukas C, Bolosiu HD, Breedveld F, Dougados M et al. EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis* 2007; 66(1): 34-45.
28. Baker SG. Surrogate endpoints: wishful thinking or reality? *J Natl Cancer Inst* 2006; 98(8): 502-503.
29. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Stat Med* 2006; 25(2): 183-203.
30. Lassere MN. The Biomarker-Surrogacy Evaluation Schema: a review of the biomarker-surrogate literature and a proposal for a criterion-based, quantitative, multidimensional hierarchical levels of evidence schema for evaluating the status of biomarkers as surrogate endpoints. *Stat Methods Med Res* 2008; 17(3): 303-340.
31. Mangiapane S, Garrido MV. Surrogate als Parameter der Nutzenbewertung [online]. 2009 [Zugriff: 28.03.2011]. (DIMDI Schriftenreihe Health Technology Assessment; Band 91). URL: http://portal.dimdi.de/de/hta/hta_berichte/hta250_bericht_de.pdf.
32. Wells GA, Tugwell P, Kraag GR, Baker PRA, Groh J, Redelmeier DA. Minimum important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 1993; 20(3): 557-560.
33. Food and Drug Administration. Guidance for industry: clinical development programs for drugs, devices, and biological products for the treatment of rheumatoid arthritis (RA) [online]. 02.1999 [Zugriff: 23.12.2009]. URL: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071579.pdf>.
34. Koller M, Neugebauer EA, Augustin M, Bussing A, Farin E, Klinkhammer-Schalke M et al. Die Erfassung von Lebensqualität in der Versorgungsforschung: konzeptuelle, methodische und strukturelle Voraussetzungen. *Gesundheitswesen* 2009; 71(12): 864-872.
35. Singh JA, Christensen R, Wells GA, Suarez-Almazor ME, Buchbinder R, Lopez-Olivo MA et al. A network meta-analysis of randomized controlled trials of biologics for rheumatoid arthritis: a Cochrane overview. *CMAJ* 2009; 181(11): 787-796.

36. European Medicines Agency. ICH Topic E 6 (R1): guideline for good clinical practice; step 5 [online]. 07.2002 [Zugriff: 28.07.2010]. URL: <http://www.emea.europa.eu/pdfs/human/ich/013595en.pdf>.
37. European Medicines Agency. Points to consider on clinical investigation of medicinal products other than NSAIDs for treatment of rheumatoid arthritis [online]. 17.12.2003 [Zugriff: 23.12.2009]. URL: <http://www.ema.europa.eu/pdfs/human/ewp/055695en.pdf>.
38. Sander O. Sicherheit der Biologika: Ergebnisse aus den Registern. *Z Rheumatol* 2010; 69(9): 788-794.
39. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis: Berichtsplan (vorläufige Version); Auftrag A10-01 [online]. 20.09.2010 [Zugriff: 05.10.2010]. URL: http://www.iqwig.de/download/A10-01_vorlaeufiger_Berichtsplan_Biologika_Zweitlinientherapie_bei_heumatoider_Arthritis.pdf.
40. Schneider M, Lelgemann M, Abholz HH, Caratti R, Flügge H, Jäniche H et al. Interdisziplinäre Leitlinie: Management der frühen rheumatoiden Arthritis. Darmstadt: Steinkopff; 2007. URL: <http://dgrh.de/fileadmin/media/Qualitaetssicherung/Leitlinien/gesamt.pdf>.
41. Smolen JS, Landewé 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.
42. Smolen JS, Aletaha D, Bijlsma JW, Breedveld FC, Boumpas D, Burmester G et al. Treating rheumatoid arthritis to target: recommendations of an international task force. *Ann Rheum Dis* 2010; 69(4): 631-637.
43. Boers M. The time has come to limit the placebo period in rheumatoid arthritis trials to 3 months: a systematic comparison of 3- and 6-month response rates in trials of biological agents. *Ann Rheum Dis* 2010; 69(1): 186-192.
44. Aletaha D, Funovits J, Keystone EC, Smolen JS. Disease activity early in the course of treatment predicts response to therapy after one year in rheumatoid arthritis patients. *Arthritis Rheum* 2007; 56(10): 3226-3235.
45. Goekoop-Ruiterman YP, De Vries-Bouwstra JK, Allaart CF, Van Zeben D, Kerstens PJ, Hazes JM et al. Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid arthritis (the BeSt study): a randomized, controlled trial. *Arthritis Rheum* 2008; 58(2 Suppl): S126-S135.
46. Grigor C, Capell H, Stirling A, McMahon AD, Lock P, Vallance R et al. Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): a single-blind randomised controlled trial. *Lancet* 2004; 364(9430): 263-269.
47. Verstappen SM, Jacobs JW, Van der Veen MJ, Heurkens AH, Schenk Y, Ter Borg EJ et al. Intensive treatment with methotrexate in early rheumatoid arthritis: aiming for remission;

Computer Assisted Management in Early Rheumatoid Arthritis (CAMERA, an open-label strategy trial). *Ann Rheum Dis* 2007; 66(11): 1443-1449.

48. Bundesministerium für Gesundheit. Bekanntmachung eines Beschlusses des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie/ AMR in Anlage 4: Therapiehinweis zu Adalimumab vom 21. November 2006. *Bundesanzeiger* 2007; 59(126): 6932.

49. Blumenauer BBTB, Cranney A, Burls A, Coyle D, Hochberg MC, Tugwell P et al. Etanercept for the treatment of rheumatoid arthritis. *Cochrane Database Syst Rev* 2003; (3): CD004525.

50. Blumenauer BBTB, Judd M, Wells GA, Burls A, Cranney A, Hochberg MC et al. Infliximab for the treatment of rheumatoid arthritis. *Cochrane Database Syst Rev* 2002; (3): CD003785.

51. European Medicines Agency. Concept paper on the need for revision of the points to consider on the clinical investigation of medicinal products other than NSAIDs in rheumatoid arthritis (CPMP/EWP/556/95 REV. 1) [online]. 20.01.2010 [Zugriff: 08.03.2011]. URL: http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/02/WC500070422.pdf.

52. Smolen JS, Boers M, Abadie EC, Breedveld FC, Emery P, Bardin T et al. Recommendations for an update of 2003 European regulatory requirements for registration of drugs to be used in the treatment of RA. *Curr Med Res Opin* 2011; 27(2): 315-325.

53. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gotzsche PC, Ioannidis JP et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. *J Clin Epidemiol* 2009; 62(10): e1-e34.

54. Higgins JPT, Green S. *Cochrane handbook for systematic reviews of interventions: version 5.0.2* [online]. 09.2009 [Zugriff: 02.11.2010]. URL: <http://www.cochrane-handbook.org>.

55. Kieser M, Hauschke D. Assessment of clinical relevance by considering point estimates and associated confidence intervals. *Pharmaceut Statist* 2005; 4(2): 101-107.

56. Revicki D, Hays RD, Cella D, Sloan J. Recommended methods for determining responsiveness and minimally important differences for patient-reported outcomes. *J Clin Epidemiol* 2008; 61(2): 102-109.

57. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden zur Bewertung von Verhältnissen zwischen Nutzen und Kosten [online]. 12.10.2009 [Zugriff: 30.11.2010]. URL: https://www.iqwig.de/download/Methodik_fuer_die_Bewertung_von_Verhaeltnissen_zwischen_Nutzen_und_Nutzen.pdf.

58. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet* 2002; 359(9308): 781-785.

59. Borenstein M, Hedges LV, Higgins JPT, Rothstein HR. Introduction to meta-analysis. Chichester: Wiley; 2009.
60. Egger M, Smith GD, Altman DG (Ed). Systematic review in health care: meta-analysis in context. London: BMJ Books; 2001.
61. Whitehead A. Meta-analysis of controlled clinical trials. Chichester: Wiley; 2002.
62. Brockwell SE, Gordon IR. A comparison of statistical methods for meta-analysis. *Stat Med* 2001; 20(6): 825-840.
63. Spiegelhalter DJ, Abrams KR, Myles JP. Bayesian approaches to clinical trials and health-care evaluation. Chichester, England Wiley; 2004.
64. Di Pietrantonj C. Four-fold table cell frequencies imputation in meta analysis. *Stat Med* 2006; 25(13): 2299-2322.
65. Hozo SP, Djulbegovic B, Hozo I. Estimating the mean and variance from the median, range, and the size of a sample. *BMC Med Res Methodol* 2005; 5(1): 13.
66. Parmar MK, Torri V, Stewart L. Extracting summary statistics to perform meta-analyses of the published literature for survival endpoints. *Stat Med* 1998; 17(24): 2815-2834.
67. Bender R, Bunce C, Clarke M, Gates S, Lange S, Pace NL et al. Attention should be given to multiplicity issues in systematic reviews. *J Clin Epidemiol* 2008; 61(9): 857-865.

3 Offenlegung potenzieller Interessenkonflikte

Im Folgenden sind die potenziellen Interessenkonflikte der Stellungnehmenden sowie weiterer Teilnehmer an der wissenschaftlichen Erörterung zusammenfassend dargestellt. Alle Informationen beruhen auf Selbstangabe der einzelnen Personen anhand des „Formblatts zur Offenlegung potenzieller Interessenkonflikte“. Das Formblatt ist unter www.iqwig.de abrufbar. Die in diesem Formblatt aufgeführten Fragen finden sich im Anschluss an diese Zusammenfassung.

3.1 Potenzielle Interessenkonflikte von Stellungnehmenden aus Organisationen, Institutionen und Firmen

Organisation / Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Abbott GmbH & Co. KG	Görtz, Anita	ja	nein	nein	nein	nein	ja
	Simianer, Stefan, Dr.	ja	nein	nein	nein	nein	nein
	Wolff, Malte, Dr.	ja	ja	ja	nein	nein	ja
Bristol-Myers Squibb GmbH & Co. KGaA	Fenzau, Monika	ja	nein	nein	nein	nein	ja
	May, Michael, Dr.	ja	ja	nein	nein	nein	ja
	Neubauer, Aljoscha, Priv. Doz. Dr.	ja	nein	nein	nein	nein	ja
Deutsche Gesellschaft für Rheumatologie	Krüger, Klaus, Prof. Dr.	nein	ja	ja	nein	nein	nein
Deutsche Rheuma-Liga Bundesverband e. V.	Faubel, Ursula	ja	nein	nein	nein	ja	nein
MSD Sharp & Dohme GmbH	Henkel, Katrin, Dr.	ja	nein	nein	nein	nein	nein
	Krobot, Karl J., Dr. PhD	ja	nein	ja	nein	nein	ja
	Peceny, Markus, Dr.	ja	nein	nein	nein	nein	ja
Pfizer Pharma GmbH	Leverkus, Friedhelm	ja	nein	nein	nein	nein	ja

Organisation / Institution / Firma	Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
	Löschmann, Peter-Andreas, Priv. Doz. Dr.	ja	nein	nein	nein	nein	ja
	Reimann, Thomas, Dr.	ja	nein	nein	nein	nein	ja
Roche Pharma AG	Iking-Konert, Christof, Priv. Doz. Dr.	ja	ja	ja	nein	ja	nein
	Neumann, Norbert	ja	nein	nein	nein	nein	ja
	Ruof, Jörg, Dr.	ja	nein	nein	nein	nein	ja
UCB Pharma GmbH	Forst, Hans-Theo, Dr.	nein	ja	nein	nein	nein	ja
	Hilf, Christof, Dr.	nein	ja	nein	nein	nein	nein
	Kreutzenbeck, Hans-Joachim, Dr.	nein	nein	nein	nein	nein	nein
Verband Forschender Arzneimittelhersteller e. V.	Throm, Siegfried, Dr.	ja	nein	nein	nein	nein	nein

3.2 Potenzielle Interessenkonflikte von weiteren Teilnehmern der wissenschaftlichen Erörterung (externe Sachverständige)

Name	Frage 1	Frage 2	Frage 3	Frage 4	Frage 5	Frage 6
Lerch, Christian, Dr.	nein	nein	nein	nein	nein	nein
Richter, Bernd, Prof. Dr.	nein	nein	nein	nein	nein	nein
Sander, Oliver, Dr.	nein	nein	ja	ja	nein	nein

Im „Formblatt zur Offenlegung potenzieller Interessenkonflikte“ wurden folgende 6 Fragen gestellt:

Frage 1: Sind oder waren Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor bei einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter¹ abhängig (angestellt) beschäftigt? Falls ja, wo und in welcher Position?

Frage 2: Beraten Sie oder haben Sie innerhalb des laufenden Jahres und der 3 Kalenderjahre davor einen Interessenverband im Gesundheitswesen oder einen vergleichbaren Interessenvertreter direkt oder indirekt beraten? Falls ja, wen und wie hoch ist / war die Zuwendung / das Honorar?

Frage 3: Haben Sie abseits einer Anstellung oder Beratungstätigkeit innerhalb des laufenden Jahres oder der 3 Kalenderjahre davor im Auftrag eines Interessenverbands im Gesundheitswesen oder eines vergleichbaren Interessenvertreters Honorare für Vorträge, Stellungnahmen, Ausrichtung und / oder Teilnahme an Kongressen und Seminaren – auch im Rahmen von Fortbildungsveranstaltungen, oder für (populär-)wissenschaftliche oder sonstige Aussagen oder Artikel erhalten? Falls ja, von wem, für welche Tätigkeiten und wie hoch war die Zuwendung / das Honorar?

Frage 4: Haben Sie abseits einer Anstellung oder Beratungstätigkeit und / oder hat die Institution², bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres und der 3 Kalenderjahre davor von einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter finanzielle Unterstützung für Forschungsaktivitäten, andere wissenschaftliche Leistungen oder Patentanmeldungen erhalten? Falls ja, von wem, für welche Tätigkeit und in welcher Höhe?

Frage 5: Haben Sie und / oder hat die Institution, bei der Sie angestellt sind bzw. die Sie vertreten, innerhalb des laufenden Jahres oder der 3 Kalenderjahre davor sonstige finanzielle oder geldwerte Zuwendungen (z. B. Ausrüstung, Personal, Reisekostenunterstützung ohne wissenschaftliche Gegenleistungen) von einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter erhalten? Falls ja, von wem, aus welchem Anlass und in welcher Höhe?

Frage 6: Besitzen Sie Aktien, Optionsscheine oder sonstige Geschäftsanteile (auch in Fonds) von einer Firma oder Institution, die zu einem Interessenverband im Gesundheitswesen oder einem vergleichbaren Interessenvertreter gehört? Falls ja, von wem und welchen Wert haben diese aktuell?

¹ Dieses Formblatt erfasst finanzielle Beziehungen zu Interessenverbänden im Gesundheitswesen oder vergleichbaren Interessenvertretern, insbesondere der pharmazeutischen Industrie und der Medizinprodukteindustrie.

² Sofern Sie in einer ausgedehnten Institution tätig sind, ist es ausreichend, die geforderten Angaben auf Ihre Arbeitseinheit (z. B.: Klinikabteilung, Forschungsgruppe etc.) zu beziehen.

4 Dokumentation der wissenschaftlichen Erörterung – Teilnehmerliste, Tagesordnung und Protokoll

4.1 Teilnehmerliste der wissenschaftlichen Erörterung

Name	Organisation / Institution / Firma / privat
Biester, Katharina	IQWiG
Faubel, Ursula	Deutsche Rheuma-Liga Bundesverband e. V.
Forst, Hans-Theo, Dr.	UCB Pharma GmbH
Görtz, Anita	Abbott GmbH & Co. KG
Haag, Susanne, Dr.	IQWiG
Henkel, Katrin, Dr.	MSD Sharp & Dohme GmbH
Hilf, Christof, Dr.	UCB Pharma GmbH
Iking-Konert, Christof, Priv. Doz. Dr.	Roche Pharma AG
Kreutzenbeck, Hans-Joachim, Dr.	UCB Pharma GmbH
Krobot, Karl J., Dr. PhD	MSD Sharp & Dohme GmbH
Krüger, Klaus, Prof. Dr.	Deutsche Gesellschaft für Rheumatologie
Lerch, Christian, Dr.	Heinrich-Heine-Universität Düsseldorf
Leverkus, Friedhelm	Pfizer Pharma GmbH
Löschmann, Peter-Andreas, Priv. Doz. Dr.	Pfizer Pharma GmbH
May, Michael, Dr.	Bristol-Myers Squibb GmbH & Co. KGaA
Mennekes, Beate	Protokollantin (Sitzungsdokumentarischer Dienst, Landtag NRW)
Neubauer, Aljoscha, Priv. Doz. Dr.	Bristol-Myers Squibb GmbH & Co. KGaA
Neumann, Norbert	Roche Pharma AG
Peceny, Markus, Dr.	MSD Sharp & Dohme GmbH
Reimann, Thomas, Dr.	Pfizer Pharma GmbH
Richter, Bernd, Prof. Dr.	Heinrich-Heine-Universität Düsseldorf
Ruof, Jörg, Dr.	Roche Pharma AG
Sander, Oliver, Dr.	Heinrich-Heine-Universität Düsseldorf

Name	Organisation / Institution / Firma / privat
Schürmann, Christoph, Dr.	IQWiG
Sieben, Wiebke, Dr.	IQWiG
Simianer, Stefan, Dr.	Abbott GmbH & Co. KG
Vervölgyi, Volker, Dr.	IQWiG
Werner, Sebastian, Dr.	IQWiG
Wieseler, Beate, Dr.	IQWiG
Windeler, Jürgen, Prof. Dr.	IQWiG
Wolff, Malte, Dr.	Abbott GmbH & Co. KG
Zens, Yvonne	IQWiG

4.2 Tagesordnung der wissenschaftlichen Erörterung

	Begrüßung und Einleitung
TOP 1	Definition der Zweitlinientherapie
TOP 2	Endpunkte a) Teilminderungen der Krankheitsaktivität (wie ACR 20, ACR 50, ACR 70) b) Radiologischer Verlauf
TOP 3	Effektmodifikatoren
TOP 4	Verschiedenes

4.3 Protokoll der wissenschaftlichen Erörterung

- Datum: 17.02.2011, 14:00 bis 15:50 Uhr
- Ort: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Dillenburger Straße 27, 51105 Köln
- Moderation: Prof. Dr. Jürgen Windeler

4.3.1 Begrüßung und Einleitung

Moderator Jürgen Windeler: Ich begrüße die große Runde zu dieser wissenschaftlichen Erörterung bei diesem schönen Wetter – es kommt die Jahreszeit, zu der man die Sonne aussperren muss. Ich will keine langen Vorreden zur Begrüßung halten – die Zeit ist begrenzt – und deswegen direkt zum Thema kommen.

Die Erörterungen finden statt und haben das Ziel, konkrete Fragen zu klären, die beim IQWiG bezüglich der eingereichten Stellungnahmen übrig geblieben sind. Das will ich ausdrücklich sagen. Es geht nicht darum, die Stellungnahmen, die wir in schriftlicher Form erhalten haben, zu wiederholen – Sie können im Grundsatz davon ausgehen, dass wir sie verstanden haben –, sondern es geht darum, die Dinge, die für uns unklar geblieben sind, mit Ihnen zu erörtern. Dazu gibt es eine Tagesordnung, wie Sie sehen, auf der die Punkte im Wesentlichen aufgelistet sind. Konkrete Fragen folgen dann gleich. Unter dem TOP „Verschiedenes“ gibt es die Möglichkeit, Dinge über die Detailpunkte hinaus, die hier aufgelistet sind, zu erörtern. Ich möchte Sie fragen, ob Sie Ergänzungswünsche zu der Tagesordnung haben.

Jörg Ruof: Vielleicht kann man unter dem letzten TOP noch die 12 Wochen, die Therapiedauer ansprechen.

Moderator Jürgen Windeler: Gut, das können wir uns aufschreiben.

Karl J. Krobot: Von unserer Seite geht es um die Frage: Sind auch indirekte Vergleiche vorgesehen?

Moderator Jürgen Windeler: Ja. – Noch ein Hinweis?

Friedhelm Leverkus: Vielleicht können wir unter dem TOP „Verschiedenes“ darüber diskutieren, ob es die Möglichkeit gibt, wenn Sie den Plan nach diesem Meeting finalisieren, noch einmal darüber zu reden. Das Indikationsgebiet ist sicherlich sehr schwierig. Daher waren vielleicht nicht alle Dinge im Detail spezifiziert, wie wir es von den sonstigen Berichtsplänen kennen.

Moderator Jürgen Windeler: Ich nehme das im Moment als Thema auf. Sie kennen natürlich das übliche Vorgehen im IQWiG.

(Friedhelm Leverkus: Klar!)

Dann muss ich Sie zu Beginn darauf hinweisen, dass die Erörterung aufgezeichnet wird. Das ist noch nicht so schlimm, sie wird aber auch veröffentlicht, und zwar im Wortlaut. Sie sind bei der Einladung auch darauf hingewiesen worden, dass Sie mit Ihrem Herkommen, mit Ihrer Teilnahme hier dieser Veröffentlichung zustimmen. Wenn sich das jetzt noch jemand überlegen möchte und anders sieht, müsste er den Raum verlassen und könnte an der Erörterung nicht teilnehmen. – Das könnte es alles geben.

In dem Zusammenhang habe ich auch die Bitte, dass Sie bei Wortmeldungen jedes Mal deutlich Ihren Namen sagen, damit die Protokollierung sowohl schriftlich als auch auf dem Band mitkommt. – Frau Wieseler, ich habe nichts vergessen?

(Beate Wieseler: Nein!)

Gut, danke. Dann komme ich zu:

4.3.2 Tagesordnungspunkt 1: Definition der Zweitlinientherapie

Ich übergebe an Frau Biester. Bitte schön.

Katharina Biester: Der Begriff der Zweitlinientherapie wird im Bereich der rheumatoiden Arthritis nicht einheitlich verwendet, was uns auch in den Stellungnahmen bestätigt wurde. Die Zweitlinientherapie als Therapie nach Versagen einer Vorbehandlung ist von der Erstlinientherapie als Therapie von behandlungsnaiven Patienten abzugrenzen. Die Frage ist, ob die Definition der Zweitlinientherapie auch eine Vorbehandlung mit mehreren krankheitsmodifizierenden Antirheumatika, darunter auch biotechnologisch hergestellte, umfassen sollte. Das würden wir jetzt gerne mit Ihnen diskutieren und Sie um Ihre Argumente bitten.

Malte Wolff: Wir sind der Meinung, dass es eine klare Dreigliedrigkeit zwischen naiven Patienten, vollständig naiven Patienten, mit synthetischen DMARDs vorbehandelten Patienten und biologisch vorbehandelten Patienten gibt. Das zeigt sich in der Zulassung, in den Fachinformationen und auch in den Veröffentlichungen, die Sie bei der EMA finden, die dies teilweise explizit als Erst-, Zweit- und Drittlinientherapie benennen. Das spiegelt sich letztlich beispielsweise auch in den EULAR-Guidelines wider. Dort wird es als Phasen bezeichnet, aber teilweise auch als Linien. Dementsprechend würden wir schon die Unterscheidung machen, dass entsprechend der Zulassung in der Erstlinientherapie MTX-naive Patienten, in der Zweitlinientherapie biologikanaive und DMARD-vorbehandelte Patienten und in der Drittlinientherapie biologikavorbehandelte Patienten sind.

Beate Wieseler: Das würde Sinn machen, wenn es auch unterschiedliche Patientengruppen charakterisieren würde. Wenn Sie sagen, die Erstlinie ist die Behandlung mit synthetischen DMARDs, macht es für Sie einen Unterschied, ob ein Patient mit einem synthetischen DMARD vorbehandelt ist und dann ein Biologikum erhält oder ob er mit mehreren synthetischen DMARDs vorbehandelt ist und dann ein Biologikum erhält?

Malte Wolff: Im Prinzip ist es so, dass von der Zulassung her, wie gesagt, die Dreigliedrigkeit gegeben ist. Was die Vorbehandlung mit DMARDs angeht – wenn ich von DMARDs spreche, meine ich synthetische DMARDs –, sind wir der Meinung, dass man die Anzahl der zur Vorbehandlung eingesetzten DMARDs in der Kategorie als Effektmodifikator nehmen könnte.

Beate Wieseler: Eine Nachfrage auch an die anderen Experten am Tisch: Was charakterisiert die Patienten im Unterschied zueinander? Ist ein Patient, der mit mehreren DMARDs vorbehandelt ist, in seinem Schweregrad der Erkrankung zum Beispiel gleichzusetzen mit einem Patienten, der nur ein DMARD bekommen hat und dann auf ein Biologikum geht? Oder wie verhält sich das im Vergleich zu einem Patienten, der ein synthetisches DMARD erhalten hat, dann auf ein Biologikum geht und dann auf ein zweites Biologikum? Sind diese Patienten in ihrem klinischen Bild tatsächlich klar voneinander abzugrenzen?

Klaus Krüger: Ich kann noch einmal präzisieren, wie die Therapie normalerweise abläuft. Nach unseren Therapieempfehlungen, die wir von der Fachgesellschaft haben, sieht es so aus, dass man mit einer Starttherapie – Methotrexat in der Regel – beginnt und dann in der Regel als Zweites eine DMARD-Kombinationstherapie hat. Es gibt eine Sonderregelung bei Patienten mit sehr schlechter Prognose und hoher Aktivität. Dort überspringt man die Kombinationstherapie. Es gibt für eine kleine Gruppe von Patienten die Möglichkeit, dass sie als zweite Therapie ein Biologikum bekommen, aber normalerweise – das zeigen auch die Zahlen vom Deutschen Rheuma-Forschungszentrum – wird eine Kombinationstherapie noch nach der DMARD-Therapie – wobei ich mehrere DMARDs meine, die Kortikoide sind außen vor, sie werden sowieso kombiniert – verabreicht. Das ist der normale Ablauf.

Natürlich kann man sagen, dass die Anzahl der Vortherapien die Ansprechwahrscheinlichkeit immer ein wenig reduziert, weil eine gewisse Selektion von Nichtansprechen auf A, Nichtansprechen auf A plus B stattfindet. Die Anzahl der Vortherapien gibt schon einen gewissen Hinweis darauf, welche Erfolgsaussicht gegeben ist.

Christof Iking-Konert: Ich möchte das, was Herr Krüger sagte, noch ergänzen. Sie sagten, es gibt eine Krankheit „rheumatoide Arthritis“. Der Begriff ist erst einmal gleich, wir sehen in der Klinik aber schon, dass es verschiedene Patienten gibt. Es mag wahrscheinlich genetisch terminiert und auch pathophysiologisch verschiedene Erkrankungen der rheumatoiden Arthritis geben, die dann unterschiedlich laufen und auch auf unterschiedliche Therapiealgorithmen ansprechen. Jetzt sind wir momentan noch nicht so weit, dass wir diese Patienten von vornherein gut charakterisieren können. Die EULAR hat allerdings in ihren Empfehlungen schon genau das aufgegriffen, was Herr Krüger gesagt hat, nämlich dass ein Patient mit einem hohen Risikofaktor – dazu zählen zum Beispiel der Antikörperstatus, der CCP-Nachweis, ein hohes CRP, das Bereits-Vorhandensein von Erosionen – früher auf eine biologische Therapie gehen sollte und ein Patient, der diese Risikofaktoren nicht hat, eher auf konventionellen DMARDs, auf Kombinationstherapien weitergeführt werden sollte.

Um das noch einmal zu konkretisieren: Es ist vom Namen her eine Krankheit, aber vom klinischen Verlauf her sehen wir verschiedene Ras bei den Patienten.

Moderator Jürgen Windeler: Gibt es weitere Wortmeldungen zu dem Punkt? – Herr Richter.

Bernd Richter: Ich möchte nur darauf hindeuten: Zum einen müssen wir uns im Rahmen des Zulassungsstatus bewegen, der schon relativ genau festschreibt, was wir in welchem Fall zu untersuchen haben, je nach Substanz. Zum anderen ist der Berichtsplanentwurf bewusst klinisch gehalten. Will heißen: Mit dem Begriff der Zweitlinientherapie, der nirgendwo wirklich scharf definiert ist – es gibt die verschiedensten Definitionen –, ist eigentlich die Nichterstlinientherapie gemeint, sodass wir zwischen Patienten, die mit synthetischen DMARDs behandelt worden sind, dann ein Versagen haben und anschließend ein Biologikum erhalten, und einer anderen Gruppe, die sich klinisch und in anderen Aspekten unterscheidet,

die einen TNF-alpha-Inhibitor hat, darunter versagt und dann mit einem anderen Präparat behandelt werden muss, differenzieren würden. Diese beiden Hauptgruppen gibt es. Ob man sie „Zweitlinie“ oder wie auch immer nennt, ist letzten Endes sekundär. Das Primäre ist, dass es nicht die Erstlinientherapie als solche ist.

Beate Wieseler: Dazu möchte ich ergänzen, dass es auch Präparate gibt, die bei sehr schweren Formen in der Erstlinie zugelassen sind. Wir wollen in der Tat nicht diese Erstlinientherapie bewerten, sondern eine Therapie nach Therapieversagen, deshalb diese Art der Abgrenzung im Berichtsplan. Es geht, wie Herr Richter ganz richtig sagt, primär um die Nichterstlinientherapie. Wie dann die Zweitlinientherapie gestaltet wird, auch als Effektmofifikator, muss man sich sicherlich anschauen; da gebe ich Ihnen vollkommen recht. Das bestimmt aber nicht, ob ein Präparat in die Bewertung einfließen sollte oder nicht. Bei der Definition der Zweitlinientherapie, wie wir sie im Berichtsplan vorgenommen haben, geht es um die Abgrenzung zu den Erstlinientherapien, die für 3 der Präparate auch zugelassen sind.

Malte Wolff: Ich möchte noch einmal nachfragen: Wie würden Sie dann mit einem Präparat verfahren, das erst nach Biologikaversagen zugelassen ist, aber nicht nach normalem synthetischem DMARD-Versagen?

Beate Wieseler: Wir würden uns die Daten für dieses Präparat anschauen und eine Aussage zum Präparat in der Situation machen. Wir werden den Zulassungsstatus berücksichtigen und in der Beschreibung des Nutzens der Präparate die Aussagen zu dem Präparat auf diese Situation beschränken.

Jörg Ruof: Es gibt 2 Fragen hinsichtlich der Zweitlinientherapie. Die eine ist: Was kommt davor? Die andere: Was kommt danach? Jetzt haben wir ein bisschen elaboriert: Was kommt davor? Es ist die Frage von ein oder 2 synthetischen DMARDs oder wie auch immer. Sie haben auf einige Produkte hingewiesen, die bereits ganz vorne eingesetzt werden können.

Zu der Frage: Was kommt danach? – Die Evidenzlage ist ganz unterschiedlich. Sie haben für die sogenannte Erstbiologikatherapie, wenn man es so nennen will, die meisten Zulassungsstudien, und Sie haben eine relativ geringe RCT-Evidenzlage in der Frage danach. Wie gedenken Sie damit umzugehen? Werden Sie das zusammenmischen oder werden Sie das getrennt beurteilen, letztendlich auch die Frage von Cycling der Biologika?

Beate Wieseler: Wir werden das sicherlich berücksichtigen. Wenn wir zu einer bestimmten Fragestellung wenig Evidenz haben, dann können wir nur entsprechende Aussagen machen. Es macht aber sicherlich Sinn, wie eben schon gesagt, dass man die Vorbehandlung – das wäre in diesem Fall auch eine Vorbehandlung, die Patienten sind bereits mit einem Biologikum vorbehandelt – auch bei der Betrachtung berücksichtigt.

Moderator Jürgen Windeler: Gibt es weitere Wortmeldungen zu diesem Punkt? – Bitte schön.

Stefan Simianer: Ich habe jetzt verstanden, dass Zweitlinie Einsatz eines s- oder bDMARDs nach dem Versagen der ersten Systemtherapie heißt. Wenn die erste Systemtherapie nicht Methotrexat war, wäre in der Situation dann Methotrexat als Zweitlinientherapie für Sie denkbar? Das geht noch einmal auf den Komplex zurück, ob der Begriff „Zweitlinientherapie“ in der Literatur und was den Zulassungsstatus und Expertenmeinungen angeht nicht tatsächlich für den Ersteinsatz eines Biologikums vorgesehen ist.

Beate Wieseler: Ich weiß nicht, ob mir Ihre Frage klar geworden ist, weil wir das MTX nicht bewerten. Wir haben gegebenenfalls MTX in der Vergleichsgruppe, aber ich bin nicht sicher, ob ich Ihre Frage unter dem Aspekt, dass wir MTX selbst nicht bewerten, verstehe.

Stefan Simianer: Die Frage zielt darauf ab, dass Patienten nicht immer automatisch primär mit MTX behandelt werden. Das ist zwar die Empfehlung, aber es gibt auch Patienten, die zuerst mit Sulfasalazin behandelt werden. Nach Ihrer Definition wäre dann die Zweitlinie im Grunde die Erstbehandlung mit MTX.

Beate Wieseler: Nein.

Stefan Simianer: Nein?

Bernd Richter: Der gesamte Körper wird im Normalfall, wie er sich auch in der Klinik darstellen würde, von allen synthetischen DMARDs gleichzeitig betrachtet, ob in Kombination oder einzelnen – dann meistens MTX –, wenn es noch Leitlinien geben würde. Beim Versagen dieser Therapie wird dann additiv meistens ein Biologikum gegeben. Das wäre eine „klassische“ Zweitlinientherapie, schließt allerdings nicht ein TNF-Versagen und dann ein anderes Biologikum aus. Deswegen haben wir gesagt: nicht Erstlinientherapie gleich Zweitlinientherapie.

Markus Peceny: Wurde es vom Auftraggeber dieses Bewertungsverfahrens klarer definiert? War das tatsächlich der Auftrag, zu sagen: „Zweitlinientherapie“?

Beate Wieseler: Der Auftrag ist als „Zweitlinientherapie mit Biologika“ benannt. Die Zweitlinientherapie mit Biologika schließt durchaus ein, dass ich vorher mehrere synthetische DMARDs verwendet habe. Das sehen wir durch den Auftrag gedeckt.

Markus Peceny: Die Frage ist, ob die nach unseren Begriffen oder in vielen Facheditorials auch als Drittlinientherapie bezeichnete Therapie nach erstem Biologikaeinsatz tatsächlich klarer von dem Auftrag umfasst war.

Beate Wieseler: Weil, wie gesagt, die ganze Nomenklatur zur Zweit- und Drittlinientherapie nicht scharf definiert ist, fassen wir unter dem Auftrag „Zweitlinientherapie mit Biologika“ die eben beschriebenen Szenarien zusammen. Das ist vom G-BA nicht in irgendeiner Art und Weise eingeschränkt worden.

Peter-Andreas Löschmann: Ich glaube, das ist genau der Punkt. Es besteht Konsens, dass man als Erstlinientherapie jede Form von synthetischen DMARDs ansehen würde. Die Frage ist: Bedeutet eine danach erfolgte Zweitlinientherapie mit einem Biologikum, die zum Therapieversagen führt, dass der nächste Stoff, vermutlich wieder ein Biologikum – das ist der Gegenstand des Auftrags – dann immer noch eine Zweitlinientherapie ist? Ich glaube, es ist so nicht haltbar, dass man sie dann quasi zusammen tut und als Zweitlinientherapie bezeichnet. Das widerspricht der Nomenklatur und, ich denke, auch dem Verständnis der Zulassungsbehörden.

Beate Wieseler: Wie gesagt, nach unserer Sichtung der Literatur und der Leitlinien gibt es keine klare Abgrenzung von Zweit- und Drittlinientherapie. Dessen ungeachtet wird man die unterschiedlichen Szenarien in der Aussage zu den einzelnen Präparaten natürlich berücksichtigen, weil sich unsere Bewertungen, das wissen Sie, im Rahmen des Zulassungsstatus bewegen.

Aljoscha Neubauer: Ich denke, es ist nur wichtig, ganz klar festzustellen, dass man, wenn wir über Zweit- oder Drittlinie reden – in der Regel geht es dann um vorheriges Versagen, TNF-alpha-Versagen –, die Gruppen gesondert betrachtet und nicht am Ende zusammenmischt. Wie wir vorhin gehört haben, gibt es einen Unterschied bei den Outcomes der Patientenpopulationen, weil sie auch unterschiedliche Wahrscheinlichkeiten haben. Das ist der Punkt, auf den es uns dabei ankommt.

Beate Wieseler: Ja, das ist angekommen. Ich kann Ihnen nur noch einmal sagen: Weil wir innerhalb des Zulassungsstatus bewerten, werden wir die Situation der Präparate, die erst nach TNF-alpha-Inhibition zugelassen sind, auch in dieser Form beschreiben.

Jörg Ruoff: Sie nehmen in Ihrem Berichtsplan Bezug auf die EULAR-Empfehlungen, die die Dreierklassifikation zitieren, die bereits mehrfach genannt wurde, unter anderem von Prof. Krüger. Das wurde kürzlich in der GREES-Publikation, also den aktualisierten Guidelines, den Empfehlungen der Expertengruppe, gewissermaßen nachgedoppelt und noch einmal ganz klar als drei Elemente und drei differenzierte Gruppen spezifiziert. In diesem Sinne wäre bei der zweiten Biologikatherapie eigentlich von einer Drittlinientherapie zu sprechen. – Ich denke, Herr Prof. Krüger, das ist auch Ihr Verständnis der Situation, oder?

Klaus Krüger: Das wird mehrheitlich so gehandhabt. Man muss sich aber nicht um Namen streiten. Wichtig ist, dass festgehalten wird, wie viele Vortherapien gelaufen sind, wenn eine Studie bewertet wird. Das ist das Entscheidende, weil die Erfolgsaussicht dann in der Tat etwas geringer wird.

Malte Wolff: Ich möchte nur noch einmal festhalten: Der Moment, wenn Sie in die Drittlinien gehen bzw. nach TNF-Versagen, ist unserer Meinung nach nicht Teil des Auftrags, den Sie bekommen haben; denn das entspricht nicht dem, was die Zulassungsbehörden haben, so wie die Dreigliedrigkeit auch in den Fachinformationen und in den EPARs von diversen

Präparaten steht. Es ist aus meiner Sicht eine Frage des Auftrags und des Zulassungsstatus. Die Evidenz ist sicherlich vorhanden – wir müssen uns nicht darüber streiten, dass sie auch unterschiedlich ist –, aber es ist durch den Auftrag nicht gedeckt, weil der „Zweitlinientherapie“ heißt.

Beate Wieseler: Ich möchte noch einmal wiederholen, dass bei der Diskussion dieses Auftrags das Ganze als Abgrenzung zur Erstlinientherapie, die für 3 Präparate möglich ist, gesehen wurde. Die Erstlinientherapie soll nicht in die Bewertung eingeschlossen werden. Ansonsten sollen die Biologika gemäß ihres Zulassungsstatus bewertet werden.

Bernd Richter: Abgesehen von semantischen Schwierigkeiten, Zweit- und Drittlinie usw., festzuhalten, sollten wir vielleicht daran denken, dass es im Interesse der Betroffenen ist, eine möglichst weitgehende Darstellung und optimale Beurteilung der Evidenzlage zu gewährleisten. Ich würde gern von den Anwesenden wissen, ob – abgesehen davon, dass wir vielleicht keine hundertprozentige Definition des Begriffs haben – etwas dagegensprechen würde, es in dem Sinne wie vorgetragen zu interpretieren und auch auszuwerten, damit man für die Betroffenen ein optimales Ergebnis, eine optimale Transparenz hat. Das ist doch das primäre Ziel des Verfahrens.

Friedhelm Leverkus: Ich habe es richtig verstanden, dass 2 unterschiedliche Analysen laufen, zwei unterschiedliche Subgruppen gebildet werden, Frau Wieseler? Wenn Sie die Zweitlinie nehmen, dann ist das eine Analyse mit der Zweitlinie, ein Versagen nach TNF-alpha wäre eine andere Analyse, und die beiden Gruppen würden nicht gepoolt werden?

Beate Wieseler: Das ist richtig. Ich würde das nicht als 2 Subgruppen beschreiben, sondern sagen: Das ist der Einsatz von Biologika nach einem Therapieversagen, also bei nicht therapienaiven Patienten. Dass wir gegebenenfalls unterschiedliche Patientenpopulationen und unterschiedliche Vorbehandlungen haben, werden wir beschreiben. Wir würden aber auch beschreiben, wenn wir unterschiedliche synthetische DMARDs in der Vorbehandlung hätten, weil es eine Rolle spielt, ob ich ein Therapieversagen nach einem synthetischen DMARD habe oder ob vorher schon verschiedene synthetische DMARDs versagt haben. Die Vorbehandlung – das haben wir auch so spezifiziert – wird bei der Bewertung sicherlich zu betrachten sein.

Friedhelm Leverkus: Aber der qualitative Sprung zwischen mehreren DMARDs und Biologika scheint doch größer zu sein. Der Unterschied sowohl nach Zulassungsstatus als auch aus medizinischer Sicht scheint doch zu sein, dass dadurch zwei unterschiedliche Populationen definiert werden. Somit wäre auch eine getrennte Betrachtung dieser Populationen sinnvoll, sofern man sie in dem Auftrag anschauen würde.

Bernd Richter: Ich wäre Ihrer Meinung. Niemand würde vernünftigerweise die beiden Populationen poolen, ganz abgesehen davon, dass wir große Schwierigkeiten haben werden, dann überhaupt etwas zu poolen, weil die Kollektive wahrscheinlich extrem heterogen sind.

Das sind ganz andere Krankheitsschweregrade und Kollektive, wenn Sie so wollen, und damit auch unterschiedliche Betrachtungsweisen. Es ist auch, wie gesagt wurde, keine Subgruppe, sondern es sind primär einfach andere Kollektive.

Markus Peceny: Um auf Ihre Frage von vorhin einzugehen, Herr Richter, ob ein Problem in der umfassenden Bewertung der Evidenzlage gesehen würde: Ich würde vielleicht ein Problem beim praktischen Einsatz der Präparate und deren Ergebnis sehen, gerade bei den Nicht-TNF-Blockern. Die anderen Biologicals wurden öfter nach TNF-Versagen untersucht, in der Praxis wird nach dem ersten TNF-Versagen aber oft ein anderer TNF-Blocker eingesetzt. Das steht auch so in den EULAR-Leitlinien, zumindest als gleichwertige Option. Letzten Endes würde da möglicherweise eine optische Evidenzlücke entstehen, die vom praktischen Leben so nicht widerspiegelt wird – belegt mit einem mangelnden Nutzen für eine Therapie, die eigentlich tagtäglich mit Erfolg eingesetzt wird. Auch wenn es formal richtig sein mag, die Evidenzlage so zu bewerten, wäre das, wenn man den tatsächlichen Nutzen bewerten will, eine starke Diskrepanz zur Praxis.

Bernd Richter: Natürlich, aber das könnte auch wunderbar eine Evidenzlücke aufzeigen, bei der gesagt werden muss: Dann muss man es vernünftig untersuchen. Denn wenn die Praxis der Evidenz fortläuft und der Bogen irgendwann zu weit auseinandergeht, kommt man in Probleme, weil man mit anderen Kombinationspartnern sehr gute Erfolge sehen würde. Dann muss man irgendwann einmal die entsprechende Untersuchung machen, sonst wird es schwierig, das als „normale“ Praxis aufrechtzuerhalten.

Jörg Ruof: Da ist unser Anliegen ganz klar, dass Sie an Ihren eigenen Richtlinien – RCT-Evidenz – festhalten. Wir planen eine Head-to-Head-Studie in dem Bereich. Es gibt eine klinische Praxis, die ihre Gründe hat, es gibt aber auch eine Evidenzlage. Unser Anliegen ist natürlich, dass Sie an Ihren eigenen Anforderungen festhalten.

Moderator Jürgen Windeler: Das Anliegen nehmen wir auf. – Gibt es weitere Bemerkungen zu der Abgrenzung und Definition der Zweitlinientherapie? – Das ist nicht der Fall. Unsererseits sind die Fragen beantwortet?

(Beate Wieseler: Ja!)

Gut. Dann kommen wir zu:

4.3.3 Tagesordnungspunkt 2: Endpunkte, a) Teilminderungen der Krankheitsaktivität (wie ACR 20, ACR 50, ACR 70)

Katharina Biester: Der TOP „Endpunkte“ teilt sich in 2 Unterpunkte auf. Im ersten geht es um Teilminderungen der Krankheitsaktivität. Im vorläufigen Berichtsplan wurde für die Krankheitsaktivität die Remission als patientenrelevanter Endpunkt definiert. In den Stellungnahmen wurde angemerkt, dass auch Teilminderungen der Krankheitsaktivität, wie sie beispielsweise durch den ACR 20 beschrieben werden, patientenrelevante Endpunkte darstellen. Wir bitten die Stellungnehmenden, zu erläutern, warum eine Teilminderung der Krankheitsaktivität aus Ihrer Sicht patientenrelevant ist.

Ursula Faubel: Auch wir hatten angemerkt, dass das aus unserer Sicht sinnvoll ist. Die Patientenvertreter, die angehört worden waren, haben bereits in der Voranhörung gesagt, dass sie das für einen wichtigen Parameter halten. Natürlich sind Patienten am meisten daran interessiert, dass es so weit wie möglich weggeht, aber für viele Patienten, die die Erkrankung schon länger haben, ist das Ziel mit den im Moment vorhandenen Möglichkeiten nicht mehr erreichbar. Für sie ist es ganz wichtig, dass wenigstens ein Zwischenziel erreicht wird. Deswegen haben wir gesagt, dass auch das Zwischenziel in irgendeiner Form als Endpunkt definiert werden muss, und hatten vorgeschlagen, es anhand des DAS zu machen, wenn er $\leq 3,2$ ist.

Stefan Simianer: Wir können das nur unterstützen. Ein Argument wäre auch, dass man bei der Erkrankung im fortgeschrittenen Stadium mit irreversiblen Veränderungen konfrontiert ist und man eine Remission in der Situation, in der eine Mischung aus Symptomatik aus den irreversiblen Schäden und der aktuellen Entzündungsaktivität vorliegt, durch eine Therapie, die primär die Entzündungsaktivität zum Ziel hat, nicht mehr erreichen kann. Aus methodischen Gründen wäre es daher aus unserer Sicht klug, gerade bei fortgeschrittenen Patienten, mit denen wir es hier zu tun haben, ein Kriterium einzuführen, das die Ausgangsaktivität als Bezug nimmt und dann eine Verbesserung auf Basis der Ausgangsaktivität misst. Das wäre bei den ACR-Kriterien der Fall, wobei wir auch mit den DAS-Kriterien sehr gut leben könnten, die allerdings Schwellenwerte zum Ziel haben und nicht eine Beziehung zum Ausgangsbefund herstellen. Wir würden vorschlagen, dass man die ganzen Composite Scores mit in die Analyse einbezieht.

Christof Iking-Konert: Ich möchte das gerne aus Therapeutensicht bestärken und bekräftigen, dass es in der klinischen Praxis genauso ist. Sie werden nicht mehr bei allen Patienten einen definierten Endpunkt, so erstrebenswert er auch sein mag, erreichen. Für viele Patienten ist die Reduktion von einem sehr hohen Krankheitsniveau – lassen Sie es ein DAS von 7 sein, das auf ein DAS von 3,5 fällt – in der Relation viel relevanter als möglicherweise die Reduktion von 3,5 dann in den Remissionsbereich unter 2,6. Es ist zwar der Endpunkt erreicht, aber die relative Reduktion und damit der Nutzen der Substanz für den Patienten ist

relativ gesehen viel, viel geringer. Also ist es für mich als Kliniker und Therapeut eine Selbstverständlichkeit, dass Sie die relative Reduktion von Krankheitsaktivität beurteilen.

Oliver Sander: Im Prinzip sehe ich das als Kliniker genauso, aber wir müssen auch Messgrößen haben. Je mehr Messgrößen wir haben, desto mehr verlieren wir uns in unterschiedlichen Messgrößen, unterschiedlichen Subpopulationen. Wir können Remission messen. Remission ist die Spitze des Eisbergs. Diese Spitze ist nur sichtbar, wenn darunter verschiedene andere nutzenrelevante Besserungen auftreten. Das heißt, wir können nicht 30 % Remission haben, aber nur 20 % ACR-20-Response.

Im Allgemeinen liegt unter der Spitze Remission immer ein großer nicht sichtbarer Berg, der sicherlich proportional zu dem ist, was wir draußen sehen. Es gibt in jeder Studie einen gewissen Proporz an Remission, der – unstrittig – nach vielen Vortherapien geringer ist, aber auch dort als Nutzen sichtbar ist und diesen auch im Ausmaß der Behandlung insgesamt sehr gut widerspiegelt, sodass man sich auf eine Zielgröße einigen kann. Dann kann man die unterschiedlichen Subpopulationen bezüglich ihrer Wahrscheinlichkeit, diese optimale Zielgröße zu erreichen, noch einmal würdigen, aber nicht verschiedene Messgrößen und verschiedene Subpopulationen, die die Wertung dann erschweren.

Beate Wieseler: Ich möchte auf 2 Punkte aufmerksam machen. Zum einen ist Remission einer von mehreren Endpunkten, die wir genannt haben. Was die Krankheitsaktivität angeht, haben wir die Remission gewählt und es auch darauf beschränkt, weil es nach unserer Kenntnis der Literatur nur für die Remission einen einigermaßen sicheren Nachweis gibt, dass die Absenkung der Krankheitsaktivität auf dieses Niveau die Gelenkzerstörung, den Funktionsverlust tatsächlich langfristig verhindern kann. Das heißt, das ist für uns das eine Maß für die Krankheitsaktivität, von dem wir relativ sicher wissen, dass der Patient davon profitiert.

Ungeachtet dessen haben wir weitere Parameter, zum Beispiel Schmerz oder eine andere Symptomatik, und bilden für den Patienten auch sichtbare Teilminderungen ab. Es ist nicht so, dass wir dadurch, dass wir den Parameter Remission wählen, für die Patienten, die die Remission eigentlich nicht mehr erreichen können, Änderungen nicht sichtbar machen können. Wir sehen die Teiländerungen in den anderen Parametern in den Funktionsfragebögen oder Symptomatikerhebungen auch für die Patienten, die keine Remission mehr erreichen können.

Gleichzeitig möchten wir für die Gesamtkrankheitsaktivität ein Maß wählen, von dem wir relativ sicher wissen, dass es tatsächlich einen Patientennutzen abbildet. Das wissen wir beim ACR 20 nicht. Wir wissen nicht, ob Patienten, die einen ACR 20 erreichen, tatsächlich einen langfristigen Nutzen davon haben. Deshalb, noch einmal, gibt es für den Komplex Krankheitsaktivität die Beschränkung auf Remission, und es erfolgt durchaus die Abbildung von Teilminderungen im Bereich Symptomatik, Funktionsfragebögen, also im Bereich der anderen Endpunkte.

Bernd Richter: Ich würde auch sagen, dass eigentlich nichts verloren geht, weil wir ein sehr umfassendes Bild davon haben, was patientenrelevant ist, und alle Teilparameter sowieso berücksichtigt werden. Es wird auch von niemandem, der klinisch ein wenig Ahnung hat, bestritten, dass so etwas wie ein ACR 20 oder 50 durchaus Patientenrelevanz hat. Das ist nicht das Problem.

Das primäre Problem ist, dass die Operationalisierung, abgesehen von der Ableitung der Größen auf spätere Funktionsverluste oder Zustände, schwierig ist. Denn wenn wir ein Remissionskriterium haben – sagen wir, ein DAS28 kleiner 2,6 –, was wahrscheinlich eher eine minimale Erkrankungsaktivität denn eine wirkliche Remission klinischer Art ist, dann haben wir einen Statuszustand. Bei einer ACR-Differenzmessung zwischen 2 verschiedenen Zeitpunkten haben wir ebenso wie bei Absenkungen von DAS die Schwierigkeit, verschiedene Patienten zu vergleichen, weil wir letzten Endes individualisieren. Jemanden, der einen ACR 50 mit einer höheren Krankheitsaktivität, mit einer schwereren rheumatoiden Arthritis hat, können Sie nicht mit einem Menschen, der dasselbe Phänomen, aber eine leichtere Form hat, vergleichen. Der Parameter ist zwar wichtig für die Zulassung etc. – so ist er ursprünglich geschaffen worden –, aber in der Operationalisierung für einen richtig patientenrelevanten Parameter schwer zu fassen, weswegen wir in erster Linie die Remissionskriterien unterschiedlichster Art befürworten würden.

Klaus Krüger: Ich möchte auch noch einmal unterstreichen, dass der DAS28 sicher der bessere Parameter ist, weil es sich bei dem anderen um eine prozentuale Besserung handelt. Wenn Sie eine ACR-20-Response bei einem Patienten haben, der 20 geschwollene Gelenke hat, dann sind es hinterher 16. Das ist nach wie vor eine völlig unakzeptable Aktivität. Deswegen ist der DAS28 mit all seinen Schwächen – wir wissen, dass die Patienten zum Teil nicht wirklich in Remission sind, wenn sie sie scheinbar erreichen – für den einzelnen Patienten sicher der relevantere Parameter, wobei man auch, wie Frau Faubel sagte, die Low Disease Activity einbeziehen muss. Das ist für viele Patienten das maximal Erreichbare, wenn die Krankheitsdauer entsprechend lang ist, wenn sie ein zusätzliches Schmerzproblem haben. Das wird man als ein relevantes Therapieziel mit einbeziehen müssen. Sie werden die ACR-Response mit berücksichtigen müssen, weil die alten Studien aus der früheren Zeit der Biologikatherapie kein DAS28, also keine Remission, erfasst haben. Der Parameter taucht in den Studien von Anfang dieses Jahrzehnts zum Teil noch gar nicht auf. Das heißt, man wird es in irgendeiner Weise noch mit berücksichtigen müssen.

Friedhelm Leverkus: Ich möchte darauf hinweisen, dass wir nicht in der Situation sind, eine Studie zu planen, den besten Endpunkt herauszusuchen und zu beurteilen, sondern wir sind in einer retrospektiven Situation. Das heißt, wir schauen: Welche Daten sind da? Welche Schlüsse können wir daraus ziehen? Hier noch einmal der Appell, dass wir die vorhandene Evidenz berücksichtigen und interpretieren und nicht vorzeitig Evidenz ausschließen, die uns auch wichtige Dinge sagen kann. – Punkt eins.

Punkt 2: Die Diskussion über relative Veränderungen oder Absolutwerte trifft nicht nur auf das Indikationsgebiet zu, sondern auf sehr viele Indikationsgebiete, wenn man zu Schmerz forscht oder zu andere Dingen. Wie gesagt, es ist sicherlich patientenrelevant, wenn ein gewisser Punkt überschritten wird, aber es ist auch patientenrelevant, wenn eine Reduktion da ist, auch wenn dieser Punkt nicht erreicht wird. Davon profitieren Patienten auch. Daher sollte man nicht vorzeitig Dinge ausschließen, sondern man sollte beides betrachten. Wenn es Unterschiede gibt, muss man sie interpretieren.

Peter-Andreas Löschmann: Ich habe eine Frage zum Vorgehen. Erwägen Sie tatsächlich, sich auf Studien zu beschränken, die Remission als Endpunkt hatten? Wenn das so wäre, glaube ich, wird die Evidenz für alle Biologika zusammengenommen sehr überschaubar.

Beate Wieseler: Wie ich eben schon ausgeführt habe, ist Remission nicht der einzige Endpunkt, den wir betrachten. Wir haben zum Beispiel Schmerz als patientenrelevanten Endpunkt aufgenommen. Jede Studie, die einen ACR 20 zeigt, verfügt natürlich über Schmerzdaten. Das heißt, wir würden diese Studien einschließen und den Schmerz als patientenrelevanten Endpunkt betrachten. Wir stellen infrage, dass eine Änderung der Krankheitsaktivität um 20 % patientenrelevant ist. Wir stellen nicht infrage, dass die Änderung des Schmerzes patientenrelevant ist, bzw. wir halten den ACR 20 nicht für ein geeignetes Kriterium, um das abzubilden.

Was die Gesamtkrankheitsaktivität angeht, halten wir die Remission für relevant, weil es in der Literatur Untersuchungen gibt, die uns zeigen, dass der Patient langfristig davon profitiert. Das haben wir für den ACR 20 nicht. Ungeachtet dessen würden wir uns natürlich die Änderung des Schmerzes, die unmittelbar patientenrelevant und Teilaspekt des ACR 20 ist, anschauen. Um Ihre Frage zu beantworten: Eine Studie, die keine Remission, aber einen ACR 20 dargestellt hat, hat innerhalb des ACR 20 die Änderung des Schmerzes und wir würden sie in dieser Form betrachten.

Jörg Ruof: Trifft die Einschränkung, die Sie gemacht haben, nur auf die Low-Hurdle Endpoints zu, also ACR 20, oder auch auf 50 und 70? Wo ist da Ihr Cut-off?

Die nächste Frage: DAS – wie sehen Sie das? Prof. Krüger hat schon ausgeführt, dass der DAS das wichtigere Kriterium ist. Wie beurteilen Sie das?

Beate Wieseler: Nach meiner Kenntnis der Literatur ist auch für den ACR 50 und 70 nicht gezeigt, dass wir abschließend einen Nutzen für den Patienten erreichen. Es ist richtig, dass wir in gewissen Fällen Überschneidungen in den Bereichen ACR 70 / Remission haben, was die verbleibende Krankheitsaktivität angeht. Das ist sicherlich eine Frage, mit der wir uns noch beschäftigen müssen. Grundsätzlich ist unsere Überlegung aber, wie gesagt, dass ausschließlich für die Remissionen, also für eine relativ niedrige absolute Krankheitsaktivität, gezeigt ist, dass der Patient langfristig davon profitiert. Die Teilminderungen können wir über die anderen Parameter abbilden.

Was die Definitionen von Remission angeht, stimme ich Herrn Leverkus zu: Wir sind in einer Situation, in der wir nicht prospektiv planen können, welche Daten erhoben werden. Wir würden uns die vorhandenen Remissionsdefinitionen anschauen und in ihrer Validität bewerten.

Moderator Jürgen Windeler: Kurzes Nachdenken. – Ja, bitte.

Stefan Simianer: Im Laufe des Gesprächs sind mehrere Punkte aufgekommen. Eingangs war ein Grund, die Remission zu nehmen: Man geht davon aus, dass die Teilverbesserungen quasi proportional zu dem Remissionsendpunkt sind und dadurch kein Informationsverlust entsteht. Das ist aber meines Wissens nicht erwiesen. Es ist nicht automatisch so, dass eine Therapie, die 20 % Remission erzeugt, automatisch 40 % Teilremission oder was auch immer hat. Solange das nicht erwiesen ist, müssen wir davon ausgehen, dass wir einen Informationsverlust haben. – Erstens.

Zweitens gibt es Arbeiten, die zumindest schlussfolgern, dass auch ein ACR 50 klinisch bedeutsam ist. Ich weiß nicht, ob das die Übersetzung für „patientenrelevant“ ist. Es gibt Evidenz, dass auch ein ACR-50-Responder eine klinisch bedeutsame Verbesserung mit sich bringt.

Wir haben uns an anderer Stelle Sorgen gemacht, dass wir Informationen verlieren, wenn wir uns einschränken. Ich kann immer noch nicht erkennen, warum man sich an der Stelle mit der Remission auf ein sehr relevantes, aber doch sehr hohes Therapieziel festlegt und den ganzen Weg, auch eine Zeit, die es braucht, um den Patienten in Remission zu bringen, ignoriert. Viele Patienten bleiben mit ihrer Response kurz vor der Remission stehen, manche kommen vielleicht auf 50 % Besserung. Ich mache mir große Sorgen, dass wir uns hier unnötig reduzieren.

Thomas Reimann: Ich habe noch eine praktische Frage. Es gibt durchaus eine ganze Reihe von Studien, in denen ACR-Kriterien, aber nicht notwendigerweise die Schmerzkriterien einzeln ausgewiesen sind. Wie wollen Sie mit diesen Studien umgehen?

Beate Wieseler: Wie gesagt, ich gehe nicht davon aus, dass wir wesentliche Informationen verlieren, weil wir Einzelaspekte der ACR-Response auswerten. Was Ihre Frage angeht – die Ausweisung der Einzelaspekte –, haben wir bei Ihnen allen die vollständigen Berichte zu den Studien angefordert. Nach meiner Erfahrung liegen die Daten in diesen Studienberichten vor. Ich habe keine Sorge, dass wir sie nicht bekommen, wenn wir die Studienberichte haben. Es ist richtig, dass das in Publikationen gegebenenfalls nicht vollständig ausgewiesen ist, aber weil wir auf Sie zugegangen sind, alle Informationen zu erhalten, gehe ich davon aus, dass uns diese Daten zur Verfügung stehen.

Aljoscha Neubauer: Die Diskussion dreht sich im Wesentlichen darum, welchen Level an Restkrankheitsaktivität wir akzeptieren. Wenn das korreliert, verlieren wir darüber natürlich

keine Informationen. Wir müssen aber berücksichtigen, dass die Studien auf eine bestimmte Zielgröße gepowert waren. Nach meiner Kenntnis war es bei vielen ACR 20. Das heißt, wir können, wenn es korreliert – was zu beweisen wäre –, auch ein sehr hartes Kriterium nehmen, müssen dann aber aufpassen, dass wir es statistisch unter Umständen nicht mehr signifikant darstellen können, weil es dafür nicht gepowert war. Das ist aber in der retrospektiven Situation nicht ein Fehler der Studien, sondern eine Definition des anderen Endpunktes, den wir dann nehmen. Dabei besteht schon die Gefahr, dass man am Ende sagt, das Ergebnis sei nicht signifikant, obwohl es in Wirklichkeit gar nicht signifikant sein kann, weil man eine ganz andere Powerzahl bräuchte. Insofern muss das zumindest klar dargestellt oder überprüft sein.

Friedhelm Leverkus: Das möchte ich unterstreichen. Aus dem Versuch, das multidimensionale Geschehen auf eine Größe abzubilden, haben wir einen zusammengesetzten Endpunkt. Es gab zum einen den Konsens, bei dem gesagt wurde: Das ist ein relevanter Endpunkt. Zum anderen werden dadurch eventuelle Schwankungen reduziert und man bekommt im Prinzip eine gemeinsame Beschreibung der ganzen Sache. – Punkt eins.

Punkt 2: Wenn wir die einzelnen Komponenten haben, zum Beispiel eine Schmerzskala oder andere Skalen, dann würden Sie auch Hedges' g anwenden. Das heißt, Sie werden dann sicherlich sehen, dass ein klinisch relevanter Unterschied vorhanden ist, der gezeigt werden muss. Das verschärft die Problematik noch einmal. Das heißt, auf der einen Seite zerrupft man die 6 oder 7 Komponenten oder wie viele man hat, in die Einzelkomponenten. Bei den Einzelkomponenten genügt es auf der anderen Seite nicht, Signifikanz zu zeigen, sondern im Prinzip muss eine klinisch relevante Überlegenheit gezeigt werden.

Beate Wieseler: Zunächst einmal ist es kein kombinierter Endpunkt in dem Sinne, dass ich Events sammle. Es ist ein Endpunkt, der nicht das Powerproblem adressiert, sondern ich muss gleichzeitig mehrere Aspekte erreichen. Es ist nicht ein klassischer Endpunkt, um das Powerproblem zu lösen. Insofern greift das Argument nicht, dass wir es noch weiter zerrupfen.

Es ist richtig, wenn wir auf einen Schmerzendpunkt schauen, würden wir die Frage nach der Relevanz der Effekte mit der auch im Berichtsplan beschriebenen Methodik stellen, die wir im Laufe des letzten Jahres ganz intensiv auf verschiedenen Veranstaltungen diskutiert haben. Mein Eindruck dort war, dass Konsens darüber besteht, dass man die Relevanzfrage stellen muss.

Noch ein Punkt zu der Powerproblematik: Es ist richtig, dass das in den einzelnen Studien gegebenenfalls auf den ACR 20 gepowert ist. Wir sind hier in der Situation einer Meta-Analyse. Wir poolen alle Studien, die vorhanden sind, was das Powerproblem verringert.

Friedhelm Leverkus: Ich glaube nicht, dass der ACR 20 gewählt worden ist, um das Powerproblem zu lösen, sondern man versucht, mit einer Variablen den klinischen Umstand zu beschreiben. Wenn dem so ist, dann spricht einiges dafür, ihn auch in Ihre Analysen einzubeziehen, weil es einmal Konsens gab, dass er etwas beschreibt, das klinisch sinnvoll ist. – Punkt eins.

Punkt 2: Wenn wir es nicht mit stetigen Variablen zu tun haben, die wir zusammenzählen, bei denen wir eine Reduktion der Fehlervarianz haben, besteht auch bei binären Variablen – die auch eine Varianz haben, bei denen auch etwas reduziert wird –, gerade bei einem multidimensionalen Geschehen, bei dem es in verschiedene Richtungen laufen kann, ein Vorteil darin, dass man eher Unterschiede sehen kann.

Punkt 3, noch einmal zurück: Auch wenn es Konsens gibt, dass man darauf schauen muss, so wird es mit dem Verfahren, das Sie anwenden, für den Einzelnen sehr viel schwieriger – das muss man auch sehen –, den Endpunkt Schmerz zu zeigen.

Malte Wolff: Habe ich Sie vorhin richtig verstanden, dass Sie ACR 50 und ACR 70 als nicht patientenrelevant einstufen?

Beate Wieseler: Wir sehen die Krankheitsaktivität an sich als patientenrelevant an. Wir denken aber, dass die Literatur zeigt, dass primär die Remission notwendig ist, damit der Patient einen Nutzen von der Therapie hat. Sie haben uns auch, so wie ich Ihre Stellungnahmen vor Augen habe, keine anderslautende Literatur eingereicht. Wir werden noch einmal prüfen, ob Sie uns dazu Literatur eingereicht haben. Unsere Literaturrecherche hat ergeben, dass wir im Zusammenhang zwischen einer Minderung der Krankheitsaktivität und einem tatsächlich langfristigen Nutzen für den Patienten die Remission sehen.

Jörg Ruoff: Zum einen noch einmal die Frage: Wie definieren Sie Remission? Mit dem DAS zum Beispiel liegt ein schönes Kriterium vor, sie zu definieren. Ich verstehe es so: ACR 70 ist für Sie keine Remission. Also ist die Frage: Wie definieren Sie die Remission?

Zum anderen habe ich Sie so verstanden, dass alles, was nicht Remission ist, aus Ihrer Sicht nicht patientenrelevant ist. Ich denke, das Anliegen von Frau Faubel am Anfang war, ganz klar zu sagen: Dank der Biologika ist Remission in erreichbare Nähe gekommen, aber das ist nicht für alle Patienten realistisch und kann daher auch von großem Nutzen sein, wenn man das Ziel nicht erreicht.

Klaus Krüger: Ich muss noch einmal ganz dringend daran appellieren, dass der Punkt Remission – ein DAS28 von 2,6 – artifiziell ist. Er ist einfach mal so festgelegt worden und natürlich ein nützliches Instrument, aber der Nutzen des Patienten ist linear. Das heißt, mit dem besser werdenden DAS, also auch mit der Low Disease Activity, hat man einen wesentlich besseren Outcome als mit der High Disease Activity, also 5,1. Nicht nur der blanke Punkt Remission als solcher, sondern schon das Ausmaß des Erfolges oder der DAS-

Reduzierung, der DAS-Besserung muss eine wichtige Rolle spielen, weil die Remission für viele, gerade langjährige Patienten eine Illusion ist, die maximal nur noch 5 oder 10 % erreichen.

Hans-Joachim Kreutzenbeck: Ich möchte auch noch einmal betonen, dass der Patient durchaus einen Nutzen hat, wenn man zum Beispiel eine niedrige Krankheitsaktivität erreicht. Das schlägt sich darin nieder, dass die Funktionsschädigung, das heißt die Schädigung der Gelenke, verlangsamt wird und dadurch hinterher die Schwächung der Gelenke, der Gelenkstatus oder wie der Patient die Gelenke gebrauchen kann, wie stark die Behinderung ist, deutlich vermindert wird. Insofern ist das für den Patienten sehr relevant, auch wenn er nicht die Remission erreicht, sondern einen niedrigen Aktivitätsstatus.

Beate Wieseler: Ich habe schon mehrfach darauf hingewiesen, dass wir nicht nur die Remission, sondern auch die Symptomatik, die gesundheitsbezogene Lebensqualität, den körperlichen Funktionsstatus – gemessen über den HAQ zum Beispiel – als Endpunkte betrachten und in den Endpunkten natürlich die ganzen Teilminderungen abgreifen und abbilden.

Ursula Faubel: Dazu habe ich direkt eine Nachfrage. Werden die Endpunkte in irgendeiner Form gewichtet oder stehen sie einfach so nebeneinander und werden alle gleichberechtigt untersucht und bewertet?

Beate Wieseler: Diese Endpunkte werden zunächst einmal gleichberechtigt nebeneinandergestellt.

Christof Iking-Konert: Aus Interesse möchte ich noch einmal nachfragen – Herr Ruof hat es schon gefragt, und Sie sind quasi nicht darauf eingegangen: Wann legen Sie fest, wie Sie Remission definieren? Wir haben jetzt 20 Jahre in der Rheumatologie diskutiert. Im März erscheinen die neuen Remissionskriterien. Wann legen Sie sich fest, oder ist es vermessen, das hier zu fragen oder eine Antwort zu erwarten?

(Heiterkeit)

Beate Wieseler: Wie Sie gesehen haben, haben wir nicht in den Berichtsplan geschrieben: Remission ist der DAS28 bei diesem oder jenem Wert. Wir werden uns die vorhandenen Daten anschauen und nach der Begründung für die Kriterien fragen.

Es ist richtig, es gibt das neue Papier von ACR, EULAR und OMERACT, in dem neue Remissionsdefinitionen, auch gut begründet, festgelegt wurden. Wir werden natürlich nicht sagen: Das ist jetzt der Stand der Diskussion und das ist für uns Remission, alle anderen Daten können wir nicht mehr berücksichtigen. Das macht keinen Sinn. Wir werden auf Basis der vorhandenen Daten prüfen, welche der Definitionen begründet und valide sind.

Christof Iking-Konert: Zu den Remissionskriterien möchte ich noch anmerken, dass gerade 2 Dinge, die Sie in Ihren Berichtsplan aufgenommen haben, nämlich Schmerz und Müdigkeit, explizit ausgenommen wurden, weil sie schwierig mit Evidenz zu belegen sind. Schmerz ist aus dem Score herausgenommen und die Morgensteifigkeit ist diskutiert worden. Auch die Fatigue ist diskutiert und explizit ausgenommen worden. Wie sehen Sie das jetzt unter der neuen Diskussion, weil Sie es explizit als relevanten Punkt aufgenommen haben?

Beate Wieseler: Das waren Rückmeldungen aus der Konsultation der Patientenvertreter, dass dies relevante Einschränkungen sind, die in der Erkrankung von den Patienten erfahren werden. Darüber hinaus gibt es auch dazu Literatur aus der OMERACT-Gruppe, dass gerade diese Aspekte für Patienten relevant sind.

Hans-Theo Forst: Frau Wieseler, wenn Sie jetzt auf andere Primärvariablen, Endpunkte übergehen, also die patientenrelevanten wie Schmerz usw., brauchen Sie eigentlich neue Relevanz- und auch Irrelevanzschwellen, damit Sie bei der Hierarchie des IQWiG etwas tun können. Wie sehen Sie das? Setzen Sie die Relevanzschwellen selber fest? Mit $d = 0,2$ kann man nach Schema F verfahren. Das wäre Step five in Ihrer Hierarchie. Das ist natürlich recht hoch gegriffen – das muss man immer wieder sagen –, weil es dann auch für Irrelevanzgrenzen gilt.

Beate Wieseler: Wir entfernen uns jetzt mehr und mehr von unserem Endpunkt.

(Hans-Theo Forst: Das hat nur Konsequenzen! Darauf wollte ich aufmerksam machen!)

Wir haben eine allgemeine Fragestunde zum Berichtsplan. Die Methodik für die Relevanzbewertung ist im Berichtsplan beschrieben. Wir würden uns freuen, wenn wir Responderanalysen auf Basis von begründeten MIDs finden. Wenn nicht, werden wir die im Berichtsplan beschriebene Relevanzgrenze ansetzen müssen, um überhaupt eine Aussage zu diesem Endpunkt machen zu können.

Peter-Andreas Löschmann: Ich möchte noch etwas zu dem Endpunkt Schmerz sagen. Es ist entscheidend wichtig, dass man sich die einzelnen Studienprotokolle selbst anguckt; denn ein Parameter – das finde ich immer sehr schade –, der nicht sehr gut erhoben wird, ist der Schmerzmittelverbrauch. Er geht dann direkt mit in Ihren Endpunkt ein. Man muss sehen, was die Protokolle vorgesehen haben. Ich habe versucht, mir einen Überblick zu verschaffen: Es gibt alles von „immer konstant halten“ bis „möglichst wenig Schmerzmittel nehmen“. Wenn Sie das explizit analysieren wollen, dann müssen Sie wissen, unter welchen Bedingungen die Schmerzskalen erhoben wurden.

Beate Wieseler: Das ist richtig. Das ist sicherlich eine Frage, die man adressieren müsste. Deshalb wäre es für uns auch wichtig, dass wir alle Protokolle vollständig von Ihnen bekommen.

Friedhelm Leverkus: Wo liegt denn Ihrer Meinung nach das Problem beim ACR 50 und ACR 70? Liegt es daran, dass Laborparameter darin sind, oder hat das keine Rolle gespielt?

Beate Wieseler: Das ist nicht das Problem. Bei der Beschäftigung mit der Frage der Krankheitsaktivität und auch mit Ihren Stellungnahmen haben wir das Ganze noch einmal dahin gehend weiterentwickelt, dass wir sagen: Das ist eigentlich kein klassischer kombinierter Endpunkt, sondern die Remission ist ein Gesamtkonzept – so hat man es eben auch bezeichnet – und die Remissionskriterien sind eine Art Messinstrument für das Gesamtkonzept. Wir sehen das Problem nicht primär darin, dass auch Laborwerte erfasst werden.

Karl J. Krobot: Sollte nicht aus methodischen Gründen bereits auf der Ebene des Berichtsplans darin stehen, wie Sie Remission im Laufe des Verfahrens operationalisieren wollen? Wir hören jetzt, dass das in der Zukunft erfolgen wird. Ich hätte aber fast erwartet, dass das bereits Gegenstand des Berichtsplans ist, darin steht und heute auch erörtert werden kann.

Beate Wieseler: Weil wir nicht in der Situation sind, die Datenerhebung prospektiv planen zu können, ist das aus unserer Sicht nicht möglich. Wir würden die Studien identifizieren, prüfen, ob die Remissionsdefinitionen, so wie sie in den Studien vorliegen, begründet sind, und dann damit arbeiten. Um alle sinnvollen Definitionen im Berichtsplan einschließen zu können, müssten wir zunächst einmal die gesamte Literatur sichten. Das ist ein Problem, das wir grundsätzlich bei der Erstellung eines systematischen Reviews, einer retrospektiven Analyse haben. Wir können diese Punkte nicht festlegen, sondern wir legen die Endpunkte fest und sichten dann die Literatur, um die Operationalisierung zu prüfen.

Bernd Richter: Methodisch wäre es auch ein falsches Vorgehen, so etwas von vornherein zu definieren, weil jede Definition angreifbar wäre, egal von welchem Remissionskriterium wir uns etwas aussuchen würden. Da es ein iterativer Prozess ist, müssen wir erst einmal jede Evidenz berücksichtigen. Wenn wir uns auf bestimmte Evidenz beschränken würden, würde man sich sofort des Informationsverlustes schuldig machen.

Zum Zweiten möchte ich noch etwas zum Begriff „ACR“ sagen: Hätten wir hier vor 10 Jahren zusammengesessen – wir müssen immer gucken, wie wir das historisch entwickelt haben –, dann wären wir alle mit ACR 20 wahrscheinlich vollkommen zufrieden gewesen, weil das therapeutische Armamentarium so gering war, dass das schon ein erheblicher Fortschritt war. Heutzutage würde sich niemand ernsthaft noch mit einer ACR 20 zufriedengeben. Das Problem ist: Wir wissen, dass wir stärker erkrankte Patienten nicht mit Macht in eine Remission reindrücken dürfen, zum Teil auch, weil dann die unerwünschten Wirkungen im Verhältnis zu dem Krankheitsbild für den Patienten eventuell sehr ungünstig wären, sodass das Konstrukt der minimalen Erkrankungsaktivität schon etwas an sich hat. Das Problem ist nur, zu operationalisieren, was das genau ist, wie wir es definieren wollen und wie wir die Assoziation zwischen einer minimalen Erkrankungsaktivität und Langzeitfolgen

sehen. Dazu sind die Daten, so wie wir es bisher gesehen haben, ziemlich schlecht. Wenn Sie Ideen einer Operationalisierung und entsprechende Daten haben, dann wäre das sehr interessant.

Moderator Jürgen Windeler: Ich habe den Eindruck, dass wir das Thema jetzt ausreichend diskutiert und auch ausreichende Informationen aufgenommen haben. – Herr Leverkus, ich bitte um Nachsicht.

Jetzt kommen wir zu:

4.3.4 Tagesordnungspunkt 2: Endpunkte, b) Radiologischer Verlauf

Katharina Biester: In den Stellungnahmen wird vorgeschlagen, die radiologische Progression als Surrogatparameter für die strukturellen Gelenkveränderungen zu berücksichtigen. Unsere erste Frage an Sie ist: Welche zusätzliche Information würde dadurch gewonnen werden?

Die zweite Frage: Ist die radiologische Progression ein valides Surrogat?

Moderator Jürgen Windeler: Herr Krüger, bitte.

Klaus Krüger: Zunächst einmal muss man sagen: Es gelingt nicht, die ursprünglich aufgeführten Parameter zu messen. Bei „Kontraktur“ haben Sie 3 Parameter aufgeführt. Dafür gibt es keine validierten Messparameter, sodass die unmittelbaren strukturellen Folgen – Kontraktur ist ja schon eine Folge – eigentlich nicht in Studien erfasst sind. – Das ist das Erste.

Zum Zweiten sind das alles Dinge, die wir heute eigentlich verhindern wollen; sie sollten zumindest bei Patienten, die neu behandelt werden, überhaupt nicht mehr auftreten. So bleibt aus meiner Sicht – ich habe das auch in dem Antrag so formuliert – nur der Parameter „Radiologische Veränderung“, der über die Gelenkspaltverschmälerung und die Erosionen einen Hinweis auf strukturelle Veränderungen gibt. Messtechnisch sehe ich keinen anderen.

Stefan Simianer: Wir haben in unserer Stellungnahme noch einmal mit Literatur unterlegt, dass die radiologische Progression langfristig sehr deutlich mit dem Funktionsstatus korreliert. Die ersten Arbeiten in der TEMPO-Studie zeigen, dass es innerhalb von 2 Jahren eine Korrelation zum Funktionsstatus gibt, sodass ich denke, dass Ihnen die Evidenz vorliegt.

Der zweite Punkt ist: Wenn Sie einen Patienten zu einem Zeitpunkt X untersuchen, den Funktionsstatus machen und Ihre Parameter für Remission oder ACR erheben, dann werden Sie immer wieder auf Patienten treffen, die an Stellen, wo sie keine klinischen Beschwerden haben, radiologische Veränderungen zeigen. Diese Diskrepanz allein ist im klinischen Alltag immer Evidenz dafür, dass die Röntgenuntersuchung eine zusätzliche Informationsquelle liefert, die Sie mit dem klassischen Parameter nicht abgreifen.

Moderator Jürgen Windeler: Sie haben gerade die Korrelation zwischen den röntgenologischen Veränderungen und Funktionsänderungen beschrieben. Die Korrelation innerhalb eines Patienten ist also für Sie die Validität des Surrogatparameters?

Stefan Simianer: Die Literatur, die ich kenne, hat Studienkohorten über viele Jahre verfolgt; die Arbeiten kommen vor allem aus den Niederlanden. Die Korrelation zum Funktionsstatus nimmt mit der Dauer der Beobachtung stetig zu.

Moderator Jürgen Windeler: Noch einmal die Frage: Das ist dann für Sie ein valider Surrogatparameter?

Stefan Simianer: Das ist die beste Evidenz, die ich gefunden habe, ja.

Moderator Jürgen Windeler: Okay, gut. – Weitere Fragen dazu?

Hans-Joachim Kreutzenbeck: Ich halte es für wichtig, die röntgenologische Progression zu betrachten, weil sie schließlich, wenn sie weiter voranschreitet, zu einer Funktionseinschränkung des Patienten führt und damit für den Patienten relevant wird. Man bekommt nicht nur durch die Betrachtung der Remission heraus – das wurde schon erwähnt –, ob in diesem Bereich etwas passiert. Das ist auch der Grund dafür, dass dieser Parameter in der täglichen Praxis verfolgt wird; denn man kann nicht allein anhand der Krankheitsaktivität sehen, ob die Funktionseinschränkung passiert. Deshalb – das ist ganz klar – werden die Patienten ca. jährlich oder teilweise, je nach Krankheitsaktivität, auch etwas häufiger röntgenologisch untersucht, um das festzustellen und dann zu vermeiden, dass es zu der Funktionseinschränkung und damit am Ende zu einer Behinderung des Patienten kommt.

Oliver Sander: Bei der Messung des Ganzen stellt sich natürlich die Frage, ob es auch Grenzwerte wie einen ACR 20 der Röntgendestruktion gibt oder ob Konsensus darüber besteht, dass null radiologische Progression ein evidenter Nutzen für den Patienten ist, aber zum Beispiel die Verhinderung eines Anstiegs des Sharp-Scores um 2 Punkte innerhalb von einem Jahr, was dann ungefähr 1 % des maximalen Scores ist, schon ein relevanter Nutzen für den Patienten, der zwar messbar ist, sich aber letztendlich vielleicht doch nicht auswirkt.

Malte Wolff: Es gibt auch im Bereich der radiologischen Progression Studien, insbesondere von OMERACT initiiert, die sich mit den MCID-Diskussionen auseinandersetzen. Beim mTSS nimmt man meines Wissens einen Bereich von 5 an, es gibt auch andere. Beim TSS sagt man: ungefähr 4 auf ein halbes Jahr oder auf ein Jahr; da müsste ich nachgucken. Es gibt solche validierten oder etablierten Grenzwerte. Sie sind nicht immer auf die Kommastelle exakt, aber die Bereiche sind in der Literatur schon, insbesondere über die OMERACT-Initiative, relativ gut verankert.

Bernd Richter: Meiner Meinung nach sprechen Sie mehr von „minimal detectable changes“ als von „important“. Bei „important“ muss man erst einmal genau definieren, was das ist. Das Problem da ist wiederum, dass es in der Rheumatologie sehr umstritten ist, ob die in den vorhandenen Kurzzeitstudien, die wir über ein Jahr oder über 2 Jahre haben – keine reinen Assoziationsstudien –, beobachteten Progressionsverzögerungen eine Relevanz bei längerfristigen Auswirkungen für die Patienten hinsichtlich Funktionseinschränkungen etc. haben. Je tiefer ich hineinschaue und je genauer ich eine kleine Differenzbildung herausarbeite, im Gegensatz vielleicht zu einer Nullprogression – null ist null, egal wie ich sie messe, natürlich mit einer gewissen Grauzone um null herum –, desto mehr muss ich fragen,

welche Auswirkungen solche kurzzeitigen Effekte langfristig haben. Das ist ein generelles Problem.

Malte Wolff: Es gibt diverse Studien, die mittlerweile zum Beispiel den HAQ in langfristige und irreversible Bestandteile unterteilen. Es ist nun einmal ein Phänomen der Erkrankung der rheumatoiden Arthritis, dass sich diese irreversiblen Bestandteile über die Zeit hinweg verstärken und aufaddieren. Das heißt, würde man nur die Minderung der Krankheitsaktivität nehmen, würde ein Teil des Wissens verloren gehen. Wie bereits vorher gesagt: Es manifestiert sich teilweise erst nach 5, 6, 7 oder auch nach 10 Jahren. Dementsprechend muss man solche Surrogate verwenden. Denn eine RCT über 10 Jahre ist schlichtweg unmöglich; da dürften wir alle übereinstimmen. Radiologische Progression misst dies eben am besten und sollte deswegen auch verwendet werden.

Hans-Joachim Kreutzenbeck: Ich möchte das noch einmal verdeutlichen. Am Anfang der Erkrankung, wenn Sie die Funktionskapazität des Patienten betrachten, wie sie sie zum Beispiel mit einem HAQ bestimmen können, können Sie viel erreichen, indem Sie Krankheitsaktivität wegnehmen. Wenn die Erkrankung voranschreitet, dann ist die Krankheitsaktivität nur noch ein Teil der Funktionseinschränkung. Der andere Teil ist die Schädigung im Bereich des Knochens, im Bereich der Gelenke. Wenn die Krankheit vorangeschritten ist, können Sie nur noch einen kleinen Teil in Bezug auf die Aktivität verbessern, weil der andere Teil der Funktionseinschränkung über die Schädigung passiert. Insofern ist es wichtig, dass man die Schädigung beseitigt und darauf achtet, es auch misst, dass das nicht passiert.

Moderator Jürgen Windeler: Gut. – Gibt es unsererseits noch Klärungsbedarf? –Nein. Dann schließe ich TOP 2 ab und komme zu:

4.3.5 Tagesordnungspunkt 3: Effektmodifikatoren

Frau Biester.

Katharina Biester: Im vorläufigen Berichtsplan wurden die folgenden 3 Effektmodifikatoren vorab festgelegt: Alter, Geschlecht und die Art der Vorbehandlung. In den Stellungnahmen wurde eine Reihe weiterer potenzieller Effektmodifikatoren genannt, die in den Berichtsplan aufzunehmen wären, darunter auch Krankheitsschwere und Krankheitsdauer. Unsere Frage an Sie ist, wie diese beiden Effektmodifikatoren zu operationalisieren wären.

Moderator Jürgen Windeler: Sollen wir zu TOP 4 gehen?

(Heiterkeit – Zuruf: Irgendjemand wird es ja geschrieben haben!)

Katharina Biester: Ich könnte sie nennen.

Klaus Krüger: Ich zum Beispiel. Es hat sich gezeigt, dass sich die Krankheitsdauer auf die Erfolgsaussicht der Therapie auswirkt. Das haben viele Subgruppenanalysen in diversen Studien ergeben. Das Gleiche gilt im Prinzip für die Ausgangsaktivität. Es ist ganz klar, dass die Erfolgsaussichten bei einer Studie, bei der zu Beginn ein durchschnittlicher DAS von 7 vorliegt, im Vergleich zu einer Studie, wo er bei 4,9 liegt, wie wir es zum Beispiel bei einigen Strategiestudien haben, dann auch unterschiedlich sind. Ich könnte Ihnen aber nicht über die Subgruppenanalysen hinaus sagen, wie man das sonst operationalisieren sollte. Es ist ein Phänomen, das man einfach beobachtet.

Malte Wolff: Fast jede Studie hat in der Baseline stehen, wie lange die Patienten krank sind. Bei einem steht vielleicht „ab Symptombeginn“ – das ist meist bei den frühen Studien der Fall –, bei einem anderen steht „Krankheitsdauer“. Jemand, der 10 Jahre krank ist, kann sich sicherlich um ein Jahr vertun, wenn er gefragt wird, trotzdem denke ich, dass die Kategorisierung in etwa möglich ist.

Wenn Sie sich Meta-Regressionen angucken, von denen wir nicht sicher wissen, ob sie stattfinden oder nicht, dann ist das üblicherweise ein Parameter, der auch kontrolliert wird. Ein Beispiel, wie man Schweregrad nicht abbilden kann, wurde von Herrn Krüger gerade genannt: die Krankheitsaktivität zu Beginn, Erosionen, die bereits vorliegen, oder Ähnliches.

Oliver Sander: Zum Schweregrad gibt es auch Einschlusskriterien für die Studien. In die üblichen Studien mit Biologika wurden im Allgemeinen nur Patienten, die schwer erkrankt sind, eingeschlossen, allein um als Effektmediator einen Rückgang der Krankheitsaktivität messen zu können. Natürlich kann man noch einmal gucken, ob es Unterschiede gibt. In die ersten Studien sind kränkere Patienten eingeschlossen worden als in die neueren Studien.

Die Krankheitsdauer ist sicherlich ein wichtiger Faktor, aber es gibt auch Menschen, die 20 Jahre eine schleichende Erkrankung haben, bei denen man sagt: Jetzt kennen wir das Mittel

besser, jetzt stellen wir Sie mal auf dieses Mittel um. Andere haben ein halbes Jahr Krankheitsverlauf, schon viele Erosionen und eine ungünstigere Prognose als der, der nach 20 Jahren gerade erst infrage käme. Es gibt auch immer das Gegenbeispiel dafür.

Noch wichtiger finde ich das Manifestationsjahr der Erkrankung und die „Zeitdauer“ bis zur ersten effektiven Therapie. Wir wissen: Je früher und je effektiver wir behandeln, desto günstiger ist der gesamte Langzeitverlauf, desto weniger Progress ist dadurch allein biologisch determiniert. Das heißt, heutzutage wissen wir, dass Menschen, die mit ihrer Erkrankung in den letzten 5 Jahren manifest geworden sind, schon früh effektiv und aggressiv behandelt wurden, was vor 20 Jahren noch nicht der Fall war. Damals hat man erst einmal 5 Jahre geguckt und schon viel verändert, was dann nicht mehr ansprechen kann, sodass sicherlich nicht nur die Krankheitsdauer, sondern auch der Zeitpunkt der Studie und letztendlich das Jahrzehnt, in dem die Untersuchung durchgeführt wurde – wahrscheinlich sogar mehr als die Krankheitsdauer allein –, ein Effektmediator sind.

Michael May: Von den Vorrednern wurde schon viel zu dem gesagt, was gemeint ist. Die meisten Redner und auch wir meinen, dass letztendlich ein Unterschied da sein muss, ob ein Patient schon 12 Jahre oder erst 6 Jahre unter Therapie steht und wie der Ausgangs-DAS bei der Untersuchung ist, ob 6, 4 oder wie auch immer. Das ist ein wesentlicher Punkt, der bei der heterogenen Population, die wir haben, Berücksichtigung finden sollte.

Friedhelm Leverkus: Ich möchte darauf hinweisen, dass es 2 verschiedene Anwendungen dieser Effektmodifikation gibt. Zum einen kann ich versuchen, wenn ich Heterogenität sehe, diese zu erklären. Zum anderen kann man versuchen nachzuweisen und tatsächlich in Subgruppen hineinzugehen und nachzuschauen, ob es in allen Subgruppen gleich wirkt. Wie ist Ihr Ansatz?

Beate Wieseler: Wir schauen an verschiedenen Stellen auf genau diese Fragen. Wenn wir unseren Studienpool haben und uns die Baselinecharakteristika der Patienten und die Studiencharakteristika anschauen, dann sind das alles Überlegungen, die dazu beitragen, die Studienpools zu bilden: Welche Studien kann man tatsächlich sinnvoll zusammenfassen? Dabei werden solche Überlegungen eine Rolle spielen.

Den zweiten Punkt haben Sie gerade auch genannt: Wenn wir eine Meta-Analyse machen und eine Heterogenität sehen, dann sind das natürlich potenzielle Effektmodifikatoren, die wir uns daraufhin anschauen würden, ob sie gegebenenfalls diese Heterogenität erklären.

Wenn wir Subgruppenanalysen machen wollen, dann haben wir genau das Problem, dass wir sie dann operationalisieren und definieren müssen. Deshalb unsere Frage: Was sind sinnvolle Operationalisierungen? Wenn ich mir Ihre Antworten vor Augen führe, dann ist die Situation hinreichend komplex, als dass es keine einfachen Operationalisierungen gäbe. Es macht keinen Sinn, zu sagen: Ich habe die Gruppe der Patienten mit einer Krankheitsdauer größer oder kleiner 5 Jahre, weil es dabei viele überlappende Aspekte gibt. Das war unsere Frage, die

wir an Sie hatten: Gibt es auf der Hand liegende Operationalisierungen, um so etwas für Subgruppen zu definieren? Mein Eindruck ist jetzt, dass es das nicht gibt.

Oliver Sander: Für die Krankheitsdauer mit relevanter Krankheitsaktivität, die sicherlich der wichtigste Effektmediator ist, ist die Zahl der Vortherapien mit Basistherapeutika zum Beispiel ein guter Marker. Es ist vielleicht nicht der alleinige, aber wenn jemand vorher schon zehn Therapien hatte, dann hat er wahrscheinlich lange eine hohe Krankheitsaktivität gehabt oder kurz eine sehr hohe Krankheitsaktivität, sodass der Rheumatologe als Behandler vieles versucht hat, um das zu unterdrücken. Derjenige, der vorher nur eine Therapie hatte – ob er nun ein halbes Jahr oder 20 Jahre krank ist –, hat wahrscheinlich eine günstigere Prognose und ein besseres Ansprechen. Ich denke, dass die Zahl der Vortherapien mit gewissen Einschränkungen, aber zumindest gleichzeitig eine gewisse Operationalisierung der Krankheitsdauer und Krankheitsaktivität ist.

Stefan Simianer: Ich möchte noch ergänzen, dass auch die Begleittherapie sicherlich ein relevanter Effektmodifikator ist, als da einmal die begleitende Methotrexatdosis wäre. Ob sie ausreichend ist, dazu wäre vielleicht ein Subgruppen- oder Schwellenwertansatz sinnvoll. Dann haben wir die Steroidbegleittherapie, die definitiv relevant ist, und wir hatten vorhin schon den Aspekt Schmerzmittel, gerade wenn Sie Schmerz als Endpunkt bevorzugen.

Ein weiterer damit verbundener Aspekt wäre das Verhältnis des Messzeitpunkts der Krankheitsaktivität zum Dosierungsintervall der Substanz, ein Problem, das ich Ihnen allerdings operationalisiert nicht beantworten kann, das aber sicherlich von großer Wichtigkeit bei der Beurteilung der Wirksamkeit einer Substanz ist.

Jörg Ruof: Ergänzend zu Herrn Simianer: Die Begleittherapie unterstützen wir absolut. Es gibt auch einige Studien, bei denen gar keine Begleittherapie, also kein DMARD ergänzend dabei ist. Biologische Monotherapien sollten also getrennt betrachtet werden.

Hans-Joachim Kreutzenbeck: Im Grunde ist das nur noch einmal ein Hinweis darauf. Wenn Sie sich die Einschlusskriterien der Studien angucken, dann sehen Sie, dass sie aufgrund der Einschlusskriterien von vornherein ganz unterschiedliche Patienten aussuchen. Daran erkennen Sie schon, welche Studien versuchen, ähnliche Patientengruppen zu untersuchen, und welche nicht.

Beate Wieseler: Da kann ich Sie beruhigen. Es gehört zu unserem normalen methodischen Repertoire, dass wir uns die Studien anschauen und prüfen, welche sinnvollerweise zu poolen sind. Im zweiten Schritt gehen wir dann bei auftretender Heterogenität noch tiefer hinein.

Moderator Jürgen Windeler: Okay. – Dann möchte ich diesen Punkt auch abschließen, und wir kommen zu:

4.3.6 Tagesordnungspunkt 4: Verschiedenes

Hierzu gab es eine Reihe von Vorschlägen und Wortmeldungen. Ich darf diejenigen, die sich am Anfang mit Themen zu Wort gemeldet hatten, bitten, sich jetzt erneut damit zu Wort zu melden. – Herr Leverkus, bitte.

Friedhelm Leverkus: In dem Berichtsplan ist noch einiges unklar und offen. Wie Herr Richter gesagt hat, ist ein iterativer Prozess notwendig, um letztlich zu dem Weg zu kommen. Daher meine Frage: Macht es in solch einer Situation nicht Sinn, dass man auch diejenigen, die hier Stellung genommen haben, in den iterativen Prozess einbezieht? Bei den Berichtsplänen, wie wir sie sonst kennen, war schon sehr viel klarer, und man konnte dazu Stellung nehmen. Wenn die Arbeit fertig ist, wenn der Draftbericht vorliegt, dann ist es schwieriger, noch zu einer Diskussion zu kommen.

Beate Wieseler: Im Vergleich zu anderen Berichtsplänen habe ich nicht den Eindruck, dass das so viel unklarer ist. Die Unschärfe, die in diesem Berichtsplan liegt, liegt in der Natur der Sache, weil wir hier ein systematisches Review machen. Deshalb kann ich nur darauf verweisen und halte es auch für ein adäquates Verfahren, dass wir weiter gehende Punkte, die dann auf den vorliegenden Daten basieren, im Stellungnahmeverfahren zum Vorbericht diskutieren.

Moderator Jürgen Windeler: Anders ausgedrückt: Es ist kein abweichendes Verfahren von dem üblichen IQWiG-Verfahren vorgesehen. – Herr Krüger.

Klaus Krüger: Das schließt also auch ein – ich weiß, dass das eine brisante Diskussion ist –, dass man komplett auf die im Falle der Biologika vorliegenden qualitativ hochwertigen Register verzichtet. Wir haben eine sehr interessante Untersuchung von Frau Zink – im letzten Jahr publiziert – gesehen: Wenn man sich das deutsche RABBIT-Register anschaut, dann würden danach nur 30 % der mit Biologika behandelten Patienten die Einschlusskriterien für kontrollierte Studien erfüllen. Das heißt, 70 % der Patienten, die in Deutschland tatsächlich mit Biologika behandelt werden, würden überhaupt nicht in randomisiert-kontrollierte Studien aufgenommen. – Das ist der eine Aspekt.

Der zweite Aspekt sind Dinge, die meines Erachtens sehr relevant für den Patienten sind, die man erst im Langzeitverlauf erfassen kann, zum Beispiel Langzeitsicherheit. Wir haben mindestens 3 sehr große Register – das deutsche, das skandinavische und das britische, es gibt natürlich noch mehr –, die 10 Jahre lang Daten zur Langzeitsicherheit der Biologika im „Real Life“ sammeln. Das halte ich für einen außerordentlich patientenrelevanten Aspekt.

Volker Vervölgyi: Im Berichtsplan steht, dass wir uns vornehmlich auf RCTs beschränken. Unsere Meinung im Moment ist: Die Aussagen, die aus Registern gezogen werden, haben im Vergleich zu Ergebnissen aus RCTs eine hohe Ergebnisunsicherheit. Die Effekte aus den Registern sind meistens nicht sehr groß, wie wir in den Publikationen und Stellungnahmen

gesichtet haben. Wenn man da die hohen Ergebniswerte danebenhält, dann ist es schwierig, noch kausale Zusammenhänge zu sehen.

Klaus Krüger: Ich gehe auch nicht davon aus, dass man die Register zur Effektivität heranziehen sollte. Dafür sind sie nicht primär konzipiert worden. Ich gehe von der Langzeitsicherheit der Therapien aus, die man meines Erachtens in randomisiert-kontrollierten Studien nicht ausreichend erfassen kann. Der Nebenaspekt ist: Es werden einfach andere Patienten behandelt.

Volker Vervölgyi: Das ist richtig. Nur, auch beim Schaden gucken wir in der Nutzenbewertung auf kausal begründete Zusammenhänge. Es geht nicht darum, seltene Nebenwirkungen – im Sinne einer Signaldetektion – aufzudecken. Daher sind wir auch da auf Evidenz von hoher Ergebnissicherheit angewiesen.

Peter-Andreas Löschmann: Ich möchte gerne Herrn Krüger beipflichten. Gerade wenn es um Schaden und Risiken geht, wissen wir aus den Registern und nicht aus kontrollierten Studien, dass beispielsweise die Infekthäufigkeit gerade zu Behandlungsbeginn größer ist als bei der längerfristigen Betrachtung. Wir können jetzt lange darüber diskutieren, woran das liegt und ob es ein tatsächlicher Effekt oder durch die Systematik bedingt ist. Ich halte es für einen tatsächlichen Effekt. Das müsste man, wenn man die Produktgruppe insgesamt und die Risiken für die Patienten, die damit behandelt werden, betrachtet, auch berücksichtigen. Sie vergeben eine ganze Menge an Möglichkeiten.

Ich verstehe Ihre Argumente, aber im Vergleich zu anderen Diskussionen, die wir am gleichen Ort geführt haben, ist es eine neue Qualität, dass wir tatsächlich für eine gesamte Gruppe von Medikamenten die Patienten prospektiv in dem Register haben. Das ist dann eine andere Qualität als ein irgendwann begonnenes Register in einem völlig heterogenen therapeutischen Umfeld. Ich würde darum bitten, noch einmal Gedanken darauf zu verwenden, ob man die Daten nicht intelligent und sinnvoll nutzen kann.

Oliver Sander: Zu dem größten Schadensereignis, das eintreten kann, nämlich der Todesfolge, habe ich mir die Mühe gemacht, alle Register anzuschauen, und hatte Schwierigkeiten mit dem Komparator, also mit den Kontrollgruppen, die die Register führen, weil sie natürlich nicht randomisiert sind, sondern einen Selektionsbias haben. Aus dem britischen Register – jetzt publiziert – gibt es die mühsame Arbeit, Vergleichsalgorithmen zu finden, die dann auch wieder, je nachdem welchen Vergleichsalgorithmus man findet, zu unterschiedlichen Ergebnissen führen.

Ich denke, das größte Problem einer Nutzenanalyse im Sinne einer Verhinderung von Schaden ist der Vergleich zu dem, was zu erwarten wäre. Die Register sind ein guter Anhalt, aber man kann sicherlich nicht in Punkt und Komma operationalisieren, wie groß der Nutzen oder Schaden allein sogar für das gut zu messende Ereignis „Todesfall“ ist. Für andere Ereignisse ist es dann noch schwieriger.

Peter-Andreas Löschmann: Für das englische Register weiß ich es nicht, aber die Daten, die ich aus dem schwedischen Register kenne, sind doch, da sie epidemiologisch erhoben worden sind, relativ solide.

Oliver Sander: Im schwedischen Register ist die Vergleichsgruppe 2, 3 Jahre früher rekrutiert worden und ist nicht die Region, sondern nur die Stadt, was sicherlich einen anderen Versorgungsgrad beinhaltet. Die Vergleichsgruppe des schwedischen Registers beinhaltet selber sogar TNF-Inhibitor-behandelte Patienten, sodass es eigentlich keine sauberere Vergleichsgruppe ist, wenn man es sehr ernst damit nimmt, und sie ist ganz anders als die britische Vergleichsgruppe.

Die britische Vergleichsgruppe ist sogar später als das Biologikaregister entstanden, wodurch es wesentlich günstigere Outcomeparameter hat als das schwedische. Daher habe ich Schwierigkeiten in der Interpretation bezüglich Mortalität der schwedischen Daten und der Würdigung dieser Vergleichsgruppe. Es gibt keine besseren, weil man das natürlich nicht randomisieren kann. Deshalb muss man sich dieser Modelle bedienen, die aber auch nur Modelle sind und nicht die Wirklichkeit widerspiegeln. Je mehr Modelle es gibt, desto eher erkennt man die Fehler, die sie tatsächlich innehaben, und umso besser erkennt man, dass in der Zeit der Behandlung der rheumatoiden Arthritis sehr viel zugunsten der Patienten passiert ist, was als Hintergrundrauschen die Auswertung sehr erschwert.

Moderator Jürgen Windeler: Gibt es über das Thema Register hinaus noch weitere Punkte zu „Verschiedenes“? – Herr Krobot.

Karl J. Krobot: Ich möchte an den letzten Gedanken von Herrn Leverkus anknüpfen und die Frage stellen: Werden denn indirekte Vergleiche durchgeführt?

Beate Wieseler: Die Methodik der indirekten Vergleiche ist im Grunde genommen noch in der Entwicklung. Es gibt noch eine ganze Reihe von Unsicherheiten im Rahmen der Methodik, weshalb wir in der Nutzenbewertung primär direkte Vergleiche einsetzen. Ein Teil der Stellungnahmen nimmt Bezug auf ein Methodenpapier, in dem wir angeblich indirekte Vergleiche vorsehen – das ist das Methodenpapier zur Kosten-Nutzen-Bewertung. Da sind wir mit der Effizienzgrenze in einer etwas anderen Situation. In der Nutzenbewertung leiten wir die Ergebnisse in der Regel aus direkten Vergleichen ab.

Karl J. Krobot: Das heißt, das können Sie für dieses Verfahren ausschließen?

Beate Wieseler: In Ausnahmefällen, in bestimmten Situationen können wir gegebenenfalls auf indirekte Vergleiche zurückgreifen. Das wird zum Beispiel jetzt im Rahmen des AMNOG geschehen. Dann muss man aber sicherlich die geringere Ergebnissicherheit berücksichtigen. Bei der Problematik von Effektmodifikation, die wir eben diskutiert haben, frage ich mich allerdings, ob es Sinn macht, indirekte Vergleiche vorzunehmen, oder ob das in diesem Therapiefeld ein Vorgehen nicht weiter erschweren würde.

Karl J. Krobot: Ich stimme Ihnen im Prinzip zu. Mein Punkt ist nur: Die Frage, ob Sie die indirekten Vergleiche durchführen oder nicht, sollte an sich vorab im Berichtsplan stehen. Wenn ja, wenn Sie planen, die Vergleiche durchzuführen, sollten Sie auch die Methodik in den Berichtsplan hineinschreiben.

Beate Wieseler: Ja.

Peter-Andreas Löschmann: Ich möchte zu dem Punkt „Behandlungsdauer“ kommen, wenn ich darf. Es ist vorgesehen, nur Studien, die länger als 24 Wochen dauern, zu berücksichtigen. Da haben wir widersprochen. Ich weiß, nicht alle Firmenvertreter sind der Meinung, dass wir auch kürzere Studiendauern untersuchen sollten. Ich halte es aus prinzipiellen Erwägungen für denkbar, dass man auch kürzere Studien durchführt. Wir vertreten ein Präparat, das relativ früh entwickelt wurde. Das heißt, es gibt auch relativ viele kurze Studien. Nach allem, was wir wissen, wird die Effektgröße der Behandlung, wenn man kürzere Studiendauern berücksichtigt, unterschätzt. Bei allen Studien, die ich kenne, die länger dauern, wird der klinische Effekt mit der Zeit ausgeprägter, stärker, egal welchen Parameter man sich ansieht. Deswegen würde ich dafür plädieren, die verfügbare Evidenz mit in die Analyse aufzunehmen, also auch kürzere Studien.

Jörg Ruoff: Noch einmal zu dem eingangs zitierten Papier, das ich Ihnen auch gleich noch geben werde, den Empfehlungen der GREES-Gruppe, die ganz klar sagt, dass der Stand inzwischen ist: Die Outcomes, Signs und Symptoms ändern sich häufig früher und können früher beurteilt werden, schon nach null bis 3 Monaten, während radiologische Veränderungen etwas länger dauern. Daher die ganz klare Empfehlung, gerade für solch patientenrelevante Punkte auch schon 3 Monate zu berücksichtigen.

Markus Peceny: Ich würde das gerne noch auf Studiendesigns erweitern, die nicht gerade nach 3 oder 6 Monaten einen Cut machen. Die neueren Studien mit Biologika beinhalten oft vorzeitige oder geplante Therapieumstellungen von Placebo auf Verum, zum Beispiel im Rahmen von „Early-Escape Design“. Das wiederum steht im Einklang mit den EMA-Leitlinien zur Durchführung von klinischen Studien in der Rheumatologie. Darin steht, dass eine dauerhafte Behandlung der rheumatoiden Arthritis mit einer nicht ausreichend wirksamen Therapie schon allein aus ethischen Gründen nicht angezeigt ist. Die Empfehlung der Leitlinien ist, die Dauer auf 3 bis 6 Monate zu verkürzen oder zu beschränken und gleichzeitig vorab schon klare Regelungen zum Abbruch einer solch unwirksamen Therapie zu definieren.

Das wird auch weiterhin gestützt; ich glaube, es ist dasselbe Editorial, das Sie gerade erwähnt haben, Smolen et al., gerade am 17. Januar 2011 online veröffentlicht, „Rheumatology“. Dort argumentiert eine international besetzte Expertengruppe, darunter sogar eine Vertreterin des BfArM, dass in klinischen Studien heutzutage die Therapiedauer der nicht wirksamen Arme auf 3 Monate beschränkt sein sollte. Dabei bezieht man sich wiederum auf die EULAR-Leitlinie, in der auch dargelegt wird, dass alle 3 bis 6 Monate ein engmaschiges Monitoring

und eine entsprechende Therapieanpassung zu erfolgen hat. Daher sollte im Berichtsplan unseres Erachtens klargestellt sein, dass solche Studien in die Bewertung aufgenommen werden, die unter Berücksichtigung der EMA- und EULAR-Empfehlungen eine vorzeitige oder geplante Therapieumstellung, zum Beispiel als „Early-Escape Design“, vorsehen.

Katharina Biester: Direkt zu dem „Early-Escape“: Es ist sicherlich so, dass die Studien nicht ausgeschlossen werden, wenn ein „Early-Escape“ durchgeführt wurde oder es Möglichkeiten gab. Das muss man bei der Bewertung der Daten sicherlich berücksichtigen. Die EMA beschreibt auch, wie Studiendesigns aussehen können, damit sie länger andauern und man nicht nur Studien von 3 Monaten Dauer machen muss.

Herr Löschmann, meinten Sie, dass alle Hersteller für die 3 Monate plädieren würden?

(Peter-Andreas Löschmann: Nein! Ich habe gesagt: Ich bin mir nicht mit allen Herstellern einig!)

Ach so, dann habe ich Sie missverstanden. Ich wollte nämlich sagen, dass das nicht so ist.

(Peter-Andreas Löschmann: Ja, ja, ich weiß!)

Dann ist das ja geklärt.

Man kann nachlesen, dass weder die EMA noch die FDA für irgendeinen Endpunkt weniger als 6 Monate Studiendauer fordern. Daran haben wir uns mit der Begründung angelehnt, dass es eine progressive Erkrankung ist, Langzeitverlauf. Wir haben uns auch die Literatur angeschaut, die Sie uns bei den Argumenten genannt haben. Das waren Leitlinien und systematische Reviews, die wohl auch 3 Monate berücksichtigt haben. Dazu ist zu sagen: Das waren Empfehlungen zur Therapiesteuerung. Die 3 Monate werden zur Therapiesteuerung herangezogen. Aus diesem Grund sehen wir, das IQWiG, keinen Anlass, die Studiendauer (das Einschlusskriterium) zu verändern.

Peter-Andreas Löschmann: Darf ich kurz widersprechen? Ich habe mir genau angeguckt, welches die FDA-Guidelines bei der Entwicklung von Etanercept waren. Die ersten Studien sind 1992 durchgeführt worden, und darin stand keineswegs definitiv, dass ein halbes Jahr zu behandeln ist. Die Formulierung war sehr flexibel. Sie ist jetzt natürlich anders, aber ein wesentlicher Teil der Evidenz und der Zulassungsunterlagen basiert auf solchen Studien, die kürzere Studiendauern haben.

Deshalb möchte ich noch einmal nahelegen, darüber nachzudenken, ob Sie diesen Punkt aufrechterhalten wollen. Denn Ihr Argument hat mich eigentlich nicht widerlegt, dass wir keinesfalls den Effekt überschätzen, wenn kürzere Behandlungsdauern berücksichtigt werden. Ich denke, das ist auch gut belegt.

Jörg Ruof: Es gibt eine Diskussion in der Literatur, inwieweit es noch ethisch ist, Patienten über 6 Monate auf einem Placebo weiterlaufen zu lassen, zitiert zum Beispiel von Maarten Boers. Die Verfügbarkeit der hochwirksamen Substanzen, die jetzt glücklicherweise da sind, ermöglicht eine solche Diskussion, sollte aber auch bei Ihnen reflektiert werden. Daher unsere ganz klare Bitte, auch kürzerfristige Trials mit zu berücksichtigen – die längeren natürlich auch, das ist die Grundlage der Evidenz. Was spricht dagegen, kürzere Trials, wo man bereits einen „Rapid Onset of Action“ oder was auch immer erkennen kann, zu berücksichtigen? Das ist zusätzliche Evidenz. Für den Patienten ist eine lang anhaltende und stabile Wirkung wichtig, aber auch ein schneller Wirkeintritt. Was spricht gegen die Einbeziehung kürzerer Trials?

Beate Wieseler: Sie haben es gerade selbst genannt. Ich denke, aus diesen Gründen wird es jetzt auch aktuell von den Zulassungsbehörden so empfohlen. Wir wollen den langfristigen Nutzen für den Patienten messen. Es geht nicht primär um ein „Early Onset“, sondern es geht darum, ob die Therapie dem Patienten auf Dauer hilft.

Die ethische Frage – darauf haben Sie auch hingewiesen – wird teilweise durch die „Early-Escape Designs“ adressiert. Aber auch da dauert die Studie 6 Monate oder länger, und nicht alle Patienten werden automatisch umgestellt, sondern es kommt abhängig vom Ergebnis zu einer Änderung der Behandlung. Inwieweit das die Interpretation der Daten erschwert, ist dann, wie gesagt, in der Auswertung zu berücksichtigen.

Das ethische Argument kann ich nicht ganz nachvollziehen. Vielleicht noch einmal zurück zu der Frage: Müssen wir hier indirekte Vergleiche berücksichtigen? Warum haben wir so wenig direkt vergleichende Studien? Sie haben das ethische Argument sofort vom Tisch, wenn Sie eine direkt vergleichende Studie machen. Es sind genügend Biologika verfügbar, die Sie als aktive Komparatoren einschließen können. Wir müssten überhaupt nicht über indirekte Vergleiche diskutieren, wenn es mehr direkt vergleichende Studien gäbe. Seit zehn Jahren gibt es Biologika auf dem Markt. Warum haben wir so wenig direkt vergleichende Studien?

Friedhelm Leverkus: Ich möchte noch einmal darauf hinweisen: Wenn Sie „Early-Escape“ akzeptieren, wo 50 % der Patienten zum Teil frühzeitig, nach drei Monaten, herausgehen, dann ist nicht ganz einsichtig, warum Sie die eigentlichen Dreimonatsstudien nicht berücksichtigen. Da sind wir fast in einem ähnlichen Bereich. – Zum Ersten.

Zum Zweiten ist es so, dass die alten EMEA-Guidelines die kurzen Studien vorgeschrieben haben. Wenn Sie jetzt so streng nur nach den neuen Guidelines vorgehen, dann bedeutet das, dass Sie gerade die Evidenz von bewährten Mitteln, die schon lange am Markt sind, einfach ausschließen. Das führt an der Stelle schon zu einem Informationsbias. Daher noch einmal die Bitte, darüber nachzudenken, auch die kürzeren Studien mit hineinzunehmen und darzustellen. Dann können Sie in der Bewertung der Studien sehen, inwieweit die vorgefundene Evidenz konsistent mit der anderen Evidenz ist, anstatt sie vorzeitig auszuschließen.

Jörg Ruof: Zum einen noch einmal zu der Frage von 12 versus 24 oder 26 Wochen, auch in Richtung Patienten: Die langfristige Remission ist ein Ziel, die langfristige Beschwerdefreiheit. Ein ebenso wichtiges Ziel ist, dass man sehr rasch beschwerdefrei ist. Das sollte – vielleicht können Sie das noch einmal kommentieren, Frau Faubel – aus meiner Sicht auch als patientenrelevant berücksichtigt werden.

Zum anderen möchte ich zu den Head-to-Head-Studien für meine Firma sagen: Wir machen 3 Head-to-Head-Studien, eine in TNF-IR, eine in der Monotherapie und eine groß angelegte Safety-Studie in den USA, sodass wir die Daten liefern werden und dann hoffen, dass sie auch Berücksichtigung finden, wenn sie kommen.

Moderator Jürgen Windeler: Möchten Sie dazu noch etwas sagen?

Ursula Faubel: Natürlich ist es für Patienten auch wichtig, dass es möglichst schnell zu einem Wirkeintritt von Therapien kommt. Das ist einer der Vorteile, die die Biologicals insgesamt gegenüber dem MTX haben. Sicher ist es aber am wichtigsten, dass es auch langfristig hilft und nicht nur eine kurzfristige Wirkung erreicht wird, wobei das bei den Biologicals in der Regel nicht der Fall ist. Es ist ja kein Problem, das man wahrnimmt, dass sie nach 6 Monaten nicht mehr wirken.

Peter-Andreas Löschmann: Wir müssen aufgrund der wissenschaftlichen Genauigkeit jetzt nicht so tun, als ob wir nicht wüssten, über welche Substanzen wir reden. Ich glaube, niemand stellt infrage – zumindest für die Substanz, die ich vertrete –, dass die Substanz dann aufhört zu wirken. Das wissen wir. Insofern kann ich die theoretischen Überlegungen gut verstehen. Eine Substanz, die man nicht so gut kennt, die nur kurz wirkt, wäre natürlich nicht wünschenswert. Keine der Substanzen, die das IQWiG jetzt zu bewerten hat, steht in dem Verdacht, nur kurz zu wirken; ich werde die Ausnahme nicht nennen, die kennen Sie selbst. Insofern sollten wir versuchen, tatsächlich die beste verfügbare Evidenz zu nutzen, anstatt theoretische Überlegungen, die für eine Substanz, die dann gar nicht untersucht wird, vielleicht zutreffen könnten, als Entscheidungskriterium zu nehmen. – Mehr sage ich zu dem Thema nicht.

Markus Peceny: Daran mag ich nur noch illustrierend anschließen, welche Effekte man in der Auswertung und auch der Evidenzbasis mit dem formal zwar begründeten, aber doch recht artifiziellen Kriterium „24 Wochen“ teilweise vergibt. Nun haben wir ausgerechnet: Eine Studie geht 22 Wochen im Vergleich, was an der Applikationsfrequenz eines Produkts von uns liegt. Der nächste sinnvolle Zeitpunkt wären 30 Wochen gewesen, was schon sehr lange ist. Das ist nun ausgerechnet eine sehr große, sehr gute Studie, die sogar Patienten mit Begleiterkrankungen und allem eingeschlossen hat, die in anderen randomisierten Studien oft nicht eingeschlossen waren, über 1000 Patienten, hervorragende Evidenzbasis. So etwas würde bei einem akademisch gefällten Cut wegfallen, was eigentlich schade wäre. Das wollte ich nur noch als Denkanstoß platzieren.

Moderator Jürgen Windeler: Gut, danke. Ich weiß nicht, ob 22 Wochen weniger akademisch gewählt sind und würde mich auch bei 3 Monaten fragen, welche akademische Grundlage das hat. Aber wir sind ja dabei, die Argumente noch einmal zu hören, auszutauschen und uns Gedanken darüber zu machen. Wir werden das in die weiteren Bewertungen aufnehmen. Mehr kann ich hier nicht dazu sagen. – Bitte schön.

Thomas Reimann: Ich möchte noch einmal auf den Punkt eingehen. Auch Herr Sander hat ausdrücklich darauf hingewiesen, wie notwendig es ist, möglichst genaue, scharfe Kriterien zu haben, um eine möglichst maximale Evidenz herausarbeiten zu können. Wenn dann ein Teil der verfügbaren Daten aufgrund eines Cut-offs nicht genutzt wird, ist das einfach bedauerlich. Bei aller Maximierung, Evidenz herauszuarbeiten, würde tatsächlich etwas verloren gehen, wenn ein Datenfundus nicht genutzt würde, der aber, wie Herr Löschmann auch ausgeführt hat, nicht das Potenzial hätte, eine Überbewertung der Wirksamkeit mit sich zu bringen aufgrund dessen, dass der Wirkeintritt in eine Separation führt, wenn wir in Richtung 24 Wochen gehen.

Moderator Jürgen Windeler: Noch einmal der Hinweis: Wir sitzen hier, damit wir Argumente aufnehmen. Ich will aber auch zum Ausdruck bringen, dass eine Maximierung von Informationen und Evidenz nicht das Ziel der Aktivitäten ist, jedenfalls nicht so, wie man es verstehen kann. Wir wollen schon konkrete Fragen mit der dafür zielgerichteten Evidenz beantworten. Deswegen würde ich mir auch die Frage stellen, die wir vielleicht noch einmal erörtern: Was heißt es in der Konsequenz, wenn man die Sechsmonatsdaten hat und, wie Sie sagen, zusätzlich die Dreimonatsdaten heranziehen soll? Heißt das dann in der Konsequenz – und wir hatten solche Diskussionen in den letzten Jahren in diesem Institut –, dass man mit den Sechsmonatsdaten keinen Effekt sieht, mit den Dreimonatsdaten hat man aber einen? Was macht man dann damit?

Ich möchte damit nur zum Ausdruck bringen: Es geht nicht darum, alles an Evidenz zu sammeln, was es so gibt, sondern es geht darum, sich – das versucht der Berichtsplan zu tun – Gedanken darüber zu machen, einigermaßen gezielt die Evidenz, die man für die Beantwortung der Fragen braucht, die hier gestellt werden sollen, heranzuziehen. Dass wir uns mit der Frage noch einmal befassen, habe ich ja zum Ausdruck gebracht. – Herr Ruof.

Jörg Ruof: Aus meiner Sicht steht die Patientenrelevanz ganz im Vordergrund. Bei Methotrexat ist die normale Therapie so: Man fängt mit einem Steroidstoß an, um die Patienten zunächst einmal sofort beschwerdefrei zu bekommen, und dann beginnt das Methotrexat über die Zeit zu wirken. Das sind die beiden Grundpfeiler der Therapie: rascher Wirkeintritt und lang anhaltende, nachhaltige Wirkung. Aus meiner Sicht ist das patientenrelevant und sollte in Ihren Kriterien Berücksichtigung finden.

Moderator Jürgen Windeler: Okay. – Weitere Wortmeldungen? – Sehe ich auch mit einem gewissen Abwarten nicht. Dann sind wir am Ende der Tagesordnung.

Ich bedanke mich dafür, dass Sie zu der wissenschaftlichen Erörterung gekommen sind. Ich bedanke mich auch ausdrücklich für Ihre Hinweise und Argumente, Ihre Antworten auf unsere Fragen.

Damit beende ich die Anhörung und wünsche Ihnen einen guten Heimweg. – Danke.

Anhang: Dokumentation der Stellungnahmen

Inhaltsverzeichnis

	Seite
A 1 Stellungnahmen von Organisationen, Institutionen und Firmen	A 2
A 1.1 Abott GmbH & Co. KG	A 2
A 1.2 Bristol-Myers Squibb GmbH & Co. KGaA.....	A 30
A 1.3 Deutsche Gesellschaft für Rheumatologie.....	A 37
A 1.4 Deutsche Rheuma-Liga Bundesverband e. V.	A 42
A 1.5 MSD Sharp & Dohme GmbH	A 46
A 1.6 Pfizer Pharma GmbH	A 53
A 1.7 Roche Pharma AG	A 60
A 1.8 UCB Pharma GmbH.....	A 93
A 1.9 Verband Forschender Arzneimittelhersteller e. V.....	A 97

A 1 Stellungnahmen von Organisationen, Institutionen und Firmen

A 1.1 Abott GmbH & Co. KG

Autoren:

Görtz, Anita

Simianer, Stefan, Dr.

Wolff, Malte, Dr.

Stellungnahme von Abbott GmbH & Co. KG zu dem vorläufigen Berichtsplan „Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis“ (Auftrag A10-01, Version 1.0 vom 20.09.2010)

Im Rahmen des schriftlichen Stellungnahmeverfahrens zum vorläufigen Berichtsplan A10-01 vom 20. September 2010 reichen wir im Namen der Abbott GmbH & Co. KG (nachfolgend Abbott Deutschland genannt) folgende Kommentierungen ein.

Allgemeine Anmerkungen

Der vom IQWiG vorgelegte Berichtsplan ist sehr umfangreich und in weiten Teilen stimmen wir den Überlegungen zu. Wie im einleitenden Teil des Berichtsplans dargelegt, handelt es sich bei der rheumatoiden Arthritis (RA) um eine progressiv verlaufende Erkrankung, die nicht nur mit einer starken persönlichen Belastung als direkte Konsequenz der Krankheitsaktivität einhergeht, sondern auch mit einer starken indirekten Belastung verbunden ist, die sich beispielsweise in einer Minderung der gesundheitsbezogenen Lebensqualität und der Arbeitsfähigkeit äußert.

Da das bei einer RA charakteristische entzündete Gewebe Zeit zur Regeneration benötigt, stimmen wir damit überein, dass eine Mindestlaufzeit von 24 Wochen für Studien angemessen ist. Dies gilt insbesondere vor dem Hintergrund, dass hohe Wirksamkeitsschwellenwerte ins Auge gefasst werden. In diesem Zusammenhang sollte berücksichtigt werden, dass viele Studien aus ethischen Gründen Early-Escape-Arme, meist in Woche 16, beinhalten. Sofern die Intention-to-Treat - Analyse (ITT) durchgeführt wurde und angemessene

Imputationsmethoden angewandt wurden, sollten diese Studien ebenfalls in den Analysen berücksichtigt werden.

Da mehrere Alternativen einer medikamentösen Behandlung existieren, erscheint der a priori vorgenommene Ausschluss von Studien mit Vergleichsarmen, die ausschließlich Placebo beinhalten (d. h. keine laufende Therapie mit sDMARDs), bei den Meta-Analysen angemessen.

Der Berichtsplan enthält jedoch auch eine Vielzahl an Abschnitten und Aussagen, deren zugrundeliegende Argumentation wir nicht nachvollziehen können. Im Folgenden führen wir eine Reihe von Kommentierungen und Vorschlägen für Bereiche auf, die sowohl aus methodischer Perspektive als auch bezüglich der berücksichtigten Endpunkte weiterer Erläuterungen bzw. Änderungen bedürfen.

Aus diesem Grund nehmen wir im Namen von Abbott Deutschland Stellung und schlagen ggf. Änderungen vor:

Kapitel 1 „Ziele der Behandlung“

Zitat

„Untersuchungen der letzten Jahre haben gezeigt, dass nur eine Remission und damit eine weitgehende Verminderung der Krankheitsaktivität das Fortschreiten der Gelenkzerstörung aufhalten kann [5,19,20]. So beschreiben z. B. Smolen & Aletaha [19], dass mit Erreichen der Remission die Knochenzerstörung aufgehalten wird.“

Kommentierung

Dieser Satz sagt aus, dass die Gelenkzerstörung nur während einer Remission gehemmt werden kann. Das Aufhalten der Knochenzerstörung mit Erreichen der Remission wird als zwangsläufig zusammengehörig dargestellt. Dies trifft auf die meisten Patienten unter Behandlung mit konventionellen DMARDs, darunter MTX, zu. Allerdings existiert bei diesen Behandlungen eine Subgruppe von Patienten, die zwar einerseits ein hohes klinisches Ansprechen auf die Behandlung aufweisen, bei denen aber andererseits ein Prozess zugrunde liegt, der weiterhin zu Gelenkzerstörung führt (vgl. Breedveld *et al.* 2006). Diese Patienten werden auch als „Silent Progressors“ bezeichnet (vgl. Wollenhaupt *et al.* 2009). Demgegenüber stehen Daten zur Behandlung mit verschiedenen TNF-Inhibitoren, die eine direkte Hemmung der Mechanismen, die zu Gelenkzerstörung führen, belegen. Im Fall von Adalimumab gibt es Nachweise, dass sowohl die Gelenkzerstörung als auch die Knorpelschädigung gehemmt werden (vgl. Keystone 2004). Die Hemmung beruht hier nicht allein auf der Unterdrückung der Entzündung, wie bei konventionellen DMARDs der Fall, und unterscheidet sich womöglich auch von der anderer TNF-Inhibitoren und Biologika, die einen anderen Wirkmechanismus aufweisen.

Somit stellt die Hemmung der Gelenkzerstörung an sich ein Therapieziel dar und sollte losgelöst von der Remission angeführt werden (vgl. Wollenhaupt *et al.* 2009). Das Vorliegen fortschreitender Gelenkzerstörung unter der Behandlung mit konventionellen DMARDs trotz des erheblichen Rückgangs der Symptome bzw. klinischen Krankheitsaktivität unterhalb festgelegte Schwellenwerte ist alarmierend, da dies letztendlich zu funktionalen Schäden bei den Patienten führen wird. Daher stellt die Hemmung der radiologischen Progression eine eigenständige und relevante Verbesserung zusätzlich zur Senkung der Krankheitsaktivität dar.

Änderungsantrag

Dieser Passus im Berichtsplan sollte dahingehend geändert werden, dass klinische Remission bzw. niedrige Krankheitsaktivität und die Hemmung der Gelenkzerstörung zwei unabhängige Therapieziele sind, obwohl sie sicherlich zusammenhängen.

Zitat

„Eine lediglich geringe Teilminderung der Krankheitsaktivität, wie sie z. B. mit manchen Responsekriterien auf pharmakologische Interventionen (wie ACR20) gemessen wird, ist hinsichtlich ihrer klinischen Relevanz nicht definiert.“

Kommentierung

Obwohl die Bedeutung des ACR20-Ansprechens als patientenrelevanter Endpunkt kontrovers diskutiert werden kann, sollte berücksichtigt werden, dass regulatorische Vorgaben dies als Endpunkt zur Messung der klinischen Symptome einer RA empfohlen haben. Dementsprechend wurde das ACR20-Ansprechen besonders in den meisten frühen RCTs als primärer Endpunkt definiert. Diese Studien waren regelhaft darauf gepowert, signifikante Unterschiede im ACR20-Ansprechen nachzuweisen.

Änderungsantrag

Aufgrund der regulatorischen Vorgaben sollte die ACR20-Response als zusätzliche Information in die Analyse aufgenommen werden (siehe auch unsere Kommentierung zu Kapitel 4.1.3).

Kapitel 2 „Ziele der Untersuchung“

Zitat

„Unter Zweitlinientherapie wird im Rahmen der vorliegenden Nutzenbewertung der Einsatz von biotechnologisch hergestellten Arzneimitteln bei Personen, die mit einem krankheitsmodifizierenden Antirheumatikum vorbehandelt sind, verstanden. Hierbei ist der erste Einsatz des jeweils zu untersuchenden Arzneimittels gemäß Zulassungsstatus zu betrachten.“

Kommentierung

Laut Fachinformationen der meisten Biologika sind diese zur Behandlung von Patienten angezeigt, die auf mindestens ein sDMARD nicht mehr oder nicht ausreichend angesprochen haben. Dies spiegelt sich auch in den Behandlungsleitlinien wider, die die Behandlung mit Biologika nach dem Versagen mindestens eines sDMARD (meist Methotrexat) empfehlen und entspricht auch der routinemäßigen klinischen Praxis.

Gemäß der aktuellen EULAR-Guidelines beinhaltet der Behandlungsalgorithmus drei Phasen mit entsprechenden Eskalationsstufen. Die erste Phase (Erstlinie) bezieht sich auf die Behandlung von sDMARD naiven Patienten, die zweite Phase auf mit sDMARD vorbehandelten und gleichzeitig Biologika-naiven Patienten und die dritte Phase auf Patienten, die bereits mit Biologika vorbehandelt

wurden. Die Empfehlungen benennen die erste Phase explizit als Erstlinientherapie (Empfehlung 14), woraus geschlossen werden kann, dass die Phasen II und III sich auf Zweitlinien- und Drittlinientherapie beziehen (vgl. Smolen *et al.* 2010).

Ebenso spiegelt sich dieses Prinzip im derzeitigen Zulassungsstatus der Biologika innerhalb des Therapiealgorithmus wider. Die Erstlinientherapie bezieht sich auf sDMARD-naive Patienten, die Zweitlinientherapie auf Patienten, die bereits mit sDMARD behandelt wurden. Somit findet sich Definition der Zweitlinientherapie sowohl in den neuesten Empfehlungen, als auch in den aktuellen Fachinformationen wieder. Sie entspricht unserem Verständnis einer Definition einer Zweitlinientherapie im Kontext dieser Nutzenbewertung.

Änderungsantrag

Zweitlinientherapie sollte als die Behandlung im Anschluss an das Versagen oder das ungenügende Ansprechen von mindestens einem sDMARD definiert werden und sich auf Patienten mit aktiver RA beziehen.

Studien, die systematisch Patienten aufnehmen, die zuvor mit Biologika behandelt wurden, sollten als Drittlinientherapie angesehen und nicht berücksichtigt werden.

Kapitel 3.1 „Zeitlicher Verlauf des Projekts“

Zitat

„Während der Erstellung des Berichtsplans wurden am 27.07.2010 Patientenvertreter des Deutschen Rheuma-Liga Bundesverbandes e. V. und der Deutschen Vereinigung Morbus Bechterew e. V. zur Festlegung patientenrelevanter Zielgrößen konsultiert.“

Kommentierung

Uns überrascht die Wahl von Patienten mit Morbus Bechterew zur Beurteilung patientenrelevanter Outcomes bei RA. So wird beispielsweise in Abschnitt 4.1.3 als Ereignis Versteifungen aufgeführt, was kein üblicher oder oft auftretender Endpunkt des Entzündungsprozesses einer RA ist, allerdings häufig bei Morbus Bechterew auftritt.

Es hat uns außerdem sehr überrascht, dass weder eine wissenschaftliche Fachgesellschaft (z. B. die DGRh) noch ein Ärzteverband (z. B. der BDRh) konsultiert wurden.

Änderungsantrag

Bitte überprüfen Sie noch einmal den Bericht, insbesondere den Abschnitt zu patientenrelevanten Endpunkten, ob sich diese auf die Erkrankung Morbus Bechterew beziehen, und passen Sie ihn gegebenenfalls an.

Kapitel 4.1.2 „Prüf- und Vergleichsintervention“

Zitat

„Rituximab (MabThera[®])“

Kommentierung

Ziel dieser Bewertung ist die Beurteilung von Biologika als Zweitlinientherapie. In Kapitel 2 „Ziele der Untersuchung“ wird Zweitlinientherapie als erster Einsatz des zu untersuchenden Arzneimittels nach einer Behandlung mit sDMARDs definiert.

Laut der europäischen Zulassung ist Rituximab (MabThera[®]) in Kombination mit Methotrexat bei der Behandlung erwachsener Patienten mit schwerer

aktiver RA angezeigt, die unzureichend auf andere DMARDs, einschließlich einer oder mehrerer Behandlungen mit Tumornekrosefaktor(TNF) - Inhibitoren, ansprachen oder diese nicht vertrugen.

Unter Berücksichtigung der in Ihrer Bewertung enthaltenen Definition für Zweitlinientherapie und der Zulassung von Rituximab zur Behandlung von RA durch die EMA, ist die Behandlung mit Rituximab als Drittlinientherapie zu klassifizieren.

Hinsichtlich der klinischen Implikationen befolgen deutsche Rheumatologen diese Einschätzung in ihren derzeitigen Behandlungsrichtlinien (vgl. Wollenhaupt *et al.* 2009). Diese beschreiben Rituximab als Therapieoption nach Versagen von sDMARDs und dem anschließenden Versagen von mindestens einem TNF-Inhibitor.

Die FDA wies die Zulassung von Rituximab für die Zweitlinientherapie zurück und bestätigte nur den Einsatz bei Patienten in der Drittlinientherapie. Gleichmaßen hat die EMA Rituximab noch nicht für die Zweitlinientherapie zugelassen.

Änderungsantrag

Bei Rituximab/MabThera[®] handelt es sich nicht um ein Präparat zur Behandlung in der Zweitlinie und daher sollte es nicht im Rahmen der vorliegenden Nutzenbewertung bewertet werden.

Zitat

„Die Anwendung der in den Studien eingesetzten Prüf- und Vergleichsinterventionen muss im Rahmen des für Deutschland gültigen Zulassungsstatus erfolgen.“

Kommentierung 1

Die Fachinformationen der Biologika unterscheiden sich hinsichtlich möglicher Kombinationen mit anderen Arzneimitteln und der Möglichkeit, sie in der Monotherapie zu verwenden. MTX ist üblicherweise das Kombinationspräparat, das in RCTs geprüft wurde. Eine Kombination mit anderen sDMARDs ist möglich, obwohl die Fachinformationen dies nicht explizit angeben. Aufgrund einer Unverträglichkeit gegenüber MTX werden solche Kombinationen relativ häufig als Alternative zu einer Monotherapie genutzt. Dies wird auch von der DGRh als patientenrelevant eingestuft (vgl. Krüger *et al.* 2008).

Änderungsantrag 1

Wir empfehlen die Berücksichtigung auch von Studien mit Kombinationstherapien, bei denen nicht MTX sondern andere sDMARDs als Kombinationspartner gegeben werden (z. B. Furst 2003). Dies gilt vor allem für Wirkstoffe, die gemäß der Fachinformation nicht zwangsläufig mit MTX kombiniert werden müssen.

Kommentierung 2

Entsprechend der Fachinformation wird die MTX-Behandlung mit 7,5 mg/Woche begonnen und dann rasch auf 15-20 mg als durchschnittliche wöchentliche Dosierung gesteigert. Dieses Behandlungsschema wird empfohlen, um MTX in der höchsten noch verträglichen Dosierung geben zu können. Es ist weithin bekannt, dass Dosissteigerungen von 2,5 oder 5 mg/Woche im Hinblick auf Wirksamkeit deutlich zu Erfolg oder Versagen einer Behandlung mit MTX beitragen. Üblicherweise sollte eine Dosierung von 15-20 mg/Woche in der klinischen Praxis bei Patienten mit hoher Krankheitsaktivität beibehalten werden.

Systematisch niedrigere Dosen führen zu einer Verzerrung bei MTX als Vergleichspräparat. Obwohl argumentiert werden kann, dass die gleiche Verzerrung bei dem Studienarm mit Kombinationstherapie der Studie existiert, kann das potentiell wirksamere Biologikum in dem Studienarm entgegengehalten werden. Dies gilt besonders für die Patienten, die zuvor nur ungenügend auf MTX angesprochen haben.

Änderungsantrag 2

Studien, in denen MTX nicht in einer durchschnittlichen Dosierung von 15 mg/Woche oder höher gegeben wird, sollten nicht in der Auswertung berücksichtigt werden, da diese eine Verzerrung zulasten des Vergleichsarmes verursachen.

Kapitel 4.1.3 „Patientenrelevante Endpunkte“

Zitat

„Remission“

Kommentierung

Remission wird als erster patientenrelevanter Endpunkt angegeben. Es ist der einzige in Abschnitt 4.1.3. aufgeführte kombinierte Endpunkt (auch unter Berücksichtigung der in Abschnitt 1 des vorläufigen Berichtsplanes gemachten Vorüberlegungen).

Wie der Berichtsplan angibt, existieren verschiedene Definitionen für „Remission“, die bis ins Jahr 1981 zurückreichen (vgl. Pinals *et al.* 1981). Damals wurden Remissionskriterien anhand eines einfachen Chart Reviews festgelegt, das mit den Einschätzungen teilnehmender Rheumatologen verglichen wurde.

Erst in den letzten 5-8 Jahren fanden validierte Remissionskriterien, die auf den verschiedenen DAS-Kriterien basieren, Einzug in die Standardberichte der RCTs. Dies geschah aufgrund der höheren Erwartungen an die Wirksamkeit von Medikamenten zur Behandlung der RA. Allerdings wurde in frühen RCTs mit z. B. Etanercept und Infliximab der DAS überhaupt nicht berücksichtigt. Erst im Zuge des Entwicklungsprogramms von Adalimumab, das Mitte bis Ende der 1990iger Jahre begann, wurden die damals neu entwickelten DAS-Kriterien zum ersten Mal berücksichtigt. Diese basierten ursprünglich auf der Erfassung von 53 Gelenken, bei denen die Druckempfindlichkeit festgehalten wurde (der sogenannte Ritchie-Score), und später auf der Erfassung von 66/68 Gelenken, wie auch die ersten DAS-Remissionskriterien mit einem Schwellenwert von 1,6 (vgl. van der Heijde 1990). Die am häufigsten angewandten DAS-Remissionskriterien haben heute einen Schwellenwert von 2,6 und basieren auf der Erfassung von 28 Gelenken. Im Gegensatz zu der Erfassung von 66/68 Gelenken werden bei der Erfassung von 28 Gelenken die Füße nicht untersucht.

Zusätzlich gibt es DAS-Scores, die BSR und CRP als Akute-Phase-Reaktanten nutzen. IL-6-Hemmer wie Tocilizumab unterdrücken die CRP-Produktion in der Leber, weshalb eine Senkung der CRP-Werte unter einer solchen Medikation nicht zwangsläufig eine klinische Wirksamkeit oder Patientenrelevanz bedeuten muss, obwohl der DAS-Score, der das CRP beinhaltet, per Konstruktion sinkt.

Wir stimmen zu, dass eine Remission der Endpunkt mit der höchsten Patientenrelevanz ist. Allerdings sollte berücksichtigt werden, dass die angewandten Remissionskriterien von Studie zu Studie unterschiedlich sein können. Um eine Verzerrung durch diese Unterschiede in berichteten Endpunkten auszuschließen, könnte die Auswertung basierend auf den berichteten Daten erfolgen. Eine angemessene Option könnte die

Verwendung der berichteten Daten in einem einheitlichen Score sein, wobei immer der gleiche Algorithmus angewandt werden müsste.

Zudem hat die ausschließliche Verwendung des DAS-Scores oder anderer Remissionskriterien einen weiteren Nachteil: Sie verwenden einen absoluten Wert minimaler Krankheitsaktivität auf. Auf den ersten Blick und vor dem Hintergrund der Patientenrelevanz erscheint dies zwar wünschenswert, allerdings hängt der Anteil der Patienten, die eine DAS-Remission während eines RCT erreichen, dadurch mehr von den Werten der Krankheitsaktivität und der Gelenkschädigung zu Studienbeginn ab als von der Wirksamkeit des Arzneimittels. Je niedriger die durchschnittliche Krankheitsaktivität zu Studienbeginn ist, desto höher ist die Wahrscheinlichkeit, dass der Patient eine Remission erreicht. Im Fall einer Langzeit-RA mit ausgeprägter Krankheitsaktivität und Gelenkschädigung zu Studienbeginn stellt das Kriterium einer niedrigen Krankheitsaktivität, wie sie durch einen DAS28 unter 3,2 definiert wird, ein anerkanntes Therapieziel dar und sollte entsprechend berücksichtigt werden (vgl. Smolen *et al.* 2010).

ACR-Responsekriterien haben hingegen den Vorteil, dass Veränderungen gegenüber Studienbeginn aufgezeigt werden, was im Zusammenhang mit der Wirksamkeit eines Arzneimittels bedeutsam ist und dazu beitragen könnte, Verzerrungen durch den unterschiedlichen Baseline-Status zu reduzieren. Für einige Patientengruppen könnte dies zudem im Alltag relevanter sein. Hinzu kommt, dass ACR-Kriterien einfacher und vergleichbarer sind, da sie in nahezu allen RCTs verwendet wurden. ACR-Response-Werte von 50 % und höher sind mit Sicherheit patientenrelevant und von klinischer Bedeutung.

Das ACR-Ansprechen erfüllt die Validitätskriterien und ist eng verbunden mit anderen patientenrelevanten Outcomes (vgl. Felson *et al.* 1993). Es wird

von Zulassungsbehörden wie der FDA oder der EMA (EMEA 2003) verlangt und ist der am häufigsten verwendete Endpunkt bei Health Technology Assessments, wie bei mehreren Cochrane-Reviews. Das ACR-Ansprechen ist damit das am besten geeignete Kriterium, um patientenrelevante Verbesserungen studienübergreifend zu vergleichen.

Änderungsantrag

Im Berichtsplan sollten die ACR50- und ACR70-Responsekriterien als bevorzugte Endpunkte für die Messung der patientenrelevanten klinischen Verbesserung (zusätzlich zu den Remissionskriterien) aufgenommen werden.

Es existieren verschiedene Scores für Krankheitsaktivität, deren Berechnung sich im Laufe der Zeit geändert hat. Zusätzlich kann die Relevanz verschiedener Scores je nach Patientengruppe variieren. Da alle diese Scores validiert sind und validierte Schwellwerte haben, sollten alle veröffentlichten Ergebnisse berücksichtigt werden. Von Vergleichen oder zusammenfassenden Darstellungen sollte jedoch aufgrund der sehr unterschiedlichen Berechnung Abstand genommen werden.

Zitat

„Symptomatik der Entzündung (insbesondere Schmerz, Fatigue, Morgensteifigkeit)“

Kommentierung

Die charakteristische Eigenschaft und Hauptsymptomatik einer aktiven rheumatischen Entzündung sind in allen Stadien der Erkrankung die Gelenkschwellung und der Gelenkschmerz. Dies gilt insbesondere für Patienten, bei denen strukturelle Schäden mit irreversiblen Veränderungen und sekundärer Arthrose sowie Behinderung und Verlust der Funktion noch

nicht eingetreten sind. Ferner ist die Beurteilung dieser Endpunkte verhältnismäßig objektiv.

Änderungsantrag

Im Berichtsplan sollten die ACR-Core-Parameter wie geschwollene und schmerzende Gelenke zur Liste der Symptome hinzugefügt werden.

Zitat

„Strukturelle Gelenkveränderungen (Deformitäten, Versteifungen, Kontrakturen)“

Kommentierung

Deformitäten, Versteifungen und Kontrakturen sind patientenrelevant, jedoch sind sie alle das Ergebnis einer lang anhaltenden, nicht kontrollierten rheumatischen Entzündung, die über einen langen Zeitraum, häufig über Jahrzehnte, fortgeschritten ist (mit Ausnahme von Versteifungen, die nicht häufig bei RA auftreten, sondern vornehmlich bei Morbus Bechterew zu finden sind). Diese Endpunkte werden als irreversibel betrachtet. Eine wirksame Behandlung ist in der Lage, die Entstehung eines solchen irreversiblen Schadens zu verhindern, aber nicht, bereits entstandenen Schaden signifikant umzukehren.

In einigen RCTs werden strukturelle Schäden als einer der wichtigsten Endpunkte geführt. Dieser wird anhand von Veränderungen in Röntgenaufnahmen von Gelenken als radiographische Progression gemessen. Die am häufigsten verwendeten Methoden umfassen den Larsen-Score, den Sharp-Score oder Modifizierungen des Sharp-Score durch van der Heijde oder Genant (vgl. Sharp *et al.* 1971, van der Heijde *et al.* 1989, Larsen *et al.* 1975 oder als Übersicht van der Heijde 2000). Scores zur radiographischen Progression nach der Methode von Sharp messen die

Gelenkspaltverengungen und Gelenkerosionen an Händen und Füßen. Beide untersuchten Bereiche hängen offensichtlich mit Knorpelrückbildung und Deformitäten der Gelenke zusammen.

Es wurde bereits gezeigt, dass der Sharp-Score auf lange Sicht mit der Funktionalität zusammenhängt. Dieser Zusammenhang nimmt mit der Krankheitsdauer zu. Eine zwölfjährige Untersuchung kam zu dem Ergebnis, dass Sharp-Scores eine unabhängige Determinante der Funktionalität sind (vgl. Drossaers-Bakker *et al.* 1999). Diese Feststellung wurde durch eine Reihe von Studien und Autoren bestätigt (vgl. Scott *et al.* 2000, Welsing *et al.* 2001, Aletaha *et al.* 2006 und Odegard *et al.* 2006). Es wurde zudem nachgewiesen, dass die radiographische Progression in Händen und Füßen auch die Progression in größeren Gelenken widerspiegelt, die nicht regelmäßig mithilfe bildgebender Verfahren untersucht werden.

Mit Daten von Patienten unter Biologikabehandlung konnten van der Heijde *et al.* (2008) zeigen, dass radiographische Progression mit einer Verschlechterung der physischen Funktion einhergeht – selbst im Zuge einer kurzen Beobachtungszeit von zwei Jahren (vgl. van der Heijde *et al.* 2008). Zudem war diese Entwicklung unabhängig von der Krankheitsaktivität. Die Auswertung bestätigte die Ansicht, dass selbst geringe durchschnittliche Veränderungen der radiographischen Progression einen messbaren Einfluss auf die Behinderung haben. Kavanaugh *et al.* (2004) konnten einen Zusammenhang zwischen radiographischer Progression und Arbeitsunfähigkeit nachweisen (vgl. Kavanaugh *et al.* 2004 und van Vollenhoven *et al.* 2010).

Eine Expertengruppe (Bruynesteyn *et al.* 2002 und 2005) hat Veränderungen von 5 Einheiten des modifizierten Total Sharp-Scores (mTSS) als minimal klinisch relevant definiert. Laut einer Empfehlung der European League against Rheumatology (EULAR) ist eine Veränderung

des Sharp-Score um 4 Einheiten als klinisch relevant anzusehen (vgl. Combe *et al.* 2007). Diese Schwellenwerte entsprechen den Smallest Detectable Differences (SDD) dieser Methoden und können für die Responderanalyse verwendet werden. Auf der anderen Seite sollte die Zahl der Patienten, die keine radiographische Progression aufweisen (in der Regel definiert als eine Veränderung von weniger als 0,5TSS), als Schwellenwert genutzt werden, da dadurch die additive und irreversible Eigenschaft der Gelenkzerstörung berücksichtigt wird.

Änderungsantrag

Radiographische Progression sollte als Surrogatparameter und Endpunkt für die Messung von strukturellen Schädigungen verwendet werden. Dieser Endpunkt ist validiert und spiegelt patientenrelevante Outcomes wider. Wie zuvor angemerkt, gilt dies auch für Patienten, die sich in Remission befinden und ist damit unabhängig von der klinischen Symptomatik.

Zitat

„Körperlicher Funktionsstatus einschließlich Aktivitäten des täglichen Lebens“

Kommentierung

Es existieren verschiedene Scores zur Messung des körperlichen Funktionsstatus bei RA. Abgesehen vom Physical Component Score (PCS) des SF-36 mit einer etablierten MCID bei RA (vgl. Kosinski *et al.* 2000) ist die am häufigsten verwendete Skala zur Messung des körperlichen Funktionsstatus in klinischen Studien der Health Assessment Questionnaire (HAQ). Er ist gut validiert und hat eine etablierte MCID von -0,22 (vgl. Wells 1993).

Änderungsantrag

Abgesehen vom PCS der SF-36 sollte der HAQ als Endpunkt zur Messung der Funktionalität und -0,22 als MCID verwendet werden.

Zitat

„Unerwünschte Arzneimittelwirkungen“

Kommentierung/ Änderungsantrag

Im Hinblick auf unerwünschte Ereignisse sollten die wichtigsten Endpunkte definiert werden.

Kapitel 4.1.4 „Studientypen“

Zitat

„Für den zu erstellenden Bericht werden daher ausschließlich RCTs als relevante wissenschaftliche Literatur in die Nutzenbewertung einfließen.“

Kommentierung

Nach der Methodik der Evidenzbasierten Medizin sollte abhängig von der Fragestellung die beste verfügbare Evidenz berücksichtigt und als Basis für die Auswertung dienen. Das bedeutet, dass im Ausgangspunkt sämtliche vorliegende Evidenz überprüft und eine Auswahl basierend auf den vorliegenden Informationen getroffen werden sollte. Der vorläufige Berichtsplan schließt jedoch andere Evidenzgrade als RCTs a priori aus.

Beispielsweise ist im Hinblick auf die Progression einer RA und die Relevanz der langfristigen Outcomes nicht nur die sofortige Wirkung der Behandlung wichtig, sondern auch das Aufrechterhalten dieser Wirkung. Diese können zweifelsohne in großen Beobachtungsstudien, offenen

Extensionsphasen usw. bewertet werden. Aus ethischen Gründen können Langzeitstudien über mehrere Jahre nur mit nicht-verblindeten Studiendesigns durchgeführt werden.

Änderungsantrag

Wir schlagen vor, andere Evidenzgrade nicht vorab auszuschließen.

Kapitel 4.1.7 „Vorgehen im Falle einer Zulassungsänderung im Projektverlauf“

Zitat

„Die jeweils vorgenommenen Änderungen werden im Vorbericht bzw. im Abschlussbericht explizit vermerkt.“

Kommentierung/Änderungsantrag

Laut dem Berichtsplan werden Änderungen im Vorbericht und im Abschlussbericht vermerkt. Wir schlagen vor, dass Änderungen im Berichtsplan, die mit dem Zulassungsstatus zusammenhängen, zeitnah nach der Durchführung der Änderung auf der Internetseite des IQWiG veröffentlicht werden sollten. Dies erhöht die Transparenz bei geringem zusätzlichem Aufwand.

Kapitel 4.3 „Informationsbewertung“

Zitat

„Umsetzung des ITT-Prinzips“

Kommentierung

Das ITT-Prinzip wird in der Regel in klinischen Studien zur RA umgesetzt. Wir weisen allerdings darauf hin, dass die Imputationsmethoden zwischen verschiedenen Studien abweichen. LOCF-Analysen kombinierter Endpunkte können Verzerrungen der Ergebnisse zwischen dem Interventions- und Kontrollarm mit sich bringen. Die Bewertung anhand der NRI-Methode wiederum schwächt diese Verzerrung ab.

Änderungsantrag

Die angewandte Imputationsmethode ist ebenfalls zu extrahieren. Diese Unterschiede sollten auch berücksichtigt werden, wenn Sie Daten aus Studien auswerten, da die Imputationsmethode zu einer Verzerrung der Ergebnisse sowohl der Meta-Analyse als auch der Meta-Regression führen kann.

Zitat

„Ergebnisgesteuerte Berichterstattung“

Kommentierung

Vor dem Hintergrund der Vergleichbarkeit der Daten, sollte auch der zeitliche Zusammenhang des Zeitpunkts der Endpunkterhebung und der Dosierungsintervalle berücksichtigt werden: Ursprünglich wurde im Rahmen von RCTs die klinische Krankheitsaktivität von Patienten an einem Tag ausgewertet, an dem der Patient auch die Prüfsubstanz erhielt. Dies hatte administrative Gründe und traf insbesondere dann zu, wenn das Arzneimittel intravenös gegeben wurde. Dies entsprach regelhaft dem Ende des Dosierungsintervalls.

Insbesondere im Fall von Arzneimitteln mit längeren Applikationsintervallen wurde vermutet, dass die klinische Aktivität nicht über den gesamten

Zeitraum gleichbleibend ist, d. h. dass sie über einige Tage nach der Gabe bis zu einem Spitzenwert ansteigen und anschließend wieder sinken würde. Zum Ende des Applikationsintervalls erreicht sie den niedrigsten Wert. Dies wurde durch Patientenberichte bestätigt und sowohl durch pharmakokinetische Daten als auch durch die allgemeinen Gesetze der Pharmakologie gestützt, und trifft auf alle untersuchten Arzneimittel zu.

Da sowohl Patienten als auch Ärzte einen Anstieg der Krankheitsaktivität (z. B. Schmerzen und Morgensteifigkeit) am Ende des Dosierungsintervalls aller Biologika bestätigen, sind wir der Meinung, dass dies patientenrelevant ist.

Eine entsprechende Abweichung des üblichen Zusammentreffens der Endpunkterhebung mit dem Ende des Applikationsintervalls könnte für neuere Studien und bei Arzneimitteln mit längeren Dosierungsintervallen relevant sein.

Änderungsantrag

Bei der Auswahl von RCTs zu Arzneimitteln mit langen Dosierungsintervallen sollte speziell auf ein Verzerrungspotential bezüglich der Wirksamkeit geachtet werden; dies gilt für den Zeitpunkt der Endpunkterhebung bezogen auf das Dosierungsintervall.

Kapitel 4.4 „Informationssynthese und -analyse“

Kommentierung

Es wurden viele Studien zu MCID bei RA durchgeführt. Für die meisten relevanten Endpunkte wurden diese MCIDs belegt. Einige MCIDs haben wir

bereits in den Kommentierungen zu Kapitel 4.1.3 benannt. Abgesehen davon sind mehrere andere Responsekriterien in der RA üblich.

Andere Methoden, wie die Anwendung von Hedges' g insbesondere im Bereich der unteren Grenze des methodisch fragwürdigen Konfidenzintervalls, können zu irreführenden Schlussfolgerungen führen. Dies trifft besonders für willkürlich gewählte Grenzwerte wie 0,2 SMD zu. Aus diesem Grund sollten sie vermieden werden.

Änderungsantrag

Es sollten ausschließlich Responsekriterien oder validierte MCIDs verwendet werden. Eine Auswertung der meisten relevanten Endpunkte kann auf diesen Kriterien basieren. Wenn eine solche Responderanalyse nicht möglich ist, sollte auf statistisch signifikante Mittelwertunterschiede als ausschlaggebendes Kriterium zurückgegriffen werden.

Bitte löschen Sie die beiden letzten Abschnitte des Kapitels 4.4.

Kapitel 4.4.1 „Gegenüberstellung der Ergebnisse der Einzelstudien“

Kommentierung

Der Nichtberücksichtigungsanteil von insgesamt 30 % ist willkürlich gewählt. Das Gleiche gilt für den Wert des maximalen Unterschieds von 15 % zwischen den Studienarmen. Hohe Patientenzahlen oder große Unterschiede zwischen Studienarmen können alternativ in dem Abschnitt zum Verzerrungsrisiko bei den Endpunkten erörtert werden.

Änderungsantrag

Dieses Unterkapitel sollte im Berichtsplan gestrichen werden.

Zitat

„Falls die für eine Meta-Analyse notwendigen Schätzer für Lage und Streuung in den Studienunterlagen nicht vorliegen, werden diese nach Möglichkeit aus den vorhandenen Informationen eigenständig berechnet beziehungsweise näherungsweise bestimmt.“

Kommentierung

Hier sollte spezifiziert werden, welche Methodik das IQWiG vorschlägt, um die Schätzer zu errechnen, und Beispiele sollten genannt werden.

Kapitel 4.4.4 „Subgruppenmerkmale und andere Effektmodifikatoren“

Zitat

„Subgruppenmerkmale und andere Effektmodifikatoren“

Kommentierung

Es gibt bisher keine Indizien, dass sich das Geschlecht auf die in Studien messbare Wirkung der Behandlung auswirkt.

Zusätzlich schlagen wir einige weitere Effektmodifikatoren vor, die ebenfalls berücksichtigt werden sollten:

- Unterschiede in der Krankheitsschwere zu Studienbeginn können auf Unterschiede in den Patientenpopulationen hindeuten und damit die Wahrscheinlichkeit des Ansprechens beeinflussen.
- Es hat sich in klinischen Studien gezeigt, dass die Krankheitsdauer ein relevanter Effektmodifikator ist (vgl. Jamal 2009 für radiographische Progression oder Nixon *et al.* 2007 für ACR-Response)

- Die Anzahl der zuvor erhaltenen DMARDs reflektiert sowohl die Krankheitsentwicklung als auch das Behandlungsschema der Vorbehandlung und kann sich auf die Verträglichkeit und die Wirksamkeit auswirken.
- Es sollte zwischen der Art der Kombinationstherapie unterschieden werden, d. h. ob MTX oder ein anderes sDMARD mit einem Biologikum kombiniert wird.
- Anwendung und Dosierung von Steroiden: Steroide gelten als sehr wirksame Antirheumatika. Daher kann die Anwendung von Steroiden zu Studienbeginn und die zeitgleiche Anwendung von Steroiden auf Patienteneigenschaften hinweisen (sehr hohe Krankheitsaktivität bei Studienbeginn). Sie kann sich außerdem entsprechend auf die klinische Wirksamkeit auswirken.
- Der Zeitpunkt der Messung der klinischen Wirksamkeit unter Berücksichtigung des Dosierungsintervalls des Arzneimittels (siehe unsere Kommentierungen zu Kapitel 4.3).

Änderungsantrag

Der Berichtsplan sollte um folgende Effektmodifikatoren ergänzt werden: der Schweregrad der Erkrankung zu Studienbeginn, die Krankheitsdauer, die Anzahl der bisher erhaltenen sDMARDs, die begleitende Medikation, die Anwendung von Steroiden und den Zeitpunkt der Messung der klinischen Aktivität unter Berücksichtigung des Dosierungsintervalls.

Bibliographie

- Aletaha D, Smolen J, Ward MM 2006: Measuring Function in Rheumatoid Arthritis. *Arthritis & Rheumatism* 54 (9): 2784-2792.
- Breedveld FC, Weisman MH, Kavanaugh AF, et al. 2006: A Multicenter, Randomized, Double-Blind Clinical Trial of Combination Therapy With Adalimumab Plus Methotrexate Versus Methotrexate Alone or Adalimumab Alone in Patients With Early, Aggressive Rheumatoid Arthritis Who Had Not Had Previous Methotrexate Treatment. *Arthritis & Rheumatism* 54 (1): 26-37.
- Bruynesteyn K, van der Heijde D, Boers, M, et al. 2002: Determination of the Minimal Clinically Important Difference in Rheumatoid Arthritis Joint Damage of the Sharp/van der Heijde and Larsen/Scott Scoring Methods by Clinical Experts and Comparison With the Smallest Detectable Difference. *Arthritis & Rheumatism* 46(4): 913-920.
- Bruynesteyn K, Boers M, Kostense P, et al. 2005: Deciding on progression of joint damage in paired films of individual patients: smallest detectable difference or change. *Ann Rheum Dis* 64: 179-182.
- Combe B, Landewe R, Lukas C, et al. 2007: EULAR recommendations for the management of early arthritis: report of a task force of the European Standing Committee for International Clinical Studies Including Therapeutics (ESCISIT). *Ann Rheum Dis* 66 (1): 34-45.
- Drossaers-Bakker KW, de Buck M, van Zeben D, et al. 1999: Long-term course and outcome of functional capacity in rheumatoid arthritis. *Arthritis & Rheumatism* 42(9): 1854-1860.
- European Medicines Agency. Points to consider on clinical investigation of medicinal products other than NSAIDs for treatment of rheumatoid arthritis [online]. 17.12.2003 [Zugriff: 07.10.2010]. URL: <http://www.ema.europa.eu/pdfs/human/ewp/055695en.pdf>.

- Felson DT, Anderson JJ, Boers M, et al. 1993: The American College of Rheumatology Preliminary Core Set of Disease Activity Measures for Rheumatoid Arthritis clinical Trials. *Arthritis & Rheumatism* 36 (6): 729-740.
- Furst DE, Schiff MH, Fleischmann RM, et al. 2003: Adalimumab, a fully Human Anti-Tumor Necrosis Factor- α Monoclonal Antibody, and Concomitant Standard Antirheumatic Therapy for the Treatment of Rheumatoid Arthritis: Results of STAR (Safety Trial of Adalimumab in Rheumatoid Arthritis). *J Rheumatol* 2563-2571.
- Jamal S, Patra K, Keystone EC 2009: Adalimumab response in patients with early versus established rheumatoid arthritis: DE019 randomized controlled trial subanalysis. *Clin Rheumatol* 28: 413-419.
- Kavanaugh A, Han C, Bala M 2004: Functional Status and radiographic Joint Damage Are Associated with Health Economic Outcomes in Patients with Rheumatoid Arthritis. *J Rheumatol* 31(5): 849-855.
- Krueger K, Manger B und die Kommission Pharmakotherapie der DGRh 2008: Leflunomid als Kombinationspartner von TNF Blockern bei Methotrexat-Unverträglichkeit [online] August 2008 [Zugriff 07.10.2010] URL: http://dgrh.de/fileadmin/media/Praxis_Klinik/Therapie-Empfehlungen/Leflunomide/leftnf_endfassung_110708.pdf
- Kosinski M, Zhao SZ, Dedhiya S, et al. 2000: Determining minimally important changes in generic and disease specific health related quality of life questionnaires in clinical trials of rheumatoid arthritis. *Arthritis & Rheumatism* 45 (7): 1478-1487.
- Larsen A 1975: How to apply Larsen Score in Evaluating Radiographs of Rheumatoid Arthritis in Longterm Studies? *J Rheumatol* 22 (10): 1974-1975.
- Nixon RM, Bansback N, Brennan A 2007: Using mixed treatment comparisons and meta-regression to perform indirect comparisons to estimate the efficacy of biologic treatments in rheumatoid arthritis. *Statist Med* 27: 1237-1254.

- Odegard S, Landewe R, Van der Heijde D, et al. 2006: Association of Early Radiographic Damage With Impaired Function in Rheumatoid Arthritis 54 (1): 68-75.
- Pinals RS, Masi, AT, Larsen RA, et al. 1981: Preliminary Criteria for Clinical Remission in Rheumatoid Arthritis. *Arthritis & Rheumatism* 24 (10):1308-1315.
- Scott DL, Pugner K, Kaarela K, et al. 2000: The links between joint damage and disability in rheumatoid arthritis. *Rheumatology* 39: 122-132.
- Sharp JT, Lidsky MD, Collins LC, et al. 1971: Methods of Scoring the Progression of Radiographic Changes in Rheumatoid Arthritis. *Arthritis & Rheumatism* 14 (6): 706-720.
- Smolen JS, Landewé R, Breedveld FC, et al. 2010: EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. *Ann Rheum Dis* 69 (6): 964-975.
- Van der Heijde DM, van Riel PL, Nuvér-Zwart IH, et al. 1989. Effects of hydroxychloroquine and sulphasalazine on progression of joint damage in rheumatoid arthritis. *Lancet* 13 (1):1036-1038.
- Van der Heijde DM, van 't Hof MA, van Riel PL, et al. 1990: Judging disease activity in clinical practice in rheumatoid arthritis: first step in the development of a disease activity score. *Ann Rheum Dis* 49: 916-20.
- Van der Heijde D 2000: Radiographic Imaging: the “gold standard” for assessment of disease progression in rheumatoid arthritis. *Rheumatology* 39 (supple. 1): 9-16.
- Van der Heijde D, Landewe R, van Vollenhoven R, et al. 2008: Level of radiographic damage and radiographic progression are determinants of physical function: a longitudinal analysis of the TEMPO trial. *Ann Rheum Dis* 67: 1267-1270.

- Van Vollenhoven RF, Cifaldi MA, Ray S, et al. 2010: Improvement in work place and household productivity for patients with early rheumatoid arthritis treated with adalimumab plus methotrexate: work outcomes and their correlations with clinical and radiographic measures from a randomized controlled trial companion study. *Arthritis Care & Research* 62 (2): 228-234.
- Wells GA, Tugwell P, Kraag GR, et al. 1993: Minimal important difference between patients with rheumatoid arthritis: the patient's perspective. *J Rheumatol* 20(3): 557-560.
- Welsing PM, van Gestel AM, Swinkels HL, et al. 2001: The relationship between disease activity, joint destruction, and functional capacity over the course of rheumatoid arthritis. *Arthritis & Rheumatism* 44 (9): 2009-2017.
- Wollenhaupt J, Alten R, Backhaus M, et al. 2009: Aktualisiertes Therapieschema der Rheumatoiden Arthritis. Ergebnisse eines Konsensprozesses deutscher Rheumatologen 2009. *Akt Rheumatol* 34: 234-239.

A 1.2 Bristol-Myers Squibb GmbH & Co. KGaA

Autoren:

Fenzau, Monika

May, Michael, Dr.

Neubauer, Aljoscha, Priv. Doz. Dr.

Bristol-Myers Squibb GmbH & Co. KGaA D-80632 München

Institut für Qualität und Wirtschaftlichkeit im
Gesundheitswesen

- Stellungnahme zum vorläufigen
Berichtsplan A10-01

Prof. Dr. med. Jürgen Windeler
Dillenburger Str. 27
51105 Köln

Ihr Ansprechpartner: [REDACTED]

Tel: [REDACTED]

Fax: [REDACTED]

E-Mail: [REDACTED]

Vorab per email: berichte@iqwig.de

München, den 26.10.2010

Auftragsnummer:	A10-01
Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie der rheumatoiden Arthritis	



Stellungnahme zum vorläufigen Berichtsplan A 10-01 „Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis“ (Version 1.0, Stand 20.09.2010, veröffentlicht am 28.09.2010)

Sehr geehrter Herr Prof. Windeler, Sehr geehrte Damen und Herren,

die Firma Bristol-Myers Squibb GmbH & Co. KGaA ist mit ihrem Produkt Abatacept an dem Verfahren A10-01 beteiligt und bedankt sich für die Möglichkeit der Kommentierung. Nach Prüfung Ihres vorläufigen Berichtsplans A10-01 dürfen wir wie folgt Stellung nehmen.

1. Anmerkungen zur projektspezifischen Methodik anhand der „PICO“ Kriterien

1.1. Klarheit der untersuchten Population

Auf Seite 4, Begriffe „Erstlinien-“ und „Zweitlinientherapie“ wird zu Recht auf den nicht konsistenten Gebrauch der Begriffe „Erstlinien-“ und „Zweitlinientherapie“ hingewiesen. Deshalb ist es wichtig klar zu definieren, wie diese Begriffe innerhalb dieses Projekts verwendet werden sollen und so die relevante Patientenpopulation/Indikation des Arzneimittels exakt zu definieren. Auf Seite 5 erscheint uns die getroffene Definition: „...der erste Einsatz des jeweils zu untersuchenden Arzneimittels gemäß Zulassungsstatus...“ als nicht ausreichend genau und darüber hinaus je nach Arzneimittel und Zulassungsstatus teilweise unterschiedlich. Dadurch besteht ein Verzerrungspotential, das sich ungünstigstenfalls bis in die Informationssynthese fortsetzt.

1.2. Intervention: Studiendauer

Die angegebene Studiendauer von mindestens 6 Monaten für die medikamentöse Intervention (Seite 8, Punkt „4.1.5 Studiendauer“) sehen wir grundsätzlich positiv. Für den Patienten ist bei einer chronischen Erkrankung unseres Erachtens der ausreichend langfristige Therapieerfolg von Bedeutung. Die vom IQWiG vorgegebene Mindestdauer für Studien von 6 Monaten (24 Wochen) entspricht auch den Anforderungen in den entsprechenden Leitlinien der europäischen und US-amerikanischen Zulassungsagenturen EMA und FDA für Zulassungsstudien, die vom IQWiG zitiert werden. Außerdem stimmen diese Anforderungen auch mit den EULAR Leitlinien überein, die darauf hinweisen, dass der Status der Krankheitsaktivität im 3. bis 6. Monat nach Therapiebeginn den langfristigen Therapieausgang vorwegnimmt. (Smolen et al 2010)

1.3. Endpunkte

Auf Seite 2 „Ziele der Behandlung“ wird hervorgehoben, dass Responsekriterien wie ACR20 lediglich auf eine geringe Teilminderung der Krankheitsaktivität hinweisen und hinsichtlich ihrer klinischen Relevanz nicht definiert seien. Dieser primäre Endpunkt wird jedoch von regulatorischer Seite gefordert und darf deshalb als solcher keinesfalls zu einem Ausschluss der jeweiligen Studie führen. Hier wäre mehr Konsistenz zu den Anforderungen der Zulassungsbehörden wünschenswert.

Weiter werden DAS/ACR-Remissionskriterien auf Seite 3 oben, als sehr harte Kriterien beschrieben. Hierbei sollte jedoch nicht außer Acht gelassen werden, dass je härter das Responsekriterium, umso geringer in der Regel die Ansprechraten. Wurde das jeweilige Kriterium nicht bei der Fallzahlplanung der Studie ausreichend berücksichtigt, so sind Analysen hierzu ggf. nicht von ausreichender statistischer Power. Entsprechend empfehlen wir an dieser Stelle und auf Seite 7, auch alle weiteren Kriterien mit einzubeziehen, die von klinischer Bedeutung sind, d.h. auch geringere Verbesserungen einzuschließen.

2. Anmerkungen zur projektspezifischen Methodik der systematischen Untersuchung

2.1. Transparenz der Entscheidungsfindung bei der Informationsbewertung

Zu Seite 10, Punkt „4.2 Informationsbeschaffung“, insbesondere Punkt „4.2.3 Selektion relevanter Studien“. Hier fehlt eine detailliertere Angabe zur Auswahl der 2 Reviewer (z.B. Erfahrungslevel in der Indikation). Das Cochrane-Handbuch (Higgins & Green, 2009, Abschnitt 7.2.4) gibt hierzu Empfehlungen. Aus Transparenzgründen sollte dieser Punkt offengelegt werden. Auch findet sich kein Hinweis, wie im Falle einer nicht durch Diskussion lösbaren unterschiedlichen Auffassung der beiden Reviewer dieser Konflikt gelöst wird (siehe z.B. Higgins & Green, 2009, Abschnitt 7.2.4). Auch hier ist aus Transparenzgründen auf ein Verfahren zu verweisen, analog zu den im Methodenpapier 3.0 genannten Dokumentationspunkten (6.1.5 Dokumentation, Seite 88, Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) Allgemeine Methoden Version 3.0, 2008). Durch den individuellen Hintergrund der Reviewer wird auch immer ein persönliches Werturteil in den Auswahl- und Bewertungsprozess einfließen. Dieses Werturteil so gering wie möglich zu halten wird durch eine objektive Methodik unterstützt. Um der verbleibenden Unsicherheit gerecht zu werden ist es empfehlenswert größt mögliche Transparenz zu gewährleisten.

Ein weiterer Punkt ist, dass im Berichtsplan die detaillierte Darstellung der Suchstrategie und Suchtermini fehlt. (Higgins & Green, 2009, Abschnitt 6.4) Aufgrund der Art des Ablaufs des IQWiG

Stellungnahme-Prozesses besteht hier so keinerlei Möglichkeit, zu diesem Punkt vor Abschluss des Vorberichts ergebnisoffen im Detail zu kommentieren.

2.2. Relevanz von Effektstärken

Das Konzept der individuellen *minimal important difference* [MID] für die Bemessung der Relevanz von Effektstärken wird auf Seite 12, „Punkt 4.4 Informationssynthese und –analyse“ angesprochen. Wir möchten klar darauf hinweisen, dass dieses im Prinzip wertvolle Konzept nicht die subjektive Bewertungskomponente der Frage „Was ist relevant?“ löst. Diesen Punkt hat Herr Prof. Windeler erst kürzlich in einem Vortrag zur Relevanz von Effektstärken am 26.1.2010 klar angesprochen (Windeler & Thomas, 2010). Der Schlussfolgerung auf Folie 18: „basierend auf einem Werturteil“ können wir uns so voll anschließen. Deshalb ist je nach gewählter Methodik diese zu validieren und die getroffenen Werturteile für alle Beteiligten transparent zu machen. So empfehlen wir die getroffenen Bewertungen auch durch einen wissenschaftlichen Expertenrat mit ausreichender Fachexpertise prüfen zu lassen.

2.3. Umgang mit Studienabbrecherquoten

Der Punkt hierzu (Seite 13, „4.4.1 Gegenüberstellung der Ergebnisse der Einzelstudien“) ist so nicht einfach zu verstehen und es sollte über eine Reformulierung nachgedacht werden um mögliche Unklarheiten zu beseitigen. Nach unserer Interpretation handelt es sich hierbei um den Umgang mit den Drop-out-Raten innerhalb einer Studie. Wir können uns der zitierten Literatur und Argumentation anschließen, welcher Prozent Wert nun als relevant zur Studienauswahl in dieser Untersuchung sein wird, wird jedoch nicht klar festgestellt. Wünschenswert wäre es insbesondere, wenn eine Klarstellung des Zeitraums aufgenommen werden würde, an dem die Drop-out-Rate nicht überschritten sein darf. Im Langzeitverlauf einer Studie steigt naturgemäß die Drop-out-Rate an. Eine Definition der Messzeitpunkte (z.B. Kriterium bei 24 Wochen) würde vermeiden, dass relevante Langzeitstudienresultate aufgrund dieses Kriteriums ausgeschlossen werden. Da eine langfristige Krankheitskontrolle von essentieller Bedeutung für die Patienten ist, erscheint uns ein potentieller Ausschluss von Langzeitdaten aufgrund naturgemäß zunehmender Drop-out-Raten nicht im Sinne des Patienten.

Weiterhin wird ausgeführt, dass alle Studien mit einem Unterschied von größer als 15 Prozentpunkten in den Nichtberücksichtigungsanteilen zwischen den Gruppen ausgeschlossen werden. Das erscheint schlüssig, dennoch wäre eine Begründung (z.B. Referenz) hilfreich.

2.4. Informationssynthese mittels Meta-Analyse

Auf Seite 13, „4.4.2 – Meta-Analysen“ wird darauf hingewiesen, dass für die statistische Auswertung primär die Ergebnisse aus Intention-to-Treat-Analysen, verwendet werden sollen und die Meta-Analysen in der Regel auf Basis von Modellen mit zufälligen Effekten erfolgen sollen. Außerdem sollen nur in begründeten Ausnahmefällen Modelle mit festen Effekten einsetzen.

Hierzu möchten wir anmerken, dass neben dem verfolgten frequentistischen Ansatz zunehmend Bayesianische Ansätze umgesetzt werden, da sie näher an der letztlich geforderten Entscheidungsunterstützung sind. Zur Diskussion um Methodenauswahl der Meta-Analyse möchten wir auf die Ihnen im Rahmen der Datenlieferung vorgelegten, noch unveröffentlichten Meta-Analyse von Mapi-Values verweisen (Mapi Values, 2010):

“Meta-analysis can be performed with traditional frequentist statistics or with Bayesian statistics. With a frequentist approach the result of the meta-analysis is a point estimate along with a (95%) confidence interval. It has to be noted that confidence intervals obtained with a frequentist approach cannot be interpreted in terms of probabilities. The 95% confidence interval does not mean there is 95% probability that the ‘true’ or population value is between the boundaries of the interval. The 95% confidence interval (CI) for the ‘true’ value is the interval which if the experiment were to be repeated on many occasions, would contain the ‘true’ value on 95% of occasions. Consequently, the frequentist approach dichotomises the results according to either ‘significant’ or ‘non significant’, based on the chosen p-value. For decision making this approach is not well suited as it does not indicate to what extent (probability) a hypothesis will be true or false.

Bayesian methods involve a formal combination of initial beliefs of the possible values of the pooled effect (prior probability distribution) with information based on the observed data (the likelihood distribution, which is comparable to a 95% CI of the frequentist output). The likelihood distribution is used to update the prior information, expressed as a probability distribution, to obtain a posterior probability distribution of the pooled effect. This posterior probability distribution attaches a probability to every possible value of the pooled effect. This means that, for decision making, a Bayesian analysis allows one to state to what extent a hypothesis may be true or false. For example: “There is an x% probability that treatment A results in a larger cholesterol reduction than treatment B”. This is in contrast to the findings with a conventional frequentist approach. In Bayesian statistics the 95% credible interval indicates the probability that the ‘true’ value lies within the interval and is thus different from a frequentist confidence interval.

The Bayesian approach naturally leads into a decision framework to support decision-making. In order not to influence the observed results by the prior distribution, an often heard critique on the Bayesian approach, a non-informative prior distribution is used for the pooled treatment effect. With such a ‘flat’ prior it is assumed that before seeing the data any value for the pooled effect is equally likely to occur. As a consequence, posterior results are not influenced by the prior distribution but totally driven by the data as with a conventional frequentist meta-analysis“.

Hinsichtlich weiterer statistischer Diskussionspunkten, insbesondere zum Thema Hedges’g und der Diskussion um die Grenze von 0,2 verweisen wir auf das Ihnen bekannte VFA-Positionspapier (VFA, 2008).

2.5. Effektmodifikatoren

Auf Seite 14, Punkt „4.4.4 Subgruppenmerkmale und andere Effektmodifikatoren“ ist bisher vorgesehen, Geschlecht, Alter und Art der Vorbehandlung als mögliche Effektmodifikatoren in die Analysen einzubeziehen. Hier sollten unbedingt noch folgende klinische Faktoren aufgenommen werden, die potentiell ebenfalls die Effekte beeinflussen können:

- Schweregrad der rheumatoiden Arthritis
- Krankheitsaktivität zum Zeitpunkt der Erhebung der Baseline Daten
- Krankheitsdauer
- Begleittherapie
- Komorbiditäten

Literatur

Bristol-Myers Squibb GmbH & Co. KGaA. Fachinformation ORENCIA 250 mg Pulver. Juli 2010

Higgins JPT, Green S (editors). Cochrane Handbook for Systematic Reviews of Interventions Version 5.0.2 [updated September 2009]. The Cochrane Collaboration, 2009. Available from www.cochrane-handbook.org.

Ludwig WD. Wie berücksichtigt und bewertet die AkdÄ unerwünschte Arzneimittelwirkungen im Rahmen der Evidenz-basierten Therapieempfehlungen. 7. Diskussionsforum 2010 Der Sozialmedizinischen Experten-Gruppe „Arzneimittelversorgung“ (SEG 6), Dortmund, 14.09.2010

Mapi Values. Systematic Literature Review and Mixed Treatment Comparison of Abatacept in the Management of Patients with Rheumatoid Arthritis who Inadequately Respond to Methotrexate - BS6186a MTC Report v6.0. September, 2010

Müller-Oerlinghausen B; Tiaden J.D. Arzneimittelkommission der deutschen Ärzteschaft - 50 Jahre im Dienste der Arzneimittelsicherheit in Deutschland. Der Internist 2002;43:482–492

Singh JA et al. A network meta-analysis of randomized controlled trials of biologics for rheumatoid arthritis: a Cochrane overview. CMAJ 2009;181(11);787-796

Smolen JS et al. EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. Ann Rheum Dis 2010;69:964-975.

VFA-Positionspapier „Anforderungen an die isolierte Nutzenbewertung“ 24.11.2008; Available from www.vfa.de

Windeler J, Thomas S. Relevanz von Effektstärken. Vortrag in Berlin am 26.01.2010; Available from www.iqwig.de

3. Angemessene Informationsbasis

3.1. Veraltete Fachinformation für Abatacept

Die angegebene Referenz zur Fachinformation von Abatacept auf Seite 18 als Referenz 32 ist veraltet. Anbei erhalten Sie die aktuellste Version der Fachinformation vom Juli 2010, die zur Indikationserweiterung verfasst wurde (Bristol-Myers Squibb, 2010).

3.2. Fokussierung auf „Nutzen“ im Berichtsplan, fehlende bzw. unscharfe Definitionen zum „Schaden“

Während im Berichtsplan die Bewertung des Nutzens relativ detailliert beschrieben wird, wird auf die Bewertung des Schadens nicht eingegangen. Der Begriff des Schadens, nähere Angaben zum Prozess oder den verwendeten Datenquellen zum Schaden bleiben weitestgehend undefiniert. Unsere Empfehlung wäre deshalb, diesen Punkt klarer vorab im Berichtsplan zu definieren. (Müller-Oerlinghausen & Tiaden 2002; Ludwig 2010)

In der Ihnen zur Verfügung gestellten Meta-Analyse des Cochrane Instituts (Singh et al 2009) wird das Schadenspotential beispielsweise über die Studienabbrüche auf Grund von Nebenwirkungen definiert. Diese Definition wird auch in dem unveröffentlichen Bericht von Mapi-Values verwendet (Mapi-Values 2010), den wir Ihnen ebenfalls im Rahmen der Datenlieferung zur Verfügung gestellt haben. Hier werden die Studienabbrüche auf Grund von Nebenwirkungen zum Zeitpunkt 24/26 Wochen für alle Vergleichsmedikamente untersucht. Den Schaden vorab zu definieren und das Schadenspotential angebracht zu analysieren, beispielsweise anhand der Studienabbrüche durch Nebenwirkungen, liegt im besonderen Interesse der Patienten.

Für Rückfragen oder eine wissenschaftliche Erörterung stehen wir Ihnen jederzeit gerne zur Verfügung.

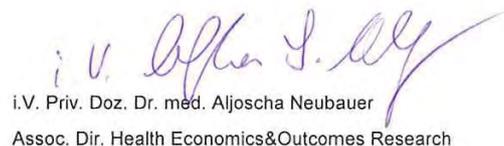
Mit freundlichen Grüßen
Bristol-Myers Squibb GmbH & Co. KGaA



ppa. Monika Fenzau
Director Market Access



i.V. Dr. Michael May
Medical Director Immunoscience&Virology



i.V. Priv. Doz. Dr. med. Aljoscha Neubauer
Assoc. Dir. Health Economics&Outcomes Research

A 1.3 Deutsche Gesellschaft für Rheumatologie

Autoren:

Krüger, Klaus, Prof. Dr.

Stellungnahme der Deutschen Gesellschaft für Rheumatologie zum vorläufigen Berichtsplan "Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis"

Nach Meinung der Deutschen Gesellschaft für Rheumatologie wird der vorläufige Berichtsplan zum Auftrag A 10-01 des IQWiG "Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis" dem Thema sehr gut gerecht und beinhaltet die für die Bearbeitung wichtigen Aspekte. Aus unserer Sicht gibt es an wenigen Stellen Interpretations- oder Ergänzungsbedarf. Auf diese Stellen wird nachfolgend in der Reihenfolge, in der sie im Manuskript Erwähnung finden, eingegangen.

1) Auf Seite 5 wird auf die Definition des Begriffes Zweitlinientherapie eingegangen. Aus der Formulierung "...Einsatz von biotechnologisch hergestellten Arzneimitteln bei Personen, die mit einem krankheitsmodifizierenden Antirheumatikum vorbehandelt sind..." geht nicht klar hervor, ob eine Bewertung ausschließlich von Zweitlinientherapien im eigentlichen Sinne des Wortes (z.B. Biologikum als zweiter Schritt im Algorithmus nach Methotrexat mono) vorgenommen werden soll, oder ob unter vorausgegangener Erstlinientherapie auch mehrere vorausgegangene Therapieschritte mit krankheitsmodifizierenden Antirheumatika (sDMARD) verstanden werden können (z.B. Biologikum als 3.Schritt nach Schritt 1 Methotrexat mono und Schritt 2 Methotrexat in Kombination mit anderen sDMARD - dieses Vorgehen wird in der Regel in der Praxis gewählt). Es sollte präzisiert werden, ob Möglichkeit 1 oder 2 (oder beide) gemeint sind.

2) Auf Seite 7 unten werden mit Recht strukturelle Gelenkveränderungen unter den patientenrelevanten Endpunkten aufgeführt. Die in Klammer genannten Parameter Deformitäten, Versteifungen und Kontrakturen allerdings werden in RCT nicht direkt erfasst, mehr noch, es existieren keine validierten Instrumente, um diese Parameter direkt zu erfassen. Die messtechnische Erfassung von Gelenkveränderungen ist hingegen radiologisch möglich (z.B. Sharp-Score mit den Komponenten "Joint Space Narrowing" als Parameter der Knorpelzerstörung und "Erosion Score" als Parameter von knöchernen Destruktionen). Der radiologische Verlauf ist auch in fast allen RCT als sekundärer oder sogar primärer Endpunkt enthalten (1 - 5 als Beispiele). Er sollte aus unserer Sicht deshalb in die Bewertung eingehen.

Indirekt schlagen sich die strukturellen Gelenkveränderungen im Funktionsstatus (gemessen meist mittels HAQ) als irreversible Komponente nieder, die sogar im Fall einer klinischen Remission zu einer Verschlechterung des HAQ führen kann. Dies konnte in einer Sekundärauswertung von sechs RCT gezeigt werden (6). Der Versuch, einen direkten Bezug zwischen Sharp-Score und HAQ herzustellen, wurde kürzlich in einer weiteren Studie mit sekundärer Auswertung zahlreicher RCT gemacht und lieferte eine numerische Beziehung (7). Zusammengefasst liefert in RCT nur die radiologische Untersuchung direkte und der HAQ indirekte Hinweise auf die als patienten-relevant identifizierten strukturellen Gelenkveränderungen, beide Messgrößen (welche sogar messtechnisch in Bezug zueinander gesetzt werden können) sollten daher Eingang in die Bewertung finden.

3) Mit Recht ist die Gesamtmortalität auf S.8 als ein ausserordentlich patientenrelevanter Endpunkt aufgeführt. Aufgrund der limitierten Studiendauer von maximal zwei Jahren werden hier jedoch mit dem Instrument des RCT keine aussagekräftigen Daten zu erheben sein. Behelfsweise müssten für diesen Parameter Kohortenstudien mit längeren Verläufen herangezogen werden (Beispiele 8 - 10). Eine weitere Auswertung, die u.a. die Mortalitätsraten thematisiert, fußt auf einer Analyse von 36 mit Adalimumab durchgeführten Studien, darunter sämtlichen relevanten RCT (11).

4) In Seite 8 unten wird klargestellt, daß ausschließlich RCT in die Nutzenbewertung einfließen. Aus unserer Sicht ist die Bewertung des Therapie-Risikos unabdingbarer Teil der Nutzenbewertung. Zwar sind unerwünschte Arzneimittelwirkungen unter den patientenrelevanten Endpunkten aufgelistet. Einige im Zusammenhang mit dem Biologika-Einsatz von Anfang als potenziell relevante Risiken, insbesondere Infektionen und Malignom-Induktion, sind mit dem Instrument des RCT aus zwei Gründen nur teilweise (Infektionsrisiko) oder gar nicht (Malignom-Risiko) erfassbar: Zum einen erstrecken sich RCT nur maximal über zwei Jahre, zum anderen ist in ihnen eine Patienten-Population vertreten, die sich (z.B. aufgrund von Studien-Ausschlusskriterien) von den Patienten deutlich unterscheidet, die in der Praxis behandelt werden (12).

Im Falle der Biologika wurden von Beginn der Anwendung an große Langzeit-Register geführt (z.B. in Deutschland RABBIT-Register), um unter Praxisbedingungen die Langzeit-Sicherheit dieser Substanzen zu verfolgen. In diesen Registern sind zum gegenwärtigen Zeitpunkt in großer Zahl Therapien mit einer Dauer von z.T. mehr als zehn Jahren erfasst (8, 13 - 15). In Schweden besitzt das entsprechende Register (12) noch dazu eine hervorragende Verknüpfung mit nationalen Datenbanken, z.B. mit einem Malignom-Register. Es wird dringend empfohlen, zur Nutzenbewertung unter dem Aspekt der Langzeitsicherheit auf dieses einmalige Reservoir an umfangreichen Daten zurückzugreifen - hier bieten RCT keinerlei Alternative.

5) In die Liste der Subgruppenmerkmale und weiteren Effektmodifikatoren (Seite 14 unten) sollten als bei Vergleich verschiedener Studien relevante Confounder die Ausgangsaktivität der Krankheit (z.B. Ausgangs-DAS 28), die Krankheitsdauer und der Serostatus (Anteil der RF- bzw. Anti-CCP-positiven Patienten) mit aufgenommen werden, um eine bessere Vergleichbarkeit zu gewährleisten.

Prof.Dr.Klaus Krüger
Sprecher der Kommission Pharmakotherapie der DGRh
für die Deutsche Gesellschaft für Rheumatologie

Literatur:

- (1) St Clair EW, van der Heijde DM, Smolen JS, Maini RN, Bathon JM, Emery P et al. Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial. *Arthritis Rheum* 2004;50(11):3432-3443.

- (2) Emery P, Breedveld FC, Hall S, Durez P, Chang DJ, Robertson D et al. Comparison of methotrexate monotherapy with a combination of methotrexate and etanercept in active, early, moderate to severe rheumatoid arthritis (COMET): a randomised, double-blind, parallel treatment trial. *Lancet* 2008;372(9636):375-382.
- (3) Lipsky PE, van der Heijde DM, St Clair EW, Furst DE, Breedveld FC, Kalden JR et al. Infliximab and methotrexate in the treatment of rheumatoid arthritis. Anti-Tumor Necrosis Factor Trial in Rheumatoid Arthritis with Concomitant Therapy Study Group. *N Engl J Med* 2000;343(22):1594-1602.
- (4) Klareskog L, van der Heijde D, de Jager JP, Gough A, Kalden J, Malaise M et al. Therapeutic effect of the combination of etanercept and methotrexate compared with each treatment alone in patients with rheumatoid arthritis: double-blind randomised controlled trial. *Lancet* 2004;363(9410):675-681
- (5) Tak PP, Rigby WF, Rubbert-Roth A, Peterfy CG, van Vollenhoven RF, Stohl W et al (2010) Inhibition of joint damage and improved clinical outcomes with rituximab plus methotrexate in early active rheumatoid arthritis: the IMAGE trial. *Ann Rheum Dis* 2010 Oct 11 Epub ahead of printing
- (6) Aletaha D, Smolen J, Ward MM (2006) Measuring Function in Rheumatoid Arthritis. Identifying Reversible and Irreversible Components. *Arthritis Rheum* 54: 2784 - 2792.
- (7) Smolen JS, Aletaha D, Grisar JC et al. (2010) Estimation of a numerical value for joint damage-related physical disability in rheumatoid arthritis clinical trials. *Ann Rheum Dis* 69: 1058 - 1064.
- (8) Carmona L, Descalzo MA, Perez-Pampin E et al. (2007) All-cause and cause-specific mortality in rheumatoid arthritis are not greater than expected when treated with tumour necrosis factor antagonists. *Ann Rheum Dis* 66: 880 - 885.
- (9) Jacobsson LT, Turesson C, Nilsson JA et al. (2007) Treatment with TNF blockers and mortality risk in patients with rheumatoid arthritis. *Ann Rheum Dis* 66: 670 - 675.
- (10) Lunt M, Watson KD, Dixon WG, Symmons DP, Hyrich KL. (2010) No evidence of association between anti-TNF treatment and mortality in patients with rheumatoid arthritis: Results from the British Society for Rheumatology Biologics Register. *Arthritis Rheum* 2010. Jul 26. [Epub ahead of print]
- (11) Burmester GR, P Mease, B A C Dijkmans, K Gordon, D Lovell, R Panaccione et al. (2009) Adalimumab safety and mortality rates from global clinical trials of six immune-mediated inflammatory diseases. *Ann Rheum Dis* 68: 1863-1869.
- (12) Zink A, Strangfeld A, Schneider M, Herzer P, Hierse F, Stoyanova-Scholz M et al. Effectiveness of tumor necrosis factor inhibitors in rheumatoid arthritis in an observational cohort study: comparison of patients according to their eligibility for major randomized clinical trials. *Arthritis Rheum* 2006;54(11):3399-3407.

- (13) Askling J, van Vollenhoven RF, Granath F, Raaschou P, Fored CM, Baecklund E et al. (2009) Cancer risk in patients with rheumatoid arthritis treated with anti-tumor necrosis factor alpha therapies: does the risk change with the time since start of treatment? *Arthritis Rheum* 60:3180-9.
- (14) Dixon WG, Watson KD, Lunt M, Mercer LK, BSRBR Control Consortium, Hyrich KL et al. (2010) The influence of anti-TNF therapy upon cancer incidence in patients with rheumatoid arthritis who have had a prior malignancy: Results from the BSR biologics register. *Arthritis Care Res* 62: 755-63.
- (15) Strangfeld A, Hierse F, Rau R, Burmester GR, Krummel-Lorenz B, Demary W et al. (2010) Risk of incident or recurrent malignancies among patients with rheumatoid arthritis exposed to biologic therapy in the German biologics register RABBIT. *Arthritis Res Ther* 12(1):R5.

A 1.4 Deutsche Rheuma-Liga Bundesverband e. V.

Autoren:

Faubel, Ursula

**Stellungnahme der Deutschen Rheuma-Liga Bundesverband
zum vorläufigen Berichtsplan
zum Auftrag A10-01: Biotechnologisch hergestellte Arzneimittel in
der Zweitlinientherapie der rheumatischen Erkrankungen**

Die Deutsche Rheuma-Liga bittet in folgenden Punkten um Überarbeitung des Berichtsplans:

4.1.3 Patientenrelevante Endpunkte

Unter 4.1.3 werden als patientenrelevante Endpunkte Remission, Symptomatik der Entzündung, Strukturelle Gelenkveränderungen, körperlicher Funktionsstatus, soziales Funktionsniveau, gesundheitsbezogene Lebensqualität, Gesamtmortalität und unerwünschte Arzneimittelwirkungen definiert. Hinsichtlich des interventionsbedingten Aufwands und der Patientenzufriedenheit wird festgestellt, dass dieser ergänzend berücksichtigt werden soll, aber ein Nutzen oder Zusatznutzen sich allein auf dieser Basis nicht ergeben kann.

In der Anhörung zu den patientenrelevanten Endpunkten wurde von den Patientenvertretern der Deutschen Rheuma-Liga ausführlich zur Frage der patientenrelevanten Endpunkte Stellung genommen. Die dort dargestellten Punkte werden im Berichtsplan wenig präzise und unvollständig wiedergegeben. Wir bitten daher zu den im Folgenden benannten Punkten um Ergänzung:

Remission: Remission wird im vorläufigen Berichtsplan nicht definiert. Somit ergibt sich die Frage, welche Definition für Remission bei Rheumatoider Arthritis verwandt wird (Definition nach Pinals et al. oder die Definition nach komplexen Scores wie ACR70 oder DAS28<2,6; Übersicht dazu: Smolen JS, Aletaha D; Bijlma J W J et al. Treating rheumatoid arthritis to target: recommendations of an international task force. *Ann Rheum Dis* 2010;69:631). Weiterhin wäre zu definieren, dass es sich um eine anhaltende Remission handelt.

Die Patientenvertreter der Deutschen Rheuma-Liga haben aber bereits in der Anhörung darauf hingewiesen, dass für viele Patienten das Ziel "Remission" nicht mehr erreichbar ist, da bereits Gelenkzerstörungen eingetreten sind. Hier ist aus Betroffenen-Sicht das Ziel der Behandlung eine niedrige Krankheitsaktivität, analog DAS28<3,2.

Symptomatik der Entzündung: Hier werden lediglich "Schmerz, Fatigue und Morgensteifigkeit" angegeben. Die den Patienten erheblich belastende Gelenkschwellung und -überwärmung wird nicht erwähnt. Der Begriff "Fatigue" gibt nur unscharf das wieder, was die Patientenvertreter benannt haben, nämlich Abgeschlagenheit, Erschöpfung, Müdigkeit und Schlaflosigkeit. Die Besserung der Schlaflosigkeit ist für viele Betroffene ein ganz wichtiger Endpunkt, insbesondere in Phasen hoher entzündlicher Aktivität. Auch sind diese Beschwerden nicht ausschließlich Ausdruck der Entzündung sondern auch ohne wesentliche

Ausprägung der entzündlichen Kardinalsymptome oftmals Folge der systemischen immunologischen Aktivität.

Die Beseitigung des Schmerzes ist aus Sicht der Patienten ein ganz wesentlicher Endpunkt, der durch die Therapie erreicht werden soll. Rheumakranke Menschen leiden dauerhaft unter Schmerzen, die durch die Biologika-Therapie in vielen Fällen ganz wesentlich gelindert werden. Die Betroffenen haben dadurch die Möglichkeit, ihrem täglichen Leben mit geringerer Schmerzbelastung nachzugehen und auch besser schlafen zu können. Die Reduktion von Schmerzen führt zu einem erheblichen Gewinn an Lebensqualität und Leistungsfähigkeit. Der Schmerz ist nicht allein an die Entzündungsaktivität gekoppelt und sollte daher als eigenständiger patientenrelevanter Endpunkt aufgeführt werden.

Strukturelle Gelenkveränderungen: Hier werden im vorläufigen Berichtsplan Deformitäten, Versteifungen und Kontrakturen benannt. Von den Patientenvertretern wurde außerdem auf das wichtige Symptom der Instabilität hingewiesen.

Soziales Funktionsniveau: Die im vorläufigen Berichtsplan benannte Definition „Teilhabe am beruflichen und sozialen Leben“ ist aus Sicht der Patientenvertreter zu unscharf. Beim sozialen Funktionsniveau sollten berücksichtigt werden: Arbeitsunfähigkeitstage bzw. Krankenhausaufenthalte, Minderung der Erwerbsfähigkeit, Pflegebedürftigkeit.

Gesundheitsbezogene Lebensqualität: Bei der gesundheitsbezogenen Lebensqualität sollte neben dem Wohlbefinden auch Depression bzw. depressive Verstimmung als Merkmale herangezogen werden.

Von den Patientenvertretern der Deutschen Rheuma-Liga wurde außerdem die **einfache und patientenfreundliche Anwendung, Einnahme und Handhabung** als patientenrelevanter Endpunkte benannt. Insbesondere für berufstätige Betroffene kann von einer nicht zu zeitaufwendigen Einnahmeform abhängen, ob sie ihrer beruflichen Tätigkeit weiter nachgehen können. Die Form der Anwendung ist daher von hoher Bedeutung für die Betroffenen. § 35 Abs. 1b SGB V benennt als Kriterium für die Nutzenbewertung ausdrücklich die Lebensqualität der Betroffenen. Zur Lebensqualität gehört auch, ob die Therapie eine soziale und berufliche Teilhabe ermöglicht. Der interventionsbedingte Aufwand für die Betroffenen sollte nach Auffassung der Deutschen Rheuma-Liga nicht als zweitrangig bewertet werden und stellt ein relevantes Kriterium in der Nutzenbewertung dar, das entsprechend als patientenrelevanter Endpunkt berücksichtigt werden muss.

Insgesamt ist unklar, wie die benannten patientenrelevanten Endpunkte gewichtet werden. Für Patienten sind Fragen des Schmerzes, der entzündlichen Aktivität und der Funktion von enormer Bedeutung, da sich hieran Lebensqualität und Teilhabe am gesellschaftlichen und beruflichen Leben oft entscheiden.

4.1.4 Studientypen

Unter 4.1.4. Studientypen wird ausgeführt, dass ausschließlich Randomisierte kontrollierte Studien (RCTs) in die Bewertung einbezogen werden und dass diese die zuverlässigsten Ergebnisse für die Bewertung des Nutzens einer medizinischen

Intervention lieferten. Unter 4.1.3. Patientenrelevante Endpunkte wird ausgeführt, dass als patientenrelevante Endpunkte auch „Unerwünschte Arzneimittelwirkungen“ verwandt werden sollen.

Aus Sicht der Deutschen Rheuma-Liga reichen RCTs nicht aus, den Parameter „Unerwünschte Arzneimittelwirkungen“ exakt abzubilden. Hierzu tragen zwei Beschränkungen der RCTs bei: Zum einen ist die Laufzeit der RCTs häufig auf 6 Monate bis ein Jahr begrenzt, so dass mittel- und langfristig auftretende Nebenwirkungen nicht erfasst werden. Vor allem aber werden zahlreiche Patienten, die in der Praxis Biologika erhalten, wegen Ein- bzw. Ausschlusskriterien nicht in RCTs eingeschlossen. Starke Funktionseinschränkung, höheres Lebensalter und Komorbidität führen zum Beispiel zum Nicht-Einschluss in RCTs. In der Versorgungspraxis erhalten aber eine Vielzahl von Patienten mit starken Funktionseinschränkungen, höherem Lebensalter und Komorbiditäten die Biologika.

In Europa wurden in Großbritannien, Schweden, Dänemark, Spanien und Norwegen nationale Register eingerichtet, um gerade den genannten Parameter Arzneimittelsicherheit besser abzubilden. Zwischen der Studienpopulation der RCTs und den Populationen in der Praxis, wie sie durch die Studienregister abgebildet werden, bestehen gravierende Unterschiede. Im Deutschen Biologika-Register RABBIT (Rheumatoid Arthritis OBservation of Biologic Therapy) wurden bislang aus über 300 teilnehmenden rheumatologischen Einrichtungen mehr als 7 500 Patienten eingeschlossen, die bis zu 10 Jahre lang beobachtet werden. Nur 21-33 % der Patienten unter Biologika in RABBIT erfüllen in Deutschland bei Therapiebeginn die Einschlusskriterien der jeweiligen Zulassungsstudien (Zink A, Strangfeld A, Schneider M et al. Effectiveness of Tumor Necrosis Factor Inhibitors in: Rheumatoid Arthritis in an Observational Cohort Study. Comparison of Patients According to Their Eligibility for Major Randomized Clinical Trials. ARTHRITIS & RHEUMATISM. Vol. 54, No. 11, November 2006, pp 3399–3407).

Da die Population der RCTs und der Registerpopulation nicht übereinstimmen, ergeben sich vor allem im Hinblick auf die Daten zu unerwünschten Arzneimittelwirkungen Unterschiede. Während beispielsweise in den RCTs die Infektionsraten unter Biologika-Therapie nicht immer statistisch signifikant erhöht sind, ist das in den Registerdaten durchaus der Fall (Listing J, Strangfeld A, Sonja Kary S et al. Infections in Patients With Rheumatoid Arthritis Treated With Biologic Agents. ARTHRITIS & RHEUMATISM. Vol. 52, No. 11, November 2005, pp 3403–3412; Curtis JR, Jain A, Askling J, Bridges SL Jr et al.. A comparison of patient characteristics and outcomes in selected European and U.S. rheumatoid arthritis registries. Semin Arthritis Rheum. 2010 Aug;40(1):2-14.).

Die RCTs spiegeln aufgrund der strengen Einschlusskriterien die Realität der Arzneimittelversorgung in der Praxis in Hinblick auf die Nebenwirkungen nicht ausreichend wider. Nach Auffassung der Deutschen Rheuma-Liga müssen die internationalen Registerdaten daher in die Analyse einbezogen werden.

Stand: 26.10.2010

A 1.5 MSD Sharp & Dohme GmbH

Autoren:

Henkel, Katrin, Dr.

Krobot, Karl J., Dr. PhD

Peceny, Markus, Dr.

**Berichtsplan A10-01: „Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie bei der rheumatoiden Arthritis“
Stellungnahme der Firma MSD Sharp & Dohme GmbH***

* Wir nehmen Bezug auf unsere email vom 08.10.2010, in der wir bestätigen, dass MSD und Herr Dr. Dr. Karl J. Krobot nunmehr Kontaktperson und Ihr Ansprechpartner zum Projekt A10-01 und den Substanzen Gollimumab (Simponi®) und Infliximab (Remicade®) sind.

Zusammenfassung

Verbesserungsbedarf sehen wir insbesondere bezüglich konkreter methodischer Vorabfestlegungen (Instrumente, kombinierte Endpunkte, indirekte Vergleiche). Solche methodischen Vorabfestlegungen sind unseres Ermessens unverzichtbarer Bestandteil wissenschaftlichen Arbeitens. Ihre Fazitrelevanz gebietet zudem ihre Festlegung bereits auf der Ebene des vorläufigen Berichtsplans, gefolgt von einem Aufruf zur Stellungnahme per IQWiG-Info-Dienst und IQWiG-Homepage. Erst dann sehen wir die Voraussetzungen für eine spezifische Einsichtnahme in die Datenlage der zu bewertenden Produkte als geschaffen.

Weitere Stellungnahmepunkte betreffen die Konkretisierung des Begriffes „Zweitlinientherapie“, die Berücksichtigung der EMA-Vorgaben bezüglich des Studiendesigns, das Vorgehen im Falle wesentlicher methodischer Änderungen im Laufe der Vorberichtserstellung, sowie die geänderte Rechtsgrundlage für Nutzenbewertungen.

Unsere Stellungnahmepunkte im Einzelnen:

1. Konkretisierung des Begriffes „Zweitlinientherapie“

Mit Zweitlinientherapie wird im vorläufigen Berichtsplan auf Seite 5 *„...der Einsatz von biotechnologisch hergestellten Arzneimitteln bei Personen, die mit einem krankheitsmodifizierenden Antirheumatikum vorbehandelt sind...“* bezeichnet. Da die meisten Studien nicht auf Patienten mit genau einer Vorbehandlung beschränkt sind, sollte die entsprechende Formulierung in *„...mit mindestens einem krankheitsmodifizierenden Antirheumatikum ...“* geändert werden.

2. Berücksichtigung der EMA-Vorgaben bezüglich des Studiendesigns

Sowohl die European Medicines Agency (EMA) als auch die Food and Drug Administration (FDA) empfehlen, anti-entzündliche Effekte einer Therapie der rheumatoiden Arthritis über einen Zeitraum von 24 Wochen zu evaluieren^{1,2}.

Eine dauerhafte Behandlung der rheumatoiden Arthritis mit einer nicht ausreichend wirksamen Therapie ist allerdings aus ethischen Gründen nicht angezeigt. Deshalb empfiehlt die EMA-Leitlinie die Dauer der Placebothherapie auf 3-6 Monate zu begrenzen und eine klare Regelung für den Abbruch der Placebothherapie zu definieren: *"Since it would be unethical to retain a patient with rheumatoid arthritis on placebo treatment indefinitely, the duration of placebo must necessarily be limited. Depending on the severity and the activity of the disease three to six months is acceptable. For ethical reasons it is recommended to provide predefined rules for withdrawal from placebo."*

Um diesem ethischen Anspruch gerecht zu werden, enthalten die Zulassungsstudien der neuen Biologika in der Regel die Möglichkeit einer vorzeitigen Therapieumstellung ("Early Escape" nach 3-6 Monaten). Andere Studien beinhalten eine geplante Umstellung von Placebo auf Verum im von der EMA empfohlenen Zeitraum.

Es sollte daher im Berichtsplan sichergestellt werden, dass die Einhaltung der EMA-Vorgaben bezüglich des Studiendesigns nicht zum Ausschluss dieser Studien aus dem Bewertungsverfahren führt.

Ein Ausschluss von Studien, die die Vorgaben von Zulassungsbehörden erfüllen, wäre auch insofern mit der Rechtsprechung unvereinbar, als es bei der Auswahl und Begutachtung von Studien nach der Einschätzung des BSG (Urteil vom 31.05.2006, Az. B 6 KA 13/05 R, BSGE 96, 261 – Clopidogrel) für Entscheidungen des Gemeinsamen Bundesausschusses zwingend erforderlich ist, den anerkannten Stand der medizinischen Erkenntnisse umfassend zu ermitteln und zu reflektieren, und zwar auf der nach dem jeweiligen Erkenntnisstand bestmöglichen Evidenzstufe. Der medizinische Nutzen hinsichtlich Qualität, Wirksamkeit und medizinischer Unbedenklichkeit eines Arzneimittels darf dabei nicht im Widerspruch zur Bewertung der Zulassungsbehörde stehen, was auch ausdrücklich im derzeit im Deutschen Bundestag beratenen "Gesetz zur Neuordnung des Arzneimittelmarktes in der Gesetzlichen Krankenversicherung" (AMNOG, BT-Drucksache 17/2413) in § 92 Abs. 2 SGB V-E so übernommen wird.

3. Ex-ante Festlegungen bei Instrumenten und kombinierten Endpunkten

Voraussetzung jeden wissenschaftlichen Arbeitens ist, die Methoden vor Kenntnis der Daten festzulegen. Dies ist in Bezug auf eine Reihe von Aspekten bisher nicht geschehen. Wir empfehlen daher nachdrücklich, dies nachzuholen und zur wissenschaftlichen Erörterung zu stellen. Erst dann kann nach unserer Auffassung Einsicht in die Daten der Biologika-Studien genommen und mit den Auswertungen begonnen werden.

Unklar und in einem vorläufigen Berichtsplan explizit zu machen ist unseres Ermessens:

- Welche Instrumente (z. B. zur Messung subjektiver Endpunkte) bzw. kombinierte Endpunkte (z. B. zur Messung der Remission) werden im Rahmen dieser Untersuchung als valide erachtet? Welche nicht? Der allgemeine Verweis, es müsse sich ausschließlich um valide Instrumente bzw. im Falle kombinierter Endpunkte ausschließlich um patientenrelevante Komponenten oder valide Surrogate handeln, ist unzureichend. Dies muss für jedes Instrument bzw. jeden kombinierten Endpunkt aus einem vorläufigen Berichtsplan transparent hervorgehen.
- Für jedes als valide erachtete Instrument bzw. für jeden als valide erachteten kombinierten Endpunkt: Welche Relevanzmaße (z. B. eine Irrelevanzschwelle für einen Gruppenunterschied; eine individuelle *minimal important difference*, etc.) werden betrachtet? Mit welcher Begründung? Mit welchen Belegen?
- Für jedes eingeschlossene Relevanzmaß eines als valide erachteten Instruments bzw. kombinierten Endpunktes: Welche numerischen Relevanzschwellen werden angesetzt? Auch hier: Mit welcher Begründung, und insbesondere mit welchen Belegen?
- Für jede als belegt erachtete numerische Relevanzschwelle eines als valide erachteten Instruments bzw. kombinierten Endpunktes: Welche Analysen (z. B. Betrachtung von Mittelwertdifferenzen oder Responderanalysen) werden vorgeschlagen?

Um Verzerrungen auszuschließen, darf ohne diese Vorarbeiten nicht mit den Auswertungen der Biologika-Studien begonnen werden. Da es sich um mitunter fazitrelevante, zentrale methodische Fragen handelt, sind diese in einem vorläufigen Berichtsplan zu konkretisieren und zur wissenschaftlichen Erörterung freizugeben. Erst dann darf der Berichtsplan unseres Erachtens finalisiert werden.

4. Ex-ante Festlegung der Vergleichsmethoden

In seiner Verfahrensordnung gibt der Gemeinsame Bundesausschuss vor, im Falle fehlender direkter Vergleichsstudien zu prüfen, „...ob placebo-kontrollierte Studien verfügbar sind, die sich für einen indirekten Nachweis einer therapeutischen Verbesserung eignen und den in Absatz 3 beschriebenen Qualitätsanforderungen entsprechen.“ [VerfO des G-BA in der Fassung vom 18. Dezember 2008, zuletzt geändert am 17. Dezember 2009: 50].

Die Allgemeinen Methoden des IQWiG zur Bewertung von Verhältnissen zwischen Nutzen und Kosten (Version 1.0, Seite 16) verweisen auf die methodischen Voraussetzungen einer *Mixed Treatment Comparison*-Meta-Analyse und mahnen an, dass „...eine vollständige Beschreibung des verwendeten Modells zusammen mit verbleibenden Unklarheiten im Berichtsplan anzustreben (ist)“.

Mit keinem Wort äußert sich hierzu allerdings der vorliegende vorläufige Berichtsplan.

Dies wiegt umso schwerer, als als hinreichend belegt gilt, dass indirekte Vergleiche im Allgemeinen und im vorliegenden Auftrag im Speziellen einer Reihe methodischer Herausforderungen höchster Fazitrelevanz unterliegen. Potentielle Confounder bezüglich der Wirksamkeit (z. B. Art der Vorbehandlung oder der Begleittherapie, Dauer und Schwere der Erkrankung) und der Sicherheit (z. B. wo und wann die Studie durchgeführt wurde) sind hinlänglich beschrieben³. Ein aktueller Cochrane-Review⁴ zum Beispiel zog daraus folgende Konsequenzen: *“Indirect comparisons of biologics were performed for efficacy and safety using a hierarchical generalized linear mixed model (GLMM) incorporating the most important study-level characteristic (i.e. type of biologic) as a fixed factor and study and study drug interaction as random factors”*.

Wir beantragen daher, dass im vorläufigen Berichtsplan transparent gemacht und zur Anhörung gestellt wird:

- Werden überhaupt indirekte Vergleiche angestrebt? Wenn nicht, was dann?
- Welche indirekten Vergleiche werden durchgeführt (nicht-adjustierte Vergleiche, adjustierte Vergleiche, Meta-Regression, *Mixed Treatment Comparison*, etc.)?
- Für welche Confounder wird wie kontrolliert?

- Welche Zielgrößen würden in indirekte Vergleiche einbezogen? Würde für multiple, indirekte Vergleiche adjustiert? Mit welcher Methode?

Dies kann und muss aufgrund der hinreichenden gut publizierten methodischen Herausforderungen beim Vergleich von Biologika bei Rheumatoider Arthritis vor Einsichtnahme in die Daten der zu bewertenden Produkte erfolgen und muss aufgrund der hohen Fazitrelevanz bereits auf Berichtsplanebene zur Anhörung gestellt werden.

5. Vorgehen bei wesentlichen methodischen Änderungen im Laufe der Vorberichtserstellung

Im vorläufigen Berichtsplan wird unter 4.1.7 ausgeführt: *„Sofern sich im Projektverlauf Änderungen im Zulassungsstatus der zu bewertenden Interventionen ergeben, werden die Kriterien für den Studieneinschluss gegebenenfalls an die neuen Zulassungsbedingungen angepasst. Die jeweils vorgenommenen Änderungen werden im Vorbericht bzw. im Abschlussbericht explizit vermerkt“.*

Dies widerspricht den Vorgaben im Methodenpapier 3.0, Seite 18: *„Sofern weitere wesentliche methodische Änderungen im Verlauf der Vorberichtserstellung erforderlich sind, erfolgen diese i. d. R. in Form eines oder mehrerer Amendments zum Berichtsplan. Nach Veröffentlichung eines Amendments besteht i. d. R. ebenfalls Gelegenheit zur Stellungnahme zu diesem Amendment unter den o. g. Bedingungen“.*

Da eine Veränderung der Kriterien für den Studieneinschluss als wesentliche methodische Änderung einzustufen ist, sind entsprechende Änderungen in Form von Amendments zum Berichtsplan mit der Gelegenheit zur erneuten Stellungnahme vorzunehmen. Ansonsten wäre die gesetzlich vorgesehene hohe Verfahrenstransparenz und angemessene Beteiligung in jedem Verfahrensstadium nicht gewährleistet.

6. Geänderte Rechtsgrundlage für Nutzenbewertungen

Die Beauftragung des IQWiG durch den Gemeinsamen Bundesausschuss mit der Durchführung einer Nutzenbewertung von biotechnologisch hergestellten Arzneimitteln in der Zweitlinientherapie der rheumatoiden Arthritis nach § 35b SGB V erscheint zum derzeitigen Zeitpunkt nicht mehr sinnvoll.

Das derzeit im Deutschen Bundestag beratene "Gesetz zur Neuordnung des Arzneimittelmarktes in der Gesetzlichen Krankenversicherung" (AMNOG, BT-Drucksache 17/2413) wird die Nutzenbewertung maßgeblich reformieren. Eine Nutzenbewertung auf einer vom IQWiG definierten Studienbasis ist in der Neufassung des § 35a SGB V-E nicht mehr vorgesehen. Vielmehr wird diese auf der Basis eines hiervon erheblich abweichenden, gesetzlich definierten Herstellerdossiers durchgeführt.

Darüber hinaus wird das Bundesministerium für Gesundheit neu festlegen, "welche Grundsätze für die Bestimmung der Vergleichstherapie gelten, in welchen Fällen zusätzliche Nachweise erforderlich sind und unter welchen Voraussetzungen Studien welcher Evidenzstufe zu verlangen sind" (Gesundheitsausschuss, Drucksache 17(14)0067, Seite 3). Dabei sollen auch Übergangsregelungen festgelegt werden. Nur so sieht der Gesetzgeber eine "rechtssichere und tragfähige Grundlage" für die Nutzenbewertung sichergestellt. Wir regen daher an, die Nutzenbewertung zurückzustellen, bis eine Klärung der zukünftigen rechtlichen Rahmenbedingungen für Nutzenbewertungen durch Gesetzgeber bzw. Ministerium erfolgt ist.

Referenzen

1. European Medicines Agency. Points to consider on clinical investigation of medicinal products other than NSAIDs for treatment of rheumatoid arthritis [online]. 17.12.2003 [Zugriff: 23.12.2009]. URL: <http://www.ema.europa.eu/pdfs/human/ewp/055695en.pdf>
2. Food and Drug Administration. Guidance for industry: clinical development programs for drugs, devices, and biological products for the treatment of rheumatoid arthritis (RA) [online]. 02.1999 [Zugriff: 23.12.2009]. URL: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm071579.pdf>
3. Smolen JS. Pharmacotherapy: How well can we compare different biologic agents for RA? *Nat Rev Rheumatol.* 2010 May;6(5):247-8.
4. Singh JA et al. Biologics for rheumatoid arthritis: an overview of Cochrane reviews. *Cochrane Database Syst Rev.* 2009 Oct 7;(4):CD007848.

A 1.6 Pfizer Pharma GmbH

Autoren:

Leverkus, Friedhelm

Löschmann, Peter-Andreas, Priv. Doz. Dr.

Reimann, Thomas, Dr.



**Stellungnahme der Firma Pfizer Pharma GmbH
(im Folgenden „Pfizer“)
zum
vorläufigen Berichtsplan A10-01 (Version 1.0) vom 20.09.2010,
Biologika – Zweitlinientherapie bei rheumatoider Arthritis**

Zu 1 Hintergrund

Begriffe „Erstlinien-“ und „Zweitlinientherapie“

Die Begriffe Erstlinien- und Zweitlinientherapie werden diskutiert, ohne dass ausdrücklich ausgeführt wird, wie diese für die Analyse definiert werden. Gemeint ist vermutlich der Einsatz von Biologika nach vorangegangener Therapie mit einem oder mehreren konventionellen DMARD. Wir möchten Sie bitten, die Definition der Zweitlinientherapie im Berichtsplan zu spezifizieren.

Zu 4.1.4 Studientypen

Vorgesehen ist die ausschließliche Evaluation von randomisierten kontrollierten Studien (RCTs), da diese „möglich und praktisch durchführbar“ sind.

Datengrundlage sollten aber alle verfügbaren Studien nach dem Prinzip der best available evidenz sein, da nicht zu jeder Fragestellung RCTs vorliegen. Z.B liegen Vergleichsuntersuchungen („head-to-head studies“) von Biologika in der Indikation rheumatoide Arthritis nicht vor.

Insbesondere für Sicherheitsendpunkte können auch nicht-randomisierte Studien wertvolle Erkenntnisse im Rahmen der Nutzenbewertung liefern.

- Wichtige Daten zur Sicherheit, Verträglichkeit und Langzeitwirksamkeit werden auch nach Abschluss der verblindeten Phase von Studien (open-label extention) erhoben. Wir halten es für geboten, auch solche Daten in die Nutzenbewertung einzubeziehen.

- Registerdatenbanken bilden eine weitere wichtige Informationsquelle zur Beurteilung von Sicherheit und Verträglichkeit von Medikamenten und bieten Informationen zu den über Jahre im Praxisalltag gesammelten Erfahrungen. In Deutschland wurde unmittelbar nach Zulassung von Etanercept und Infliximab eine prospektive Kohortenstudie - das Deutsche Biologika-Register „RABBIT Register“ - gemeinsam durch das Deutsche Rheumaforschungszentrum und die beteiligten Unternehmen ins Leben gerufen.

Alle später zugelassenen Substanzen wie das Adalimumab und alle weiteren Biologika wurden nach der Zulassung in das Register aufgenommen. In dem Register wird auch eine Kontrollgruppe von Patienten mitgeführt die mit konventionellen DMARD behandelt wird. Auch vergleichende Untersuchungen zu verschiedenen Aspekten der Therapie mit Biologika werden regelmäßig publiziert^{1,2}. Darüber hinaus existieren umfangreiche publizierte Daten, die in den Safety-Registern anderer Länder erhoben wurden^{3,4}.

- Weiter lassen sich in den Registern therapeutisch wichtige Erkenntnisse gewinnen, wie beispielsweise die Tendenz zur Dosissteigerung über die Behandlungsdauer -⁵.
- Gerade bei den, die Sicherheit von Biologika betreffenden Zielgrößen (z. B. unerwünschte Ereignisse, Infektions- und Tumorrisiko) ist eine Betrachtung auch offener Studien und Register wichtig ⁶.

Auch ein im Auftrag des Bundesministeriums für Gesundheit erstelltes Diskussionspapier hält fest, dass man gerade bei einer umfassenden Nutzenbewertung unter Berücksichtigung der Schadenspotentiale (Verträglichkeit) auf Studien „anderen Zuschnitts kaum verzichtet... (kann)... (Phase 4 Studien, Registerstudien, Anwendungsbeobachtungen)“ ⁷.

Zu 4.1.5 Studiendauer

Unter Hinweis auf die Richtlinien der FDA aus dem Jahr 1999 und der EMA aus dem Jahr 2003 sollen nur Studien berücksichtigt werden die 24 Wochen oder länger dauerten. Hierzu ist zu bemerken, dass die FDA Richtlinien diese zeitliche Beschränkung keineswegs apodiktisch fordern und die der EMA erst lange nach der Zulassung von Enbrel im Jahr 1999 veröffentlicht wurden. Die klinische Entwicklung

von Etanercept wurde aber schon im Jahr 1992 begonnen. Gegen die Berücksichtigung kürzer dauernder Studien kann es keine Einwände geben, da der Therapieerfolg nicht überschätzt, sondern tendenziell eher unterschätzt wird. Die Krankheitsaktivität korreliert 3-6 Monate nach Therapiebeginn auch mit dem Langzeitverlauf und der Prognose der Erkrankung⁸. In einer retrospektiven Analyse mehrerer RCTs wurde gezeigt, dass die bessere Krankheitskontrolle der Interventionsgruppe bezüglich Krankheitszeichen und -symptomen bereits nach drei Monaten beurteilt werden kann⁹. Im Einklang mit diesen Befunden empfehlen die Treat-to-Target-Expertengruppe¹⁰, EULAR¹¹ und die DGRh¹² die klinische Beurteilung der Therapie nach 3 Behandlungsmonaten.

Als **Zielgrößen** für die patientenrelevanten Therapieziele werden das ACR-Remissionskriterium und der Disease Activity Score (DAS) Remissions-schwellenwert genannt. Wir möchten Sie bitten, diese patientenrelevanten Endpunkte im Kapitel 4.1.3 aufzuführen.

Zu 4.4. Informationsbewertung

Das IQWiG sieht im vorläufigen Berichtsplan die Bewertung der klinischen Relevanz eines beobachteten Patientennutzens vor. Dabei soll die Relevanzbewertung je nach Verfügbarkeit der Daten auf Basis von Mittelwertsdifferenzen und / oder anhand von Responderanalysen unter primärer Berücksichtigung von validierten MIDs (minimal important differences) erfolgen.

Liegen für (komplexe) Skalen keine validierten Grenzen vor, so bezieht sich das IQWiG auf Hedges' g von 0,2. Hedges' g ist eine Modifikation von Cohen's d. Es ist ein rein statistisches Maß, das zur Vereinfachung der Fallzahlab-schätzung beschrieben wurde. Cohen bemerkt, dass mit der Einführung der Konvention, Effekte in klein, mittel und groß zu kategorisieren, nicht die Wichtigkeit (klinische Relevanz) von kleinen, mittleren oder großen Effekten präjudiziert ist. Er warnt vor einer Missdeutung und betont die Relativität der Kategorien in verschiedenen Anwendungsfeldern¹³.

Die europäische Arzneimittelbehörde EMA lehnt bei der Bestimmung eines „noninferiority margins“ solche Effektgrößen als alleinige Grundlage ab und fordert stattdessen eine Interpretation im klinischen Kontext: „It is not appropriate to use effect size (treatment difference divided by standard deviation) as justification for the choice of non-inferiority margin. This statistic provides information on how difficult a difference would be to detect, but does not help justify the clinical relevance of the difference, and does not ensure that the test product is superior to placebo.“¹⁴

Das IQWiG will – falls Responderauswertungen nicht zur Verfügung stehen – einen Test mit verschobenen Nullhypothesen anwenden, indem es fordert, dass das zum beobachteten Effekt korrespondierende Konfidenzintervall vollständig oberhalb der Relevanzschwelle liegt, damit von einer relevanten Effektstärke ausgegangen werden kann. Dieser Test ist bei klinischen Prüfungen, außer im Nicht-Unterlegenheits-Design, unüblich. Bei dieser Vorgehensweise ist der Nachweis, dass eine MID statistisch signifikant überschritten wird, nur mit sehr großen Fallzahlen zu führen. Die MID, die durchaus relevant ist, kann praktisch nicht nachgewiesen werden. Chuang Stein et al¹⁵ kommen zu dem Schluss, dass dieser Ansatz nicht zu empfehlen ist.

Auf der Patientenebene lassen sich unter Umständen mit verschiedenen Methoden die klinisch bedeutende Differenzen wissenschaftlich determinieren. Auf der Gruppenebene ist dies aber nicht möglich. Die Festlegung einer Irrelevanz oder Relevanzgrenze auf der Gruppenebene beinhaltet immer ein Werturteil.

Zu 4.4.2. Meta-Analysen

Randomeffekt Modelle sollten nicht generell das Modell der Wahl sein. Nach Whitehead sollte für die Wahl des Modells neben der Größe der erwarteten Heterogenität die Anzahl der Studien berücksichtigt werden. Bei einer kleinen Anzahl von Studien, wie bei der Nutzenbewertung erscheinen REMs weniger geeignet. Falls die Heterogenität nicht zu groß ist, sollten Fixed Effect Models angewandt werden.

Literaturverzeichnis:

- ¹ Strangfeld A, et al, Comparative effectiveness of TNFalpha inhibitors in combination with either methotrexate or leflunomide. *Ann Rheum Dis*, published online January 6, 2009
- ² Strangfeld et al. Risk of Herpes Zoster in Patients, *JAMA* 2009—301, 7, 737
- ³ Dixon WG et al, Drug-specific risk of tuberculosis in patients with rheumatoid arthritis treated with anti-TNF therapy Results from the British Society for Rheumatology Biologics Register (BSRBR), *Ann Rheum Dis* 2010, 69, 522-528
- ⁴ Curtis et al, A Comparison of Patient Characteristics and Outcomes in Selected European and US Rheumatoid Arthritis Registries . *Semin Arthritis Rheum.* 2010,4 0, 2-14
- ⁵ Finckh,A et al, Evidence for differential acquired drug resistance to anti-tumour necrosis factor agents in rheumatoid arthritis , *Ann Rheum Dis* 2006, 65,746-752
- ⁶ Strangfeld A, Assessing cancer risk of cytokine inhibitors in RA, *Nature Reviews Rheumatology* 2010, 6, 126-127
- ⁷ Antes et al., 2007: Kommentierende Synopse zur Kosten-Nutzen-Bewertung für Arzneimittel, Freiburg, Essen, Greifswald, Kiel, Oktober 2007
- ⁸ Aletaha, D; Funovits, J; Keystone, EC; Smolen, JS: Disease activity early in the course of treatment predicts response to therapy after one year in rheumatoid arthritis patients. *Arthritis Rheum* 56, 2007, 10, 3226-3235
- ⁹ Boers, M, The time has come to limit the placebo period in rheumatoid arthritis trials to 3 months: a systematic comparison of 3- and 6-month response rates in trials of biological agents. *Ann Rheum Dis* 69, 2010, 1, 186-192
- ¹⁰ Smolen, JS; Aletaha, D; Bijlsma, JW; Breedveld, FC et al., Treating rheumatoid arthritis to target: recommendations of an international task force. *Ann Rheum Dis* 69, 2010, 4, 631-637
- ¹¹ Smolen, JS; Landewe, R; Breedveld, FC; Dougados, et al., EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs. *Ann Rheum Dis* 69, 2010, 6, 964-975
- ¹² Schneider, M; Lelgemann, M; Abholz, H; Caratti, R; Flügge, C; Jäniche, H; Kunz, R; Krüger, K; Rehart, S; Specker, C: Interdisziplinäre Leitlinie Management der frühen rheumatoiden Arthritis. 2007.
- ¹³ Cohen J. *Statistical power analysis for the behavioral sciences*, 2nd Edition. Taylor & Francis Group, New York 1988
- ¹⁴ EMA – Committee for medicinal products for human use (CHMP). Guideline on the choice of inferiority margin 2005
- ¹⁵ Chuang-Stein C, et al. The role of the minimum clinically important difference and its impact on designing a trial. *Pharmceut. Statist.* 2010; Oct. 8.

A 1.7 Roche Pharma AG

Autoren:

Iking-Konert, Christof, Priv. Doz. Dr.

Neumann, Norbert

Ruof, Jörg, Dr.

**Stellungnahme der Roche Pharma AG zum vorläufigen IQWiG-
Berichtsplan A10-01:**

**„Biotechnologisch hergestellte Arzneimittel in der
Zweitlinientherapie der rheumatoiden Arthritis“**

Inhaltsverzeichnis

Abkürzungsverzeichnis inkl. Erläuterung wichtiger Begriffe	3
1 Anmerkungen zur projektspezifischen Methodik	5
1.1 Berücksichtigung des DAS28-Remissionskriteriums (ad 4.1.3)	5
1.2 Berücksichtigung der Krankheitsaktivität als patientenrelevanten Endpunkt (ad 4.1.3)	9
1.3 Messinstrumente für die vorgegebenen patientenrelevanten Endpunkte (ad 4.1.3)	12
1.4 Beschreibung der Validierung von kombinierten Endpunkten mit Surrogatparametern (ad 4.1.3)	12
1.5 Berücksichtigung von Studien mit einer Dauer von mindestens 12 Wochen (ad 4.1.5)	13
1.6 Berücksichtigung von bestimmten Post-hoc-Subgruppenanalysen (ad 4.1.8)	14
1.7 Lage des Konfidenzintervalls bei Bewertung der klinischen Relevanz (ad 4.4)	17
1.8 Ergänzung des Jahrs des Studienbeginns als Effektmodifikator (ad 4.4.4)	18
1.9 Verlängerung der Frist für die Stellungnahme zum Vorbericht (ad 3.1)	19
2 Verweise auf qualitativ angemessene Unterlagen, insbesondere bislang unpublizierte Daten	20
2.1 Berücksichtigung der aktuellen Fachinformationen	20
2.2 Head-to-Head-Studien von Roche	20
3 Literaturverzeichnis	21
4 Anhang	27
4.1 Tabelle 2: Internationaler Standard	27
4.2 Remissionskriterium DAS28<2,6 als valider Surrogat für radiographische Progressionshemmung	28

Abkürzungsverzeichnis inkl. Erläuterung wichtiger Begriffe

Abkürzung	Bedeutung	Erläuterung
ACR	American College of Rheumatology	
AHRQ	Agency for Healthcare Research and Quality	
BSG	Blutkörperchensenkungsgeschwindigkeit	
COX	Cyclooxygenasen	
CADTH	Canadian Agency for Drugs and Technologies in Health	
CDAI	Clinical Disease Activity Index	
CRP	C-reaktives Protein	
DAS	Disease Activity Score	
DGRh	Deutsche Gesellschaft für Rheumatologie	
DMARD	Disease-Modifying Antirheumatic Drug (Krankheitsmodifizierendes Antirheumatikum)	
sDMARD-IR	Synthetic Disease-Modifying Antirheumatic Drug: inadequate response	Unzureichendes Ansprechen auf oder Unverträglichkeit von mindestens einem synthetisch hergestellten DMARD; dient zur Charakterisierung der Vorbehandlung der RA-Patienten
EMA	European Medicines Agency	
Eta	Etanercept	
EULAR	European League Against Rheumatism	
FACIT-F	Functional Assessment of Chronic Illness Therapy Fatigue Scale	
G-BA	Gemeinsamer Bundesausschuss	
HAQ	Health Assessment Questionnaire	
HAQ-DI	Disability Index des HAQ	
HAS	Haute Autorité de Santé	
hsCRP	hochsensibles C-reaktives Protein	
HTA	Health Technology Assessment	
Inf	Infliximab	
IR	inadequate response	Unzureichendes Ansprechen oder Unverträglichkeit
ITT	Intention-to-treat-Population	
LOCF	last observation carried forward	
MDS	Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen	
MTX	Methotrexat	
MTX-IR	Methotrexat: inadequate response	Unzureichendes Ansprechen auf oder Unverträglichkeit von Methotrexat; dient zur Charakterisierung der Vorbehandlung der RA-Patienten
NICE	National Institute for Health and Clinical Excellence	
NSAR	Nichtsteroidale Antirheumatika	
RA	Rheumatoide Arthritis	
RCT	Randomised controlled trial (Randomisierte und kontrollierte Studie)	

Abkürzung	Bedeutung	Erläuterung
SDAI	Simplified Disease Activity Index	
SF-36	Short Form 36	
SJC	Swollen joint count	
TJC	Tender joint count	
TNF	Tumornekrosefaktor	
TNF-IR	TNF inhibitors: inadequate response	Unzureichendes Ansprechen auf oder Unverträglichkeit von TNF-alpha-Hemmern; dient zur Charakterisierung der Vorbehandlung der RA-Patienten
T2T	Treat-to-Target	
TSS	Total Sharp Score	
VAS	Visual Analogue Scale	
WHO	World Health Organisation	

Die Roche Pharma AG nimmt wie folgt Stellung zum vorläufigen IQWiG-Berichtsplan A10-01: „Biotechnologisch hergestellte Arzneimittel in der Zweitlinientherapie der rheumatoiden Arthritis“ (Version 1.0 vom 20.09.2010)

1 Anmerkungen zur projektspezifischen Methodik

1.1 Berücksichtigung des DAS28-Remissionskriteriums (ad 4.1.3)

Das IQWiG führt die Remission als patientenrelevanten Endpunkt an (siehe Kap. 4.1.3 des vorläufigen Berichtplans). Die Kriterien zur Messung der Remission werden allerdings nicht spezifiziert. Das Remissionskriterium Disease Activity Score 28 (DAS28) $<2,6$ sollte als gut validiertes Scoringverfahren explizit zur Beurteilung der Remission und der DAS28 entsprechend zur Beurteilung der Krankheitsaktivität herangezogen werden. Nachfolgend wird die Bedeutung des Remissionskriteriums DAS28 $<2,6$ und des DAS28 dargestellt.

Bedeutung des Remissionskriteriums DAS28 $<2,6$ und des DAS28 insgesamt

Zur Beurteilung einer Remission gibt es unterschiedliche Schwellenwerte in Abhängigkeit vom angewandten Scoringverfahren (DAS44, DAS28, Simplified Disease Activity Index (SDAI), Clinical Disease Activity Index (CDAI) oder American College of Rheumatology (ACR) Remissionskriterien). So definiert beispielsweise der DAS28 einen Wert von unter 2,6 als klinische Remission^{1,2}.

Der DAS ist der bekannteste, unter den wissenschaftlichen Gesellschaften und Rheumatologen weitest verbreitete und akzeptierte und am besten validierte Index zur Beurteilung der Krankheitsaktivität³. Er wurde bereits vor 20 Jahren entwickelt und wurde ursprünglich an 44 Gelenken erhoben, später jedoch aufgrund einer besseren Anwendbarkeit zum DAS 28 modifiziert^{4,5} und zugleich zur Vergleichbarkeit therapeutischer Interventionen mit nicht-biologischen und biologischen Disease-Modifying Antirheumatic Drugs (krankheitsmodifizierende Antirheumatika; DMARDs) verwendet. Zudem ist er der erste Score, der absolute Werte der Erkrankungsaktivität nutzt und dahingehend gegenüber den ACR Kriterien mit relativen Änderungen eine patientenorientierte Innovation darstellt.

Der DAS28 wird an 28 Gelenken erhoben und berechnet sich nach einer validierten⁵ Formel anhand der Einzelkomponenten Anzahl druckschmerzhafter Gelenke (Tender joint count; TJC), Anzahl geschwollener Gelenke (Swollen joint count; SJC), Blutkörperchensenkungsgeschwindigkeit (BSG) und Einschätzung der globalen Gesundheit (mittels Visual Analogue Scale (VAS)) durch den Patienten. Mit den Angaben des DAS28 ist die Krankheitsaktivität gut darstellbar und erlaubt den Vergleich von Studienergebnissen, sowie die individuelle Verlaufskontrolle der Erkrankung für jeden Patienten über den Erkrankungszeitraum. Zudem

wurden verschiedene Aktivitätsgrenzen des DAS28 evaluiert und es kann zwischen hoher ($>5,1$), moderater ($>3,2-5,1$), niedriger ($<3,2$) und keiner Krankheitsaktivität ($<2,6$ = klinische Remission) unterschieden werden¹. Das Erreichen eines niedrigen DAS Wertes ($<3,2$) bzw. einer klinischen Remission ($<2,6$) ist das postulierte Therapieziel⁶. Der Hintergrund für dieses Vorgehen liegt darin, dass gezeigt werden konnte, dass durch eine niedrige Krankheitsaktivität bzw. Remission in der frühen Phase der Rheumatoiden Arthritis (RA)-Therapie (1-2 Jahre) die langfristige Prognose der Patienten verbessert werden kann⁷. Die radiologische Progression, die im Verlauf der RA Erkrankung deutlich zunimmt und letztendlich zu einem Funktionsverlust der Gelenke führt, kann durch das Erreichen eines niedrigen DAS28 Wertes bzw. einer DAS28 Remission aufgehalten werden⁷. Die Grenzwerte des DAS28 für niedrige Krankheitsaktivität ($<3,2$) und klinische Remission ($<2,6$) sind in Bezug auf die radiologische Progression validiert^{7, 8}. Darüber hinaus zeigt eine eigene Analyse (siehe Kap. 4.2. im Anhang) folgendes: Das Remissionskriterium $\text{DAS28} < 2,6$ ist ein valider Surrogat für die radiographische Progressionshemmung und zeigt vergleichbare Korrelationen wie SDAI und CDAI. Für diese Parameter liegen allerdings weniger Vergleichsdaten aus randomisierten, klinischen Studien (RCTs) vor².

Patientenrelevanz der Einzelkomponenten des DAS28

„Zur Erhebung der Remission als patientenrelevanter Endpunkt sind solche Remissionskriterien geeignet, in die ausschließlich patientenrelevante Komponenten oder valide Surrogate eingehen“ (siehe vorläufiger Berichtsplan; Abschnitt: Ziele der Behandlung (Seite 3)).

Der DAS28 ist ein von der European League Against Rheumatism (EULAR) entwickeltes und validiertes System. Die im DAS28 enthaltenen Einzelkomponenten sind Bestandteile des international (ACR / EULAR / World Health Organisation (WHO)) akzeptierten Core-Sets zur Beurteilung der Krankheitsaktivität. Folgende Einzelkomponenten sind im DAS28 enthalten:

- Anzahl geschwollener Gelenke
- Anzahl druckschmerzhafter Gelenke
- Globale Einschätzung der Krankheitsaktivität / des Gesundheitsstatus durch den Patienten
- Akutphase Reaktanten, C-reaktives Protein (CRP)/BSG

Die „Anzahl druckschmerzhafter Gelenke“ findet sich implizit als Kriterium für die „Symptomatik der Entzündung“ im vorläufigen Berichtsplan wieder und ist damit aus unserer Sicht als patientenrelevant akzeptiert.

Die weiteren drei Einzelkomponenten des DAS28 sind ebenso patientenrelevant. Nachfolgend die Begründung:

Anzahl geschwollener Gelenke:

Geschwollene Gelenke sind direkter Ausdruck der entzündlichen Gelenkerkrankung⁹ und so unmittelbar für den RA-Patienten von Relevanz. Die Anzahl der geschwollenen Gelenke sind mit der Gelenkschädigung assoziiert und ein Prädiktor für die radiologische Progression⁹. Daher erfassen auch die neuen 2010 ACR/EULAR Klassifikationskriterien für Rheumatoide Arthritis¹⁰ die Gelenkbeteiligung als eine der vier übergeordneten Kriterien. Die Gelenkbeteiligung bezieht sich auf jedes geschwollene oder druckschmerzhaftes Gelenk¹⁰. Diese Klassifikationskriterien sollen es ermöglichen, Patienten frühzeitig zu erkennen, die ein hohes Risiko für eine persistierende und erosive RA haben. Hieraus leitet sich direkt die Patientenrelevanz ab.

Globale Einschätzung der Krankheitsaktivität / des Gesundheitsstatus durch den Patienten

Die subjektive Einschätzung der Krankheitsaktivität / Gesundheitsstatus durch den Patienten drückt direkt die vom Patienten erlebte Krankheitslast aus, spiegelt wider wie der Patient selbst seine Lebenssituation beurteilt und ist daher direkt und in hohem Maße patientenrelevant.

Akutphase Reaktanten, CRP/BSG:

Die BSG und das C-reaktive Protein korrelieren sehr gut mit klinischen Parametern der Krankheitsaktivität. Sie erwiesen sich in klinischen Prüfungen von Basistherapeutika als änderungssensitiv und sind gute Prädiktoren für das Ausmaß der röntgenologischen Progression¹¹. CRP korreliert darüber hinaus mit der Erkrankungsaktivität und Funktionseinschränkung¹². Ebenso korrelieren die CRP-Werte über die Zeit mit der Zerstörung der Gelenke⁹. Die BSG wurde auch als unabhängiger Prädiktor für den frühen Knochendichteverlust beschrieben¹³. Daher erfassen auch die neuen 2010 ACR/EULAR Klassifikationskriterien für RA¹⁰ Akutphase Reaktanten wie BSG oder CRP. Diese Klassifikationskriterien sollen es noch besser ermöglichen, Patienten frühzeitig zu erkennen, die ein hohes Risiko für eine persistierende und erosive RA haben. Es gibt hinreichende Evidenz, dass das hochsensible C-reaktive Protein (hsCRP) direkt mit dem kardiovaskulären Risiko assoziiert ist und somit eine erhebliche Relevanz für den Patienten bedeutet^{14, 15}, da die kardiovaskuläre Mortalität der führende Sterblichkeitsgrund bei der RA ist. Aus den oben ausgeführten Angaben leitet sich direkt die Patientenrelevanz ab.

DAS28 als Composite Score

Die einzelnen Komponenten des „Core-Sets“ reflektieren unterschiedliche Aspekte der RA. Die RA ist heterogen und Composite Scores erfassen diese Heterogenität am besten⁶. Das Treat-to-Target Expertengremium betrachtet die Composite Scores als ein in der klinischen Praxis notwendiges Instrument. Das Expertengremium ist überzeugt, dass der Einsatz von Composite Scores der beste Weg ist, die

Krankheitsaktivität und das Ansprechen auf eine Therapie zu beurteilen⁶. Die Krankheitsaktivität der RA kann nicht ausreichend alleine durch die Einzelparameter des Core-Sets abgebildet werden. In die mathematische Berechnung des Composite Scores DAS 28 gehen die Einzelkomponenten in validierter Gewichtung ein. Der DAS28-Score ist zudem in den Arzneimittelrichtlinien des Gemeinsamen Bundesausschusses (G-BA) als Parameter für das Ansprechen der Therapie anerkannt¹⁶. Erst in der Zusammenschau der Einzelparameter der Zeichen einer RA und der subjektiven Sicht des Patienten kann die Krankheitsaktivität bestimmt und in ihrem Verlauf bewertet werden.

Internationales Vorgehen

Das IQWiG hat bisher Cochrane Reviews und Health Technology Assessments (HTAs) ausländischer HTA-Institute als internationalen Standard herangezogen (z.B. Nutzenbewertung der Glitazone¹⁷). Bei folgenden HTA-Instituten mit englischsprachigen HTAs wurden HTAs zur Behandlung der RA mit Biologika gefunden (Recherche am 28.07.10): National Institute for Health and Clinical Excellence (NICE) (England/Wales; mehrere HTAs), Canadian Agency for Drugs and Technologies in Health (CADTH) (Kanada; 1 HTA¹⁸), Agency for Healthcare Research and Quality (AHRQ) (USA; 1 HTA¹⁹). Zusätzlich wurden mehrere Cochrane Reviews zur Behandlung der RA mit Biologika gefunden. Vom NICE wurden beispielhaft die HTAs zu den Roche-Produkten ausgewählt: ein HTA für mehrere Biologika und ein HTA für ein Biologikum (single HTA)^{20, 21}. Von den identifizierten Cochrane Reviews wurden beispielhaft die Cochrane Reviews zu den Roche-Produkten und ein Cochrane Review zu sechs Biologika ausgewählt^{22, 23, 24}. Basierend auf diesen sieben Übersichtsarbeiten (HTAs / Cochrane Reviews) kann der internationale Standard abgebildet werden. Die Tabelle 2 (siehe Anhang) stellt die Mindeststudiendauer sowie die Berücksichtigung von ausgewählten patientenrelevanten Endpunkten und Messinstrumenten in diesen Übersichtsarbeiten dar. In fünf von sieben Übersichtsarbeiten werden die DAS- / DAS28-Remissionskriterien akzeptiert. Wenn diese auch durch das IQWiG berücksichtigt werden, steht die Nutzenbewertung im Einklang mit dem Großteil internationaler HTAs und Cochrane Reviews.

Fazit: Der Stellenwert des DAS28 zur Beurteilung und Dokumentation der Krankheitsaktivität ist sowohl für klinische Studien als auch für die klinische Praxis international anerkannt. Er wird von den Kassenärztlichen Vereinigungen bzw. den Kostenträgern zur Abrechnung der rheumatologischen Grundpauschale gefordert. Sowohl die Einzelkomponenten des DAS28 als auch der DAS28 in seiner Gesamtheit sind patientenrelevant. Der DAS28 Grenzwert ($DAS28 < 2,6$) zur Definition einer klinischen Remission ist in zahlreichen Untersuchungen validiert und seine Korrelation zu patientenrelevanten Endpunkten wie Gelenkzerstörung (inkl. radiographische Progressionshemmung) und Entzündungsaktivität validiert. Deshalb bitten wir nachdrücklich, dass im Kapitel 4.1.3. der patientenrelevante Endpunkt Remission folgendermaßen spezifiziert wird: „*Remission (Remissionskriterium: $DAS28 < 2,6$)*“.

1.2 Berücksichtigung der Krankheitsaktivität als patientenrelevanter Endpunkt (ad 4.1.3)

Im Kapitel 4.1.3 wird der patientenrelevante Endpunkt „Krankheitsaktivität“ nicht aufgelistet, obwohl es auf Seite 2 des vorläufigen Berichtsplans heißt: „Das primäre Ziel der Behandlung der rheumatoiden Arthritis ist die Reduktion der Krankheitsaktivität auf ein Niveau, auf dem die Patienten weitgehend frei von Krankheitssymptomen sind, und auf welchem die Gelenkzerstörung maßgebend verzögert oder verhindert wird (Remission).“ Der kombinierte Endpunkt Krankheitsaktivität, gemessen mit DAS 28, ACR 50 und ACR70, fasst mehrere patientenrelevante Parameter zur Symptomatik, zur Funktionsfähigkeit oder zur gesundheitsbezogenen Lebensqualität zusammen und ist somit selbst patientenrelevant und in der Nutzenbewertung zu berücksichtigen. Nachfolgend wird zuerst die Patientenrelevanz der Krankheitsaktivität allgemein und anschließend (ergänzend zu Kapitel 1.1) spezifisch bezüglich der Messinstrumente ACR50 / ACR70 einschließlich ihrer Komponenten dargestellt.

Krankheitsaktivität

Die Krankheitsaktivität ist ein guter Prädiktor für die zu erwartende Gelenkschädigung und die Funktionseinschränkung^{1, 7, 8, 25, 26, 27, 28, 29, 30}. Zur Bewertung einer erfolgreichen Therapie bei einer bestehenden RA wird die Reduktion der Krankheitsaktivität herangezogen⁶.

Wie im vorläufigen Berichtsplan dargestellt wird die Krankheitsaktivität, gemessen mit ACR50 und ACR70, DAS oder DAS28, in zahlreichen Leitlinien und Übersichtsarbeiten als anerkannter Standard beschrieben.

So werden in den Leitlinien von öffentlichen nationalen HTA-Instituten und der Deutschen Gesellschaft für Rheumatologie (DGRh)³¹ die Messinstrumente DAS und DAS28 als Verlaufskontrolle für den klinischen Alltag empfohlen. Die aktuelle NICE Clinical Guideline³² sowie die französische Haute Autorité de Santé (HAS)³³ legen für die Verlaufskontrolle den DAS28 zugrunde.

Die europäische Zulassungsbehörde European Medicines Agency (EMA) gibt in ihren Empfehlungen zur klinischen Entwicklung von RA-Medikamenten explizit den DAS28 und den ACR als Messinstrumente vor³⁴. Auch in der nun anstehenden Revision der „Points to Consider“ werden ACR70 und DAS als Messinstrumente zur Wirksamkeit favorisiert³⁵.

In allen sieben internationalen Übersichtsarbeiten wird die Krankheitsaktivität als patientenrelevanter Endpunkt berücksichtigt (siehe Anhang: Tabelle 2). Als Messinstrumente für die Krankheitsaktivität werden in allen Übersichtsarbeiten ACR50 und ACR70 und in sechs Übersichtsarbeiten DAS oder DAS28 akzeptiert.

DAS28

Die Patientenrelevanz der Einzelkomponenten des DAS28 und des DAS28 in seiner Gesamtheit ist im Kapitel 1.1. dargestellt.

ACR50 und ACR70

Die Composite Scores ACR50 und ACR70 und deren Einzelkomponenten gehören zu den wichtigsten Wirksamkeitsparametern bei der RA-Therapie in RCTs, die ständig überprüft und angepasst werden¹⁰. Sie spiegeln das relative Maß einer klinischen Verbesserung wider. ACR50 und ACR70 sind Endpunkte, die sehr hohe Anforderungen an die statistische Power und Überlegenheit einer RA-Therapie stellen und wurden umfassend validiert³⁶.

Patientenrelevanz der Einzelkomponenten des ACR50 und ACR70

Tabelle 1: Einzelkomponenten von ACR50, ACR70 und DAS28 als Maß der Krankheitsaktivität

Einzelkomponenten	ACR 50 / 70	DAS28
1. Anzahl geschwollener Gelenke	X	X
2. Anzahl druckschmerzhafter Gelenke	X	X
3. Globale Einschätzung der Krankheitsaktivität durch den Arzt	X	
4. Globale Einschätzung der Krankheitsaktivität / des Gesundheitsstatus durch den Patienten	X	X
5. Schmerz (mittels visuelle Analogskala (VAS))	X	
6. HAQ-Funktionsbehinderungsindex (HAQ)	X	
7. Akutphase Reaktanten, CRP/BSG	X	X

Der vorläufige Berichtsplan sieht unter 4.1.3 acht patientenrelevante Endpunkte vor:

- I. Remission
- II. Symptomatik der Entzündung (insbesondere Schmerz, Fatigue, Morgensteifigkeit)
- III. Strukturelle Gelenkveränderungen (Deformitäten, Versteifungen, Kontrakturen)
- IV. Körperlicher Funktionsstatus einschließlich Aktivitäten des täglichen Lebens
- V. Soziales Funktionsniveau (Teilhabe am beruflichen und sozialen Leben)
- VI. Gesundheitsbezogene Lebensqualität
- VII. Gesamtmortalität
- VIII. Unerwünschte Arzneimittelwirkungen

Zwei Einzelkomponenten des ACR50 / ACR 70, „Anzahl druckschmerzhafter Gelenke“ und „Schmerz (mittels VAS)“ finden sich implizit als Kriterien für die „Symptomatik der Entzündung“ im vorläufigen Berichtsplan des IQWiG wieder und sind somit als patientenrelevant akzeptiert.

Die weiteren fünf Einzelkomponenten des ACR50 / ACR70 sind ebenso patientenrelevant. Nachfolgend die Begründung:

1. Anzahl geschwollener Gelenke

Siehe dazu Kapitel 1.1

3. Globale Einschätzung der Krankheitsaktivität durch den Arzt

Die Einschätzung der allgemeinen Krankheitsaktivität unter Berücksichtigung aller Auswirkungen der Arthritis durch den behandelnden Arzt, wurde im Rahmen einer Analyse des SDAI und seiner Einzelkomponenten validiert¹. Anders als bei der subjektiven Einschätzung der Krankheitsaktivität durch den Patienten, in die auch Beschwerden durch Komorbiditäten eingehen, bewertet der Arzt spezifisch die auf der rheumatischen Erkrankung beruhenden Symptome³⁷.

Damit wird die Krankheitsaktivität der RA nach klinischen Kriterien beurteilt und objektiviert. Zwischen der globalen Beurteilung des Patienten und des Arztes (bestimmt mittels VAS) besteht eine enge Korrelation, auch wenn beim Patienten die Bewertung seiner Schmerzen und beim Arzt die Bestimmung der Anzahl geschwollener Gelenke dominieren³⁸.

4. Globale Einschätzung der Krankheitsaktivität / des Gesundheitsstatus durch den Patienten

Siehe dazu Kapitel 1.1

6. HAQ-Funktionsbehinderungsindex (HAQ)

Der Health Assessment Questionnaire (HAQ) und seine Varianten wie der Disability Index des HAQ (HAQ-DI) sind etablierte und validierte Messinstrumente zur Beurteilung des körperlichen Funktionsstatus³⁹. In einer Studie wurde die Sensitivität des HAQ-DI auf strukturelle Schäden, Schmerz und psychologische Faktoren bestätigt⁴⁰. Er ist einer der besten Prädiktoren für das Langzeitergebnis einer RA-Therapie⁴¹.

Der HAQ wird in allen sieben Übersichtsarbeiten akzeptiert (siehe Anhang: Tabelle 2).

7. Akutphase Reaktanten, CRP/BSG

Siehe dazu Kapitel 1.1

Fazit: Die Krankheitsaktivität ist ein international anerkannter, patientenrelevanter Endpunkt. Sowohl die Einzelkomponenten der Messinstrumente DAS28, ACR50 und ACR70 als auch die Messinstrumente in ihrer Gesamtheit sind patientenrelevant. Deshalb bitten wir nachdrücklich, dass in Kapitel 4.1.3 folgender patientenrelevanter Endpunkt ergänzt wird: *Krankheitsaktivität (insbesondere gemessen mit DAS28, ACR50 und ACR70)*

1.3 Messinstrumente für die vorgegebenen patientenrelevanten Endpunkte (ad 4.1.3)

Wir möchten darauf hinweisen, dass mit dem Functional Assessment of Chronic Illness Therapy Fatigue Scale (FACIT-F)⁴², dem HAQ^{39, 40} und dem Short Form 36 (SF-36)⁴³ valide Messinstrumente für die vorgegebenen patientenrelevanten Endpunkte Fatigue, körperlicher Funktionsstatus einschließlich Aktivität des täglichen Lebens bzw. gesundheitsbezogene Lebensqualität zur Verfügung stehen.

1.4 Beschreibung der Validierung von kombinierten Endpunkten mit Surrogatparametern (ad 4.1.3)

Das Kapitel 4.1.3. zu den patientenrelevanten Endpunkten im vorläufigen Berichtsplan endet mit folgendem Satz: „Kombinierte Endpunkte mit Surrogatparametern werden betrachtet, falls diese anhand der gegebenen Daten validiert werden können.“ Die Methodik für diese Validierung sollte detailliert dargestellt werden. Deshalb bitten wir, dass die Methodik für den Fall einer mündlichen Anhörung den Teilnehmern spätestens zeitgleich mit der Einladung zur Anhörung mitgeteilt wird, damit die Teilnehmer zu dieser Methodik Stellung nehmen können. Folgende Punkte sollten mindestens dargestellt werden:

- Was versteht das IQWiG unter „gegebenen Daten“?
- Welche gegebenen Daten werden für die Validierung herangezogen?
- Welche Methodik wird für die Validierung verwendet?
- Wann werden die kombinierten Endpunkte mit Surrogatparametern als valide angesehen? Dies sollte a priori festgelegt werden.

Fazit: Die Methodik für die Validierung von kombinierten Endpunkten mit Surrogatparametern sollte detailliert dargestellt und den Stellungnehmenden im Falle einer mündlichen Anhörung spätestens zeitgleich mit der Einladung zur Anhörung mitgeteilt werden.

1.5 Berücksichtigung von Studien mit einer Dauer von mindestens 12 Wochen (ad 4.1.5)

Die im vorläufigen Berichtsplan vorgesehene Mindeststudiendauer von 24 Wochen ist nicht in Übereinstimmung mit den Empfehlungen mehrerer Fachgesellschaften zur Therapiesteuerung. Die geforderte Mindeststudiendauer sollte 12 Wochen betragen. Nachfolgend die Begründung, dass eine kürzere Studiendauer als 24 Wochen aussagekräftige Ergebnisse mit Patientenrelevanz generiert.

Änderung der Krankheitsaktivität nach 3 Monaten

Das Ziel einer antirheumatischen Therapie ist es, die Krankheitsaktivität schnell und deutlich zu reduzieren. Von verschiedenen Fachgesellschaften wie der DGRh³¹, der EULAR⁴⁴ oder dem Treat-to-Target (T2T) Expertengremium⁶ wird zur Beurteilung der Krankheitsaktivität u.a. der DAS28 mit seinen Trennwerten von <2,6 (Remission) und niedrige Krankheitsaktivität (<3,2) angegeben. Diese Empfehlungen treffen auch eine Aussage darüber, zu welchem Zeitpunkt der Erfolg einer eingeleiteten Therapie bezüglich seiner Wirkung auf die Krankheitsaktivität erstmalig beurteilt werden kann. Von der EULAR⁴⁴ wird eine Beurteilung nach 3 Monaten, von der DGRh nach 3 Monaten und von der T2T Expertengruppe ebenfalls nach 3 Monaten empfohlen.

Darüber hinaus führt der G-BA in seinem Therapiehinweis zu Adalimumab¹⁶ folgendes aus: „Ein Ansprechen auf die Therapie ist bereits nach 1 bis 2 Wochen zu erwarten. Soweit auch nach 3 Monaten kein deutliches klinisches Ansprechen (klinische Symptomatik, DAS-Score, Labor) zu verzeichnen ist, ist die Therapie mit Adalimumab abzusetzen“. Somit wird auch vom G-BA eine Behandlungsdauer von 3 Monaten zur Beurteilung des Therapieerfolges anerkannt. Dieses ist deckungsgleich zu den Empfehlungen der Fachgesellschaften.

Diese Aussagen werden auch von einer Untersuchung von Boers⁴⁵ bestätigt. Boers zeigte in einer retrospektiven Analyse mehrerer randomisierter Phase-III-Studien, dass die bessere Krankheitskontrolle der Verumgruppe bezüglich Krankheitszeichen und -symptomen bereits nach drei Monaten beurteilt werden kann. Folglich ist eine dreimonatige kontrollierte Studienphase zur Beurteilung der Aktivität einer Medikation ausreichend⁴⁵.

Der Status der Krankheitsaktivität schon 3 Monate nach Therapiebeginn korreliert auch mit dem Langzeitverlauf und der Prognose der Erkrankung⁴⁶.

Zusammenfassend kann gesagt werden, dass eine engmaschige Kontrolle des Therapieerfolges beziehungsweise der Medikamentenwirkung möglich ist und so die Prognose verbessert werden kann^{46, 47, 48, 49}.

Internationales Vorgehen

In drei von sieben Übersichtsarbeiten erfolgt keine Einschränkung der Studiendauer^{21, 23, 24}, in zwei Übersichtsarbeiten beträgt die Mindeststudiendauer 12 Wochen^{19, 20} (siehe Anhang: Tabelle 2). Wenn das IQWiG Studien mit einer Mindestdauer von 12 Wochen berücksichtigt, steht die Nutzenbewertung im Einklang mit dem Großteil internationaler HTAs und Cochrane Reviews.

Fazit: In internationalen HTAs und Cochrane Reviews wird überwiegend eine kürzere Mindeststudiendauer vorgegeben als im IQWiG-Berichtsplan. Darüber hinaus kann der Therapieerfolg bereits nach 12 Wochen (3 Monaten) mittels Evaluation der Entzündung und der Krankheitsaktivität beurteilt werden. Relevante Evidenz aus Studien mit kürzerer Dauer sollte berücksichtigt werden. Wir bitten daher um folgende Umformulierung im Kapitel 4.1.5: *„Eingeschlossen werden Studien mit einer Mindestdauer von 3 Monaten (12 Wochen).“*

1.6 Berücksichtigung von bestimmten Post-hoc-Subgruppenanalysen (ad 4.1.8)

Entsprechende Subgruppenanalysen von Studien, bei denen das Einschlusskriterium E1 (Population) nicht zu 100% erfüllt ist, werden vom IQWiG betrachtet (siehe vorläufiger Berichtsplan Kap. 4.1.8). Eine Berücksichtigung der Subgruppenanalysen von Studien mit einer Teilerfüllung der Einschlusskriterien E2 (Prüfintervention) und E3 (Vergleichsintervention) wird nicht explizit dargestellt. Das IQWiG sollte Post-hoc-Subgruppenanalysen berücksichtigen, die entsprechend der Einschlusskriterien E1, E2, E3 und E4 (Endpunkte) definiert wurden, wenn das Gesamtkollektiv der Studie diese Einschlusskriterien nicht vollständig erfüllt. Nachfolgend die Begründung für die Berücksichtigung dieser Post-hoc-Subgruppenanalysen.

Post-hoc-Subgruppenanalysen in bisherigen Nutzenbewertungen

Das IQWiG hat in anderen Nutzenbewertungen (z.B. von Antihypertensiva und Memantin) Post-hoc-Subgruppenanalysen akzeptiert oder deren Durchführung zur Erfüllung der Zulassungskriterien oder der Verbesserung der Studienqualität selbst gefordert^{50, 51}.

Anforderungen an Subgruppenanalysen

In einem umfassenden Gutachten des Fachbereichs Evidenz-basierte Medizin des Medizinischen Dienstes des Spitzenverbandes Bund der Krankenkassen (MDS) vom 25.3. 2004 wird die Aussagekraft von Post-hoc- und A-

priori-Subgruppenanalysen untersucht⁵². **A-priori-Subgruppenanalysen** sind vor Studienbeginn festgelegt und im Studienprotokoll dokumentiert oder nach Studienbeginn ohne Einsicht in die bereits erhobenen Daten definiert und in einem Amendment zum Studienprotokoll festgehalten. Alle anderen üblichen Formen der Subgruppenanalysen sind datengetrieben und werden als **Post-hoc-Subgruppenanalysen** bezeichnet. Nachfolgend werden zuerst die Ergebnisse des Gutachtens für A-priori-Subgruppenanalysen dargestellt und dann für bestimmte Post-hoc-Subgruppenanalysen überprüft.

Die Schlussfolgerungen des Gutachtens lauten: A-priori-Subgruppenanalysen können confirmatorisch sein, falls vier Kriterien (siehe Tabelle auf Seite 16) erfüllt sind. Weiterhin sollte eine Begründung für die Subgruppenbildung vor der Auswertung angegeben werden. Falls die Randomisierung nicht stratifiziert nach der subgruppen-bildenden Variablen durchgeführt werden kann, sollte es keine Anhaltspunkte für Confounding für die Therapiegruppen innerhalb einer Subgruppe geben⁵².

Erfüllung der Anforderungen für durch die Einschlusskriterien des IQWiG definierte Post-hoc-Subgruppenanalysen

Die o.g. Punkte und Kriterien des Gutachtens werden für solche Post-hoc-Subgruppenanalysen überprüft, die entsprechend der Einschlusskriterien der Nutzenbewertung (siehe Kapitel 4.1. des vorläufigen Berichtsplans) definiert wurden, wenn das Gesamtkollektiv der Studie diese Einschlusskriterien nicht vollständig erfüllt.

Hierzu ist zunächst festzustellen, dass eine solche Subgruppe typischerweise durch ein externes Kriterium (die Einschlusskriterien des IQWiG), unabhängig von den Daten, definiert wird. Es handelt sich dann nicht um eine datengetriebene post-hoc-Subgruppenanalyse.

Solche Post-hoc-Subgruppenanalysen können die Kriterien für A-priori-Subgruppenanalysen erfüllen:

Kriterien für konfirmatorische A-priori-Subgruppenanalysen	Prüfung der Kriterien für Post-hoc-Subgruppenanalysen, die durch die Einschlusskriterien des IQWiG definiert sind
1) Die Subgruppenanalyse gehört zu einer adäquat durchgeführten aussagekräftigen Studie.	Es gibt keine Probleme des multiplen Testens, da nur eine Untergruppe ausgesondert wird und kein Vergleich mit der Restgruppe durchgeführt werden soll. Insofern sind auch keine Interaktionstests nötig und die Kriterien 1)-3) können als erfüllt angesehen werden.
2) Es werden nur wenige Subgruppenanalysen durchgeführt und es wird möglichst eine Korrektur für multiples Testen vorgenommen.	
3) Eine adäquate statistische Methodik wird angewendet (Interaktionstests, kein separates Auswerten der Subgruppen).	
4) Die Subgruppen-bildende Variable ist eine Baseline-Variable, d.h., sie darf nicht nach Therapiebeginn erhoben werden. Ansonsten könnte der Verlauf der Therapie die Zugehörigkeit zur Subgruppe beeinflussen haben.	Ob Kriterium 4) gilt, hängt von den Einschlusskriterien ab. Es muss gesichert sein, dass diese nicht die Zugehörigkeit zu den Therapiegruppen beeinflussen.

In diesen Post-hoc-Subgruppenanalysen wurde die Randomisierung in der Regel nicht stratifiziert nach der interessierenden Subgruppe durchgeführt. Nichtsdestoweniger wird beim Randomisieren davon ausgegangen, dass auch Subgruppen, die kritisch bezüglich unbekannter Confounder sind, gleichmäßig auf die Behandlungsgruppen verteilt sind. In der Praxis ist es vor allem bei kleineren Fallzahlen empfehlenswert, auf die Ausgeglichenheit der Therapiegruppen bezüglich relevanter Confounder zu prüfen. Diese Prüfung stellt auch für randomisierte Studien im Allgemeinen eine qualitätsorientierte Herangehensweise dar.

Fazit: Das IQWiG hat in der Vergangenheit selbst die Erstellung von bestimmten Post-hoc-Subgruppenanalysen, z.B. zur Erfüllung der Zulassungskriterien, gefordert. Es ist statistisch unbedenklich eine Subgruppe eines RCT auszuwerten, falls diese durch unabhängige äußere Kriterien wie die Einschlusskriterien des IQWiG definiert ist. Solche Post-hoc-Subgruppenanalysen erfüllen die Kriterien, die an konfirmatorische A-priori-Subgruppenanalysen gerichtet sind. Deshalb bitten wir nachdrücklich, dass in Kapitel 4.1.8 folgender Satz eingefügt wird: *„Post-hoc-Subgruppenanalysen, die entsprechend der Einschlusskriterien E1, E2, E3 und E4 (siehe Kapitel 4.1.1 - 4.1.3) definiert wurden, werden berücksichtigt, wenn das Gesamtkollektiv der Studie diese Einschlusskriterien nicht vollständig erfüllt.“*

1.7 Lage des Konfidenzintervalls bei Bewertung der klinischen Relevanz (ad 4.4)

Wenn Responderauswertungen nicht verfügbar sind, geht das IQWiG nur dann von einer klinisch relevanten Effektstärke aus, wenn das Konfidenzintervall vollständig oberhalb der Irrelevanzschwelle liegt. Wir schlagen folgendes, besser geeignetes Kriterium zum Nachweis der klinischen Relevanz vor: Der beobachtete Effekt liegt oberhalb der Relevanzschwelle (d.h. des Schwellenwerts für die klinische Relevanz) und die Null liegt nicht im Konfidenzintervall, d.h. ein statistisch signifikanter Effekt wurde beobachtet. Nachfolgend die Begründung für diese Änderung:

Die klinische Relevanz von Therapieeffekten wird in neuerer Zeit intensiv diskutiert (z.B. auf dem Symposium „IQWiG im Dialog“ und dem Workshop „Meta-Analysen bei der Arzneimittelbewertung“ auf der Tagung der Deutschen Gesellschaft für Medizinische Informatik, Biometrie und Epidemiologie (GMDS) 2010). Eine systematische Sichtung, Klassifikation und Bewertung der methodischen Konzepte wurde von S. Thomas auf der GMDS-Tagung vorgestellt⁵³. Insbesondere die Einführung eines „minimalen Relevanzdelta“ δ_{\min} (unterhalb dieser Schwelle ist der Effekt sicher irrelevant) gibt Probleme auf. Eine solche „Irrelevanzschwelle“ gilt beim einzelnen Patienten in der Responderanalyse als sinnvoll, bei der Anwendung auf Gruppenebene aber als problematisch. Dieses δ_{\min} ist aber kaum festzulegen. Vorschläge gehen von $\delta_{\min}=0$ bis $\delta_{\min}=\delta_{\text{rel}}$ der Relevanzschwelle. Das IQWiG schlägt vor, bei bekannter Relevanzschwelle δ_{rel} daraus ein zugehöriges δ_{\min} zu berechnen. Wir teilen den von anderer Seite begründeten Standpunkt, dass $\delta_{\min}=0$ die einzige klinisch vertretbare Wahl der Irrelevanzgrenze auf Gruppenniveau ist. In diesem Sinne ist das vorgeschlagene Kriterium zum Nachweis der klinischen Relevanz ausreichend.

Wird gefordert, dass die untere Konfidenzgrenze über einem bestimmten Schwellenwert $\delta_{\min}>0$ liegt, so ist dieses Kriterium normalerweise nicht in die Planung der zulassungsrelevanten klinischen Studien eingegangen, da diese auf statistische Signifikanz (d.h. Null nicht im Konfidenzintervall) geplant sind. Damit wäre die Power eines solchen Kriteriums unverträglich klein. Post-hoc Kriterien für den „Erfolg“ von klinischen Studien festzusetzen ist statistisch zumindest fragwürdig.

Das vorgeschlagene Kriterium für die klinische Relevanz ist konsistent mit Kieser und Hauschke (2005)⁵⁴, die den sogenannten *probably clinically significant effect* empfehlen: Das Ergebnis ist signifikant und der beobachtete Effekt liegt oberhalb des Schwellenwertes. Die untere Konfidenzgrenze kann dabei unterhalb des Schwellenwertes liegen.

Fazit: Das vorgeschlagene Kriterium ist für die Bewertung der klinischen Relevanz besser geeignet als die Einführung einer schwer zu rechtfertigenden Irrelevanzschwelle. Daher bitten wir nachdrücklich um folgende Änderung im Kapitel 4.4: Der zweimal im Kapitel vorkommende Halbsatz „Liegt das zum beobachteten Effekt korrespondierende Konfidenzintervall vollständig oberhalb dieser Irrelevanzschwelle“ wird durch folgenden

Halbsatz ersetzt „Ist das Ergebnis statistisch signifikant und liegt der beobachtete Effekt oberhalb dieser Relevanzschwelle“.

1.8 Ergänzung des Jahrs des Studienbeginns als Effektmodifikator (ad 4.4.4)

Als mögliche Effektmodifikatoren werden Geschlecht, Alter und Art der Vorbehandlung genannt. Der Effektmodifikator Jahr des Studienbeginns sollte ergänzt werden. Nachfolgend die Begründung für diese Änderungen:

Seit der Zulassung des ersten Biologicals zur Behandlung der RA (in Deutschland im Jahr 1999) haben sich die Therapiealgorithmen, die Diagnostik und die Fachkenntnisse der behandelnden Ärzte innerhalb und außerhalb der Kliniken stark gewandelt. Das hat zur Folge, dass sich auch die demografischen Daten und das Erkrankungsprofil der in klinische Studien eingeschlossenen Patienten deutlich verändert haben⁵⁵. Insbesondere die Patienten mit IR (inadequate response) auf sDMARDs oder TNF-alpha-Inhibitoren (sDMARD-IR- und TNF-IR-Patienten) sind über die Jahre mehr und mehr austherapiert und sprechen daher auf neue Therapien schlechter an. Es ist daher zu erwarten, dass die Ergebnisse von neueren Studien eher schlechter ausfallen werden als solche von älteren Studien.

Tendenz von einer sDMARD-Monotherapie hin zu einer Kombinationstherapie mit unterschiedlichen Kombinationspartnern

Wurden in Deutschland im Jahr 1997 nur ca. 8 % der RA-Patienten mit einer sDMARD-Kombinationstherapie behandelt, waren es 2007 bereits ca. 23 %. Zudem erhöhte sich zwischen den Jahren 2000 und 2007 der Anteil der mit Leflunomid behandelten Patienten von 5 % auf 12,2 %. Cyclooxygenasen (COX)-Inhibitoren wurden hingegen im gleichen Zeitraum erst verstärkt und dann wieder weniger häufig eingesetzt (2000: 10,7% / 2003: 20,3% / 2007: 8,3%). Nichtsteroidale Antirheumatika (NSAR) wiederum wurden über die Jahre sehr unterschiedlich aber eher weniger eingenommen (1998: 48,8% / 2004: 25,2% / 2007: 36,4%)⁵⁶. Hinzu kommt, dass in RA-Zentren Methotrexat (MTX) erst ab Mitte der 90iger-Jahre international vermehrt eingesetzt wurde. Zuvor dominierten Sulfasalazin und intramuskuläres Gold. Zudem wird heute MTX früher und aggressiver eingesetzt, noch bevor es zu Gelenkschäden gekommen ist⁵⁷.

Veränderte Demografie der RA-Patienten

Studien mit sDMARD-IR- oder TNF-IR-Patienten legen meist als ein Einschlusskriterium ≥ 6 SJC / TJC fest. Der Anteil dieser Patienten an den RA-Patienten sank von Jahr zu Jahr dank der optimierten Therapie von 43 % / 46 % (1997) auf 8 % / 16 % (2007)⁵⁶. Folglich sind Patienten, die heute eine mittelgradige oder schwere RA

haben und nicht mehr auf sDMARDs und/oder TNF-alpha-Inhibitoren ansprechen, deutlich intensiver vorbehandelt, haben vermehrt Gelenkerosionen und somit einen damit einhergehenden irreversiblen HAQ-Score. Diese Patienten werden daher auch in klinischen Studien weniger auf eine Prüfmedikation ansprechen.

Fazit: Bei der Bewertung und dem Vergleich von Studienergebnissen sollte das Jahr des Studienbeginns beachtet werden. In Abhängigkeit vom Jahr des Studienbeginns und der damit zusammenhängenden Vorbehandlung können sich die Charakteristika der Studienpopulation deutlich unterscheiden. Deshalb bitten wir, dass in Kapitel 4.4.4 das Jahr des Studienbeginns als weiterer Effektmodifikator ergänzt wird.

1.9 Verlängerung der Frist für die Stellungnahme zum Vorbericht (ad 3.1)

In seinem Leitfaden zum Stellungnahmeverfahren zu Vorberichten sieht das IQWiG eine Stellungnahmefrist von mindestens 4 Wochen vor. Für den Vorbericht A10-01 ist eine Verlängerung der Stellungnahmefrist auf mindestens 8 Wochen wünschenswert.

Die Erstellung einer Stellungnahme für einen Vorbericht ist sehr aufwändig. Die Gründe dafür sind zum einen die Komplexität und der Umfang der vorläufigen Nutzenbewertung und zum anderen die Vielzahl der zu bearbeitenden Punkte, v.a. die Prüfung und Erarbeitung von Änderungswünschen inklusive Begründung für

- die Umsetzung der im Berichtsplan beschriebenen Methodik
- die Bewertung der Studienqualität
- die Bewertung und Interpretation der eingeschlossenen Studien
- die Bewertung und Interpretation der Ergebnisse der Meta-Analysen (inklusive der klinischen Relevanz)
- die Schlussfolgerungen.

Darüber hinaus muss in die Stellungnahme auch die Expertise der Kollegen auf internationaler Ebene einfließen, die an der Planung, Durchführung und Auswertung der klinischen Studien wesentlich beteiligt waren. Dazu müssen der Vorbericht und relevante Entwürfe der Stellungnahme übersetzt werden. Diese Übersetzungen sind zeitintensiv.

Fazit: Die Erstellung einer Stellungnahme für einen Vorbericht ist sehr aufwändig. Daher bitten wir nachdrücklich, dass folgender Absatz in Kapitel 3.1 eingefügt wird: „Die Frist für die Stellungnahme zum Vorbericht beträgt bei diesem Projekt acht Wochen.“

2 Verweise auf qualitativ angemessene Unterlagen, insbesondere bislang unpublizierte Daten

2.1 Berücksichtigung der aktuellen Fachinformationen

Wir bitten darum, dass die aktuelle Fachinformation von RoACTEMRA[®] (Stand Juni 2010)⁵⁸ berücksichtigt wird.

2.2 Head-to-Head-Studien von Roche

Wir möchten darauf hinweisen, dass Roche

- eine Head-to-Head-Studie (Monotherapie; WA19924) kürzlich gestartet hat, sowie
- eine weitere Head-to-Head-Studie in der Kombinationstherapie im TNF-IR-Segment und
- eine Head-to-Head-Safety-Studie mit Schwerpunkt Cardiovascular Outcome plant.

Die Head-to-Head-Studie in der Monotherapie hat folgenden Titel:

A multi-center, randomized, blinded, parallel-group study of the reduction of signs and symptoms during monotherapy treatment with Tocilizumab 8 mg/kg intravenously versus adalimumab 40 mg subcutaneously in patients with rheumatoid arthritis (interne Studien-Nr.: WA19924, siehe auch Studienliste zu Tocilizumab)

Die Ergebnisse dieser Studien werden wichtige Evidenz zur Beurteilung der biotechnologisch hergestellten Antirheumatika generieren. Wir bitten Sie daher, diese Studien bei der Beurteilung der Evidenzlage zu vermerken.

3 Literaturverzeichnis

1. Aletaha, D; Ward, MM; Machold, KP; Nell, VP; Stamm, T; Smolen, JS: **Remission and active disease in rheumatoid arthritis: defining criteria for disease activity states.** In: Arthritis Rheum 52 (2005) Nr. 9, S. 2625-2636.
2. Aletaha, D; Smolen, JS: **The Simplified Disease Activity Index (SDAI) and the Clinical Disease Activity Index (CDAI): A review of their usefulness and validity in rheumatoid arthritis.** In: Clinical Exp Rheumatol 23 (Suppl.39) (2005), S. 100-108.
3. Duhm, B; Kraus, A; Kutil, B; Cerny, H; Mustak, M; Walter, M; Gaugg, M; Erlacher, L: **Aktivitätsscores bei chronischer Polyarthritis und Morbus Bechterew.** 2007, <http://www.springermedizin.at/img/db/docs/1898.pdf> (21.10.10).
4. Prevoo, ML; van 't Hof, MA; Kuper, HH; van Leeuwen, MA; van de Putte, LB; van Riel, PL: **Modified disease activity scores that include twenty-eight-joint counts. Development and validation in a prospective longitudinal study of patients with rheumatoid arthritis.** In: Arthritis Rheum. 38 (1995) Nr. 1, S. 44-48.
5. van der Heijde, DM; van 't, HM; van Riel, PL; van de Putte, LB: **Development of a disease activity score based on judgment in clinical practice by rheumatologists.** In: J.Rheumatol 20 (1993) Nr. 3, S. 579-581.
6. Smolen, JS; Aletaha, D; Bijlsma, JW; Breedveld, FC; Boumpas, D; Burmester, G; Combe, B; Cutolo, M; de, WM; Dougados, M; Emery, P; Gibofsky, A; Gomez-Reino, JJ; Haraoui, B; Kalden, J; Keystone, EC; Kvien, TK; McInnes, I; Martin-Mola, E; Montecucco, C; Schoels, M; van der Heijde, D: **Treating rheumatoid arthritis to target: recommendations of an international task force.** In: Ann Rheum Dis 69 (2010) Nr. 4, S. 631-637.
7. Smolen, JS; Han, C; van der Heijde, DM; Emery, P; Bathon, JM; Keystone, E; Maini, RN; Kalden, JR; Aletaha, D; Baker, D; Han, J; Bala, M; St Clair, EW: **Radiographic changes in rheumatoid arthritis patients attaining different disease activity states with methotrexate monotherapy and infliximab plus methotrexate: the impacts of remission and tumour necrosis factor blockade.** In: Ann.Rheum.Dis. 68 (2009) Nr. 6, S. 823-827.
8. Welsing, PM; Landewe, RB; van Riel, PL; Boers, M; van Gestel, AM; van der Linden, S; Swinkels, HL; van der Heijde, DM: **The relationship between disease activity and radiologic progression in patients with rheumatoid arthritis: a longitudinal analysis.** In: Arthritis Rheum 50 (2004) Nr. 7, S. 2082-2093.
9. Smolen, JS; Aletaha, D: **Developments in the clinical understanding of rheumatoid arthritis.** In: Arthritis Res Ther 11 (2009) Nr. 1, S. 204.
10. Aletaha, D; Neogi, T; Silman, AJ; Funovits, J; Felson, DT; Bingham, CO, III; Birnbaum, NS; Burmester, GR; Bykerk, VP; Cohen, MD; Combe, B; Costenbader, KH; Dougados, M; Emery, P; Ferraccioli, G; Hazes, JM; Hobbs, K; Huizinga, TW; Kavanaugh, A; Kay, J; Kvien, TK; Laing, T; Mease, P; Menard, HA; Moreland, LW; Naden, RL; Pincus, T; Smolen, JS; Stanislawski-Biernat, E; Symmons, D; Tak, PP; Upchurch, KS; Vencovsky, J; Wolfe, F; Hawker, G: **2010 Rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative.** In: Arthritis Rheum 62 (2010) Nr. 9, S. 2569-2581.

11. Rau, R; Genth, E: **Wirksamkeitskriterien der Therapie bei rheumatischen Erkrankungen**. 2007, <http://dgrh.de/qualitaetsmanual4.html> (21.10.10).
12. Emery, P; Gabay, C; Kraan, M; Gomez-Reino, J: **Evidence-based review of biologic markers as indicators of disease progression and remission in rheumatoid arthritis**. In: Rheumatol Int 27 (2007) Nr. 9, S. 793-806.
13. Boyesen, P; Hoff, M; Odegard, S; Haugeberg, G; Syversen, SW; Gaarder, PI; Okkenhaug, C; Kvien, TK: **Antibodies to cyclic citrullinated protein and erythrocyte sedimentation rate predict hand bone loss in patients with rheumatoid arthritis of short duration: a longitudinal study**. In: Arthritis Res Ther 11 (2009) Nr. 4, S. R103.
14. Lavie, CJ; Milani, RV; Verma, A; O'Keefe, JH: **C-reactive protein and cardiovascular diseases--is it ready for primetime?** In: Am.J Med.Sci. 338 (2009) Nr. 6, S. 486-492.
15. Ridker, PM: **Clinical application of C-reactive protein for cardiovascular disease detection and prevention**. In: Circulation 107 (2003) Nr. 3, S. 363-369.
16. Gemeinsamer Bundesausschuss: **Bekanntmachungen eines Beschlusses des Gemeinsamen Bundesausschusses über eine Änderung der Arzneimittel-Richtlinie/ AMR in Anlage 4: Therapiehinweis zu Adalimumab vom 21. November 2006**. 2007, http://www.g-ba.de/downloads/39-261-346/2006-11-21-AMR4-Adalimumab_BAnz.pdf
17. IQWiG: **Glitazone zur Behandlung des Diabetes mellitus Typ 2. Dokumentation und Würdigung der Stellungnahmen zum Berichtsplan A05-05A Version 1.0**. 2007, <http://www.iqwig.de/index.139.html> (22.4.08).
18. CADTH: **Biological Response Modifier Agents for Adults with Rheumatoid Arthritis**. 2010, http://www.cadth.ca/media/pdf/TR_RA_Clinical_and_Economic_Overview_e.pdf (5.8.10).
19. Donahue, KE; Gartlehner, G; Jonas, DE; Lux, LJ; Thieda, P; Jonas, B; Hansen, RA; Morgan, LC; Williams, SC; Lohr, KN: **Comparative Effectiveness of Drug Therapy for Rheumatoid Arthritis and Psoriatic Arthritis in Adults. Comparative Effectiveness Review No. 11**. 2007, <http://www.effectivehealthcare.ahrq.gov/reports/final.cfm>. (5.8.10).
20. Malottki, K; Barton, P; Tsourapas, A; Uthman, A; Liu, Z; Routh, K; Connock, M; Jobanputra, P; Moore, D; Fry-Smith, A; Chen, Y-F: **Adalimumab, etanercept, infliximab, rituximab and abatercept for the treatment of rheumatoid arthritis after failure of a TNF inhibitor: a systematic review and economic evaluation**. 2009, <http://www.nice.org.uk/nicemedia/live/12135/47685/47685.pdf> (5.8.10).
21. Meads, C; Jit, M; Tsourapas, A; Ashfaq, K; Connock, M; Fry-Smith, A; Jobanputra, P: **Tocilizumab for the treatment of rheumatoid arthritis. A single technology appraisal. Evidence Review Group Report**. 2009, <http://www.nice.org.uk/nicemedia/live/12033/45575/45575.pdf> (5.8.10).
22. Lopez-Olivo, MA; Amezcaga, M; McGahan, L; Suarez-Almazor, ME: **Rituximab for rheumatoid arthritis**. In: Cochrane Database of Systematic Reviews: Protocols (CD007356)(2008) Nr. 4,
23. Singh, JA; Berg, S; Lopez-Olivo, MA: **Tocilizumab for rheumatoid arthritis**. In: Cochrane Database of Systematic Reviews (CD008331)(2010) Nr. 1,

24. Singh, JA; Christensen, R; Wells, GA; Suarez-Almazor, ME; Buchbinder, R; Lopez-Olivo, MA; Tanjong, GE; Tugwell, P: **Biologics for rheumatoid arthritis: an overview of Cochrane reviews**. In: Cochrane Database Syst Rev (CD007848)(2010) Nr. 4,
25. Combe, B; Cantagrel, A; Goupille, P; Bozonnat, MC; Sibilia, J; Eliaou, JF; Meyer, O; Sany, J; Dubois, A; Daures, JP; Dougados, M: **Predictive factors of 5-year health assessment questionnaire disability in early rheumatoid arthritis**. In: J Rheumatol 30 (2003) Nr. 11, S. 2344-2349.
26. Dawes, PT; Fowler, PD; Clarke, S; Fisher, J; Lawton, A; Shadforth, MF: **Rheumatoid arthritis: treatment which controls the C-reactive protein and erythrocyte sedimentation rate reduces radiological progression**. In: Br J Rheumatol 25 (1986) Nr. 1, S. 44-49.
27. Drossaers-Bakker, KW; de, BM; van, ZD; Zwinderman, AH; Breedveld, FC; Hazes, JM: **Long-term course and outcome of functional capacity in rheumatoid arthritis: the effect of disease activity and radiologic damage over time**. In: Arthritis Rheum 42 (1999) Nr. 9, S. 1854-1860.
28. Smolen, JS; van der Heijde, DM; St Clair, EW; Emery, P; Bathon, JM; Keystone, E; Maini, RN; Kalden, JR; Schiff, M; Baker, D; Han, C; Han, J; Bala, M: **Predictors of joint damage in patients with early rheumatoid arthritis treated with high-dose methotrexate with or without concomitant infliximab: results from the ASPIRE trial**. In: Arthritis Rheum 54 (2006) Nr. 3, S. 702-710.
29. van der Heijde, DM; van Riel, PL; van Leeuwen, MA; van 't Hof, MA; van Rijswijk, MH; van de Putte, LB: **Prognostic factors for radiographic damage and physical disability in early rheumatoid arthritis. A prospective follow-up study of 147 patients**. In: Br J Rheumatol 31 (1992) Nr. 8, S. 519-525.
30. van Leeuwen, MA; van der Heijde, DM; van Rijswijk, MH; Houtman, PM; van Riel, PL; van de Putte, LB; Limburg, PC: **Interrelationship of outcome measures and process variables in early rheumatoid arthritis. A comparison of radiologic damage, physical disability, joint counts, and acute phase reactants**. In: J Rheumatol 21 (1994) Nr. 3, S. 425-429.
31. Schneider, M; Lelgemann, M; Abholz, H; Caratti, R; Flügge, C; Jäniche, H; Kunz, R; Krüger, K; Rehart, S; Specker, C: **Interdisziplinäre Leitlinie Management der frühen rheumatoiden Arthritis, 2. überarbeitete Auflage**. Steinkopff Darmstadt, 2006.
32. National Institute for Health and Clinical Excellence: **Rheumatoid arthritis - The management of rheumatoid arthritis in adults**. 2009, <http://www.nice.org.uk/nicemedia/live/12131/43327/43327.pdf> (30.9.10).
33. Haute Autorité de Santé: **Rheumatoid arthritis: Diagnosis and initial management**. 2007, http://www.has-sante.fr/portail/upload/docs/application/pdf/2009-10/ra_diagnosis_and_initial_management-guidelines.pdf (30.9.10).
34. European Medicines Agency: **Concept paper on the need for revision of the points to consider on the clinical investigation of medicinal products other than NSAIDs in rheumatoid arthritis (CPMP/EWP/556/95 REV. 1)**. 2010, http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2010/02/WC500070422.pdf (30.9.10).
35. The European Agency for the Evaluation of Medicinal Products: **Points to consider on clinical investigation of medicinal products other than NSAIDs for treatment of rheumatoid**

arthritis.

2003,

http://www.ema.europa.eu/docs/en_GB/document_library/Scientific_guideline/2009/09/WC50003439.pdf (30.9.10).

36. Felson, DT; Anderson, JJ; Lange, ML; Wells, G; LaValley, MP: **Should improvement in rheumatoid arthritis clinical trials be defined as fifty percent or seventy percent improvement in core set measures, rather than twenty percent?** In: Arthritis Rheum 41 (1998) Nr. 9, S. 1564-1570.
37. Harrington, JT: **The uses of disease activity scoring and the physician global assessment of disease activity for managing rheumatoid arthritis in rheumatology practice.** In: J Rheumatol 36 (2009) Nr. 5, S. 925-929.
38. Studenic, P; Smolen, JS; Aletaha, D: **Perception of RA Disease Activity by Patients and Physicians: Reason for and Estimators of Discrepancies.** Poster presented at ACR 2010 (2010) .
39. Bruce, B; Fries, JF: **The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation.** In: J Rheumatol 30 (2003) Nr. 1, S. 167-178.
40. Wolfe, F: **A reappraisal of HAQ disability in rheumatoid arthritis.** In: Arthritis Rheum 43 (2000) Nr. 12, S. 2751-2761.
41. Wolfe, F: **The determination and measurement of functional disability in rheumatoid arthritis.** In: Arthritis Res 4 Suppl 2 (2002), S. S11-S15.
42. Repping-Wuts, H; van, RP; van, AT: **Fatigue in patients with rheumatoid arthritis: what is known and what is needed.** In: Rheumatology.(Oxford.) 48 (2009) Nr. 3, S. 207-209.
43. Maurischat, C; Ehlebracht-Konig, I; Kuhn, A; Bullinger, M: **[Structural validity of the Short Form 36 (SF-36) in patients with rheumatic diseases].** In: Z.Rheumatol 64 (2005) Nr. 4, S. 255-264.
44. Smolen, JS; Landewe, R; Breedveld, FC; Dougados, M; Emery, P; Gaujoux-Viala, C; Gorter, S; Knevel, R; Nam, J; Schoels, M; Aletaha, D; Buch, M; Gossec, L; Huizinga, T; Bijlsma, JW; Burmester, G; Combe, B; Cutolo, M; Gabay, C; Gomez-Reino, J; Kouloumas, M; Kvien, TK; Martin-Mola, E; McInnes, I; Pavelka, K; van, RP; Scholte, M; Scott, DL; Sokka, T; Valesini, G; van, VR; Winthrop, KL; Wong, J; Zink, A; van der Heijde, D: **EULAR recommendations for the management of rheumatoid arthritis with synthetic and biological disease-modifying antirheumatic drugs.** In: Ann Rheum Dis 69 (2010) Nr. 6, S. 964-975.
45. Boers, M: **The time has come to limit the placebo period in rheumatoid arthritis trials to 3 months: a systematic comparison of 3- and 6-month response rates in trials of biological agents.** In: Ann Rheum Dis 69 (2010) Nr. 1, S. 186-192.
46. Aletaha, D; Funovits, J; Keystone, EC; Smolen, JS: **Disease activity early in the course of treatment predicts response to therapy after one year in rheumatoid arthritis patients.** In: Arthritis Rheum 56 (2007) Nr. 10, S. 3226-3235.
47. Goekoop-Ruiterman, YP; de Vries-Bouwstra, JK; Allaart, CF; van, ZD; Kerstens, PJ; Hazes, JM; Zwinderman, AH; Rooday, HK; Han, KH; Westedt, ML; Gerards, AH; van Groenendaal, JH; Lems, WF; van Krugten, MV; Breedveld, FC; Dijkmans, BA: **Clinical and radiographic outcomes of four different treatment strategies in patients with early rheumatoid**

- arthritis (the BeSt study): A randomized, controlled trial.** In: Arthritis Rheum 58 (2005) Nr. 2 Suppl, S. S126-S135.
48. Grigor, C; Capell, H; Stirling, A; McMahon, AD; Lock, P; Vallance, R; Kincaid, W; Porter, D: **Effect of a treatment strategy of tight control for rheumatoid arthritis (the TICORA study): a single-blind randomised controlled trial.** In: Lancet. 364 (2004) Nr. 9430, S. 263-269.
49. Verstappen, SM; Jacobs, JW; van der Veen, MJ; Heurkens, AH; Schenk, Y; ter Borg, EJ; Blaauw, AA; Bijlsma, JW: **Intensive treatment with methotrexate in early rheumatoid arthritis: aiming for remission. Computer Assisted Management in Early Rheumatoid Arthritis (CAMERA, an open-label strategy trial).** In: Ann Rheum Dis 66 (2007) Nr. 11, S. 1443-1449.
50. IQWiG: **Vergleichende Nutzenbewertung verschiedener antihypertensiver Wirkstoffgruppen als Therapie der ersten Wahl bei Patienten mit essentieller Hypertonie, Abschlussbericht A05-09.** 2009, http://www.iqwig.de/download/A05-09_Abschlussbericht_Antihypertensive_Wirkstoffgruppen_als_Therapie_der_ersten_Wahl.pdf (26.7.10).
51. IQWiG: **Memantin bei Alzheimer Demenz, Abschlussbericht A05-19C.** 2009, http://www.iqwig.de/download/A05-19C_Abschlussbericht_Memantin_bei_Alzheimer_Demenz.pdf (26.7.10).
52. Janatzek, S: **Aussagekraft von Subgruppenanalysen. G-2 Gutachten Fachbereich Evidenzbasierte Medizin.** 2004, <http://www.mds-ev.de/media/pdf/subgruppen-gutachten.pdf> (20.8.10).
53. Thomas, S: **Klinische Relevanz von Therapieeffekten: Systematisch Sichtung, Klassifizierung und Bewertung methodischer Konzepte. Vortrag auf der 55. GMDS Jahrestagung am 6.9.2010.** 2010, <http://www.gmds2010.de/cms/wordpress/wp-content/uploads/fohlen/ws1-thomas.pdf> (20.10.10).
54. Kieser, M; Hauschke, D: **Assessment of clinical relevance by considering point estimates and associated confidence intervals.** In: Pharmaceut Statist 4 (2005) Nr. 2, S. 101-107.
55. Uhlig, T; Kvien, TK: **Is rheumatoid arthritis really getting less severe?** In: Nat.Rev.Rheumatol 5 (2009) Nr. 8, S. 461-464.
56. 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.** In: Ann Rheum Dis 69 (2010) Nr. 10, S. 1803-1808.
57. Sokka, T; Envalds, M; Pincus, T: **Treatment of rheumatoid arthritis: a global perspective on the use of antirheumatic drugs.** In: Mod.Rheumatol 18 (2008) Nr. 3, S. 228-239.
58. Roche; Chugai: **Fachinformation RoActemra (Stand Juni 2010).** 2010, www.fachinfo.de (5.7.10).
59. Hafstrom, I; Albertsson, K; Boonen, A; van der Heijde, D; Landewe, R; Svensson, B: **Remission achieved after 2 years treatment with low-dose prednisolone in addition to disease-modifying anti-rheumatic drugs in early rheumatoid arthritis is associated with reduced joint destruction still present after 4 years: an open 2-year continuation study.** In: Ann.Rheum.Dis. 68 (2009) Nr. 4, S. 508-513.

60. St Clair, EW; van der Heijde, DM; Smolen, JS; Maini, RN; Bathon, JM; Emery, P; Keystone, E; Schiff, M; Kalden, JR; Wang, B; Dewoody, K; Weiss, R; Baker, D: **Combination of infliximab and methotrexate therapy for early rheumatoid arthritis: a randomized, controlled trial.** In: Arthritis Rheum. 50 (2004) Nr. 11, S. 3432-3443.
61. van der Heijde, D; Klareskog, L; Landewe, R; Bruyn, GA; Cantagrel, A; Durez, P; Herrero-Beaumont, G; Molad, Y; Codreanu, C; Valentini, G; Zahora, R; Pedersen, R; MacPeck, D; Wajdula, J; Fatenejad, S: **Disease remission and sustained halting of radiographic progression with combination etanercept and methotrexate in patients with rheumatoid arthritis.** In: Arthritis Rheum. 56 (2007) Nr. 12, S. 3928-3939.
62. Svensson, B; Boonen, A; Albertsson, K; van der Heijde, D; Keller, C; Hafstrom, I: **Low-dose prednisolone in addition to the initial disease-modifying antirheumatic drug in patients with early active rheumatoid arthritis reduces joint destruction and increases the remission rate: a two-year randomized trial.** In: Arthritis Rheum. 52 (2005) Nr. 11, S. 3360-3370.
63. IQWiG: **Allgemeine Methoden. Version 3.0.** 2008, http://www.iqwig.de/download/IQWiG_Methoden_Version_3_0.pdf (26.7.10).

4 Anhang

4.1 Tabelle 2: Internationaler Standard

Review	Reviewtyp	Betrachtung von patientenrelevanten Endpunkten / Messinstrumenten							Studiendauer	
		Krankheitsaktivität	ACR20	ACR50	ACR70	DAS oder DAS28	Krankheitsremission anhand DAS oder DAS28	EULAR Response-Kriterien	HAO	Mindeststudiedauer in Wochen (entsprechend den Ein- und Ausschlusskriterien)
Lopez et al. 2008	Cochrane Review (Protokoll zur Methodik)	ja	ja	ja	ja	ja	nein	ja	ja	24 (6 Monate)
Singh et al. 2009	Cochrane Review	ja	ja	ja	ja	ja	ja	ja	ja	Keine Beschränkung
Donahue et al. 2007	Comparative Effectiveness Review AHRO	ja	ja	ja	ja	ja	ja	ja	ja	12 (3 Monate)
CADTH 2010	CADTH Therapeutic Review	ja	ja	ja	ja	nein	nein	nein	ja	24
Malotki et al. 2009	NICE HTA	ja	ja	ja	ja	ja	ja	ja	ja	12
Singh et al 2010	Cochrane Review	ja	ja	ja	ja	ja	ja	ja	ja	Keine Beschränkung
Meads et al. 2009	NICE: single HTA	ja	ja	ja	ja	ja	ja	nein	ja	Keine Beschränkung
Zusammenfassung		7-mal ja	7-mal ja	7-mal ja	7-mal ja	6-mal ja, 1-mal nein	5-mal ja, 2-mal nein	5-mal ja, 2-mal nein	7-mal ja	3-mal keine Beschränkung; 2-mal 12 Wochen; 2-mal 24 Wochen

AHRQ: Agency for Healthcare Research and Quality; CADTH: Canadian Agency for Drugs and Technologies in Health; NICE: National Institute for Health and Clinical Excellence; k.A.: keine Angabe; n.r.: nicht relevant

Quellen: Lopez et al. 2008, Singh et al. 2010, Donahue et al. 2007, CADTH 2010, Malotki et al. 2009, Singh et al 2010, Meads et al. 2009^{18, 19, 20, 21, 22, 23, 24}

4.2 Remissionskriterium $DAS28 < 2,6$ als valider Surrogat für radiographische Progressionshemmung

Im vorläufigen Berichtsplan (S. 2) sieht das IQWiG die Remission als primäres Ziel der Behandlung der RA. Dazu führt das IQWiG weiter aus: „Untersuchungen der letzten Jahre haben gezeigt, dass nur eine Remission und damit eine weitgehende Verminderung der Krankheitsaktivität das Fortschreiten der Gelenkzerstörung aufhalten kann [5,19,20].“ Basierend auf dieser Feststellung des IQWiG wird der Zusammenhang zwischen der Remission, definiert als „ $DAS28 < 2,6$ “, und der Verhinderung des Fortschreitens der Gelenkzerstörung (radiographische Progressionshemmung) analysiert.

Ausgangspunkt der Analyse ist die vom IQWiG zitierte (Quelle [20]) systematische Übersichtsarbeit zum Zusammenhang der Remission mit der radiographischen Progression. Aus dieser Übersichtsarbeit wurden die Studien ausgewählt, die die folgenden Kriterien erfüllen:

- Definition der Remission gemäß $DAS28 < 2,6$
- RCT
- Dauer der Remission: „Minimum ein Zeitpunkt“

Aus diesen Studien wurden die Daten zum prozentualen Anteil der Patienten mit einer $DAS28$ -Remission und zum prozentualen Anteil der Patienten mit Progressionshemmung (keine bzw. keine messbare radiographische Progression) am Studienende extrahiert, graphisch dargestellt und analysiert.

Ergebnis:

Drei Studien erfüllten die Kriterien: die BARFOT-Studie (Hafström et al 2009; Svensson et al. 2005), die TEMPO-Studie (van der Heijde et al. 2007) und die ASPIRE-Studie (Smolen et al. 2009; St.Clair 2004)^{7, 59, 60, 61, 62}.

Diese drei Studien unterscheiden sich in wichtigen Charakteristika (siehe Tabelle 3). Zwar wird in allen drei Studien die radiographische Progression mit der von van der Heijde modifizierten Sharp-Methode (total sharp score (TSS)) gemessen. Dennoch werden unterschiedliche Grenzen für die radiographische Progressionshemmung verwendet: die kleinste messbare Differenz beträgt 5,8 in der BARFOT-Studie und 9,03 in der ASPIRE-Studie; in der TEMPO-Studie wird für ein Anhalten der Progression die Grenze 0,5 verwendet. Darüber hinaus ist die Studiendauer (1, 2 bzw. 3 Jahre) und die mittlere Erkrankungsdauer der Patienten (<1 Jahr bzw. 7 Jahre) verschieden. Aufgrund dieser Unterschiede können die 3 Studien nicht zusammengefasst werden. Daher ist keine Meta-Regression möglich.

Tabelle 3: Wichtige Charakteristika der drei Studien

	BARFOT-Studie	TEMPO-Studie	ASPIRE-Studie
Grenze für die radiographische Progressionshemmung bzgl. des total sharp scores (TSS)	5,8 (kleinste messbare Differenz)	0,5 (Anhalten der Progression)	9,03 (kleinste messbare Differenz)
Studiendauer [Jahre]	2	3	1
Mittlere Erkrankungsdauer der Patienten zu Studienbeginn [Jahre]	0,5	7	0,9

Quellen: BARFOT-Studie: Svensson et al. 2005; TEMPO-Studie: van der Heijde et al. 2007; ASPIRE-Studie: St.Clair 2004^{60, 61, 62}

Die 3 Studien werden einzeln betrachtet. Die Ergebnisse dieser Studien werden in Tabelle 4 dargestellt und in der Abbildung 1 veranschaulicht. In der Abbildung wird für jeden Behandlungsarm der Studie dem prozentualen Anteil der Patienten mit einer DAS28-Remission (Wert auf der x-Achse) der entsprechende prozentuale Anteil der Patienten mit einer radiographischen Progressionshemmung (Wert auf der y-Achse) zugeordnet. Zusätzlich wird für jede Studie eine Regressionsgerade dargestellt. In allen 3 Studien gibt es eine strenge monotone Zunahme des prozentualen Anteils der Patienten mit einer radiographischen Progressionshemmung mit der Zunahme des prozentualen Anteils der Patienten mit einer DAS28-Remission (vgl. Abb. 1). Die Wahrscheinlichkeit bei unabhängigen Beobachtungen, zufällig einen solchen Befund zu bekommen ist $1/32$ (die Wahrscheinlichkeit, dass der Punkt des nächsten Behandlungsarms oberhalb des Punkts des Ausgangs-Behandlungsarms liegt, ist $1/2$; für die TEMPO- und ASPIRE-Studie beträgt die Wahrscheinlichkeit, dass der Punkt des 2. Behandlungsarms über dem Punkt des 1. Behandlungsarm und der Punkt des 3. Behandlungsarms über dem Punkt des 2. Behandlungsarm liegt jeweils $1/2 * 1/2 = 1/4$; die Gesamtwahrscheinlich für die 3 Studien beträgt $1/4 * 1/4 * 1/2 = 1/32$). Dieses entspricht bei robuster Modellierung (z.B. Vorzeichenstest) einer falsch-positiven Wahrscheinlichkeit von 0,031, also ist das Ergebnis statistisch signifikant auf dem 5%-Niveau.

Tabelle 4: Zusammenhang der DAS28-Remission mit der radiographischen Progressionshemmung

	Behandlungsarm 1	Behandlungsarm 2	Behandlungsarm 3
TEMPO-Studie	MTX	Eta	MTX+Eta
Prozentualer Anteil der Patienten mit DAS28<2,6 (nach 3 Jahren; LOCF)	18,9	20,6	40,3
Prozentualer Anteil der Patienten mit radiographischer Progressionshemmung (Δ TSS \leq 0,5; nach 3 Jahren)	51	61	76
BARFOT-Studie	Kein Prednisolon	Prednisolon	
Prozentualer Anteil der Patienten mit DAS28<2,6 (nach 2 Jahren)	32,8	55,5	
Prozentualer Anteil der Patienten mit radiographischer Progressionshemmung (Δ TSS \leq 5,8; nach 2 Jahren)*	60,7	74,1	
ASPIRE-Studie	MTX	MTX+Inf (3mg)	MTX+Inf (6mg)
Prozentualer Anteil der Patienten mit DAS28<2,6 (nach 1 Jahr)	15,0	21,2	31,0
Prozentualer Anteil der Patienten mit radiographischer Progressionshemmung (Δ TSS \leq 9,03; nach 1 Jahr)*	89,0	96,1	98,9

* Eigene Berechnung aus dem Kehrwert (100% - x); Eta: Etanercept; Inf: Infliximab; MTX: Methotrexat; LOCF: last observation carried forward; TSS: total sharp score;

Quellen: BARFOT-Studie: Svensson et al. 2005; TEMPO-Studie: van der Heijde et al. 2007; ASPIRE-Studie: St.Clair 2004^{60, 61, 62}

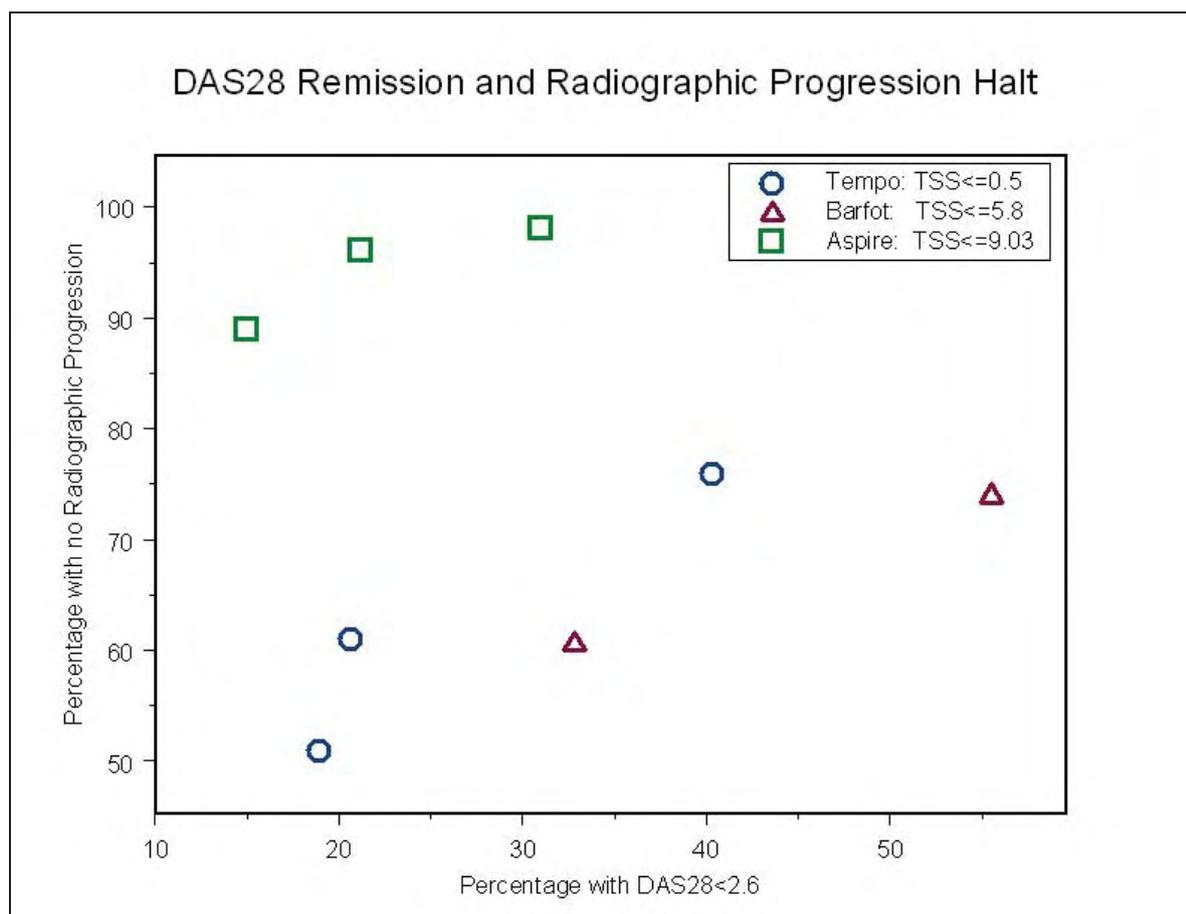


Abb. 1: Zusammenhang der DAS28-Remission mit der radiographischen Progressionshemmung

Die Ergebnisse zeigen die Validität des Surrogats „Remissionskriterium DAS28<2,6“. Zur Anerkennung der Validität von Surrogaten durch das IQWiG sind i.d.R. „eindeutige Belege aus Interventionsstudien eines plausiblen, starken, konsistenten und gleichgerichteten Zusammenhangs zwischen der Änderung des Surrogats und der Änderung des patientenrelevanten Endpunkts erforderlich“⁶³. Das Remissionskriterium DAS28<2,6 erfüllt diese Anforderungen des IQWiG: Der Zusammenhang zwischen der Änderung der DAS28-Remission und der Änderung des patientenrelevanten Endpunkts radiographische Progressionshemmung ist

- plausibel, weil die mit der DAS28-Remission einhergehende Unterdrückung der Gelenkentzündung zu einer verminderten Zytokinproduktion führt und dadurch die Osteoklastenaktivierung reduziert wird; dies verhindert beziehungsweise reduziert die radiologische Progression
- konsistent, weil es keine Ausnahme gibt
- gleichgerichtet, weil die Messwerte beider Parameter zunehmen
- stark, weil die relevanten Maßzahlen starke Unterschiede in beiden Parametern zeigen

Fazit: Das Remissionskriterium $\text{DAS28} < 2,6$ ist ein valider Surrogat für die radiographische Progressionshemmung.

A 1.8 UCB Pharma GmbH

Autoren:

Forst, Hans-Theo, Dr.

Hilf, Christof, Dr.

Kreutzenbeck, Hans-Joachim, Dr.



Stellungnahme der UCB Pharma GmbH zum vorläufigen Berichtsplan A10-01, Version 1.0 (Biologika – Zweitlinientherapie bei rheumatoider Arthritis, Stand: 20.09.2010)

Nach sorgfältigem Lesen des oben genannten vorläufigen Berichtsplans bitten wir das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) um die im Folgenden beschriebenen Änderungen bzw. Präzisierungen:

Auf Seite 7 sind unter 4.1.3 die zu berücksichtigenden patientenrelevanten Endpunkte aufgelistet. Hier sollte unter dem Punkt „Remission“ auch die Definition durch den *Disease Activity Score* (DAS-28) genannt werden. Als weitere Endpunkte sollten neben den bereits genannten Endpunkten „Symptomatik der Entzündung“ und „Strukturelle Gelenkveränderungen“ alle ACR-Scores eingefügt werden.

Die Scores des „*American College of Rheumatology*“ erfassen entsprechend ihrer Definition¹, patientenrelevante klinische Veränderungen. Hierbei sehen wir auch den ACR20 Score als zu berücksichtigenden Endpunkt an, da dieser beispielsweise mit der 20-prozentigen Verringerung der Zahl der druckschmerzhaften Gelenke eindeutig patientenrelevant ist. Die klinische Validität wird in einer Untersuchung des „*American College of Rheumatology*“ aus dem Jahre 2007 bestätigt².

Ebenso sollten radiologisch erfasste Gelenkveränderungen als valide Surrogatparameter als Endpunkte aufgenommen werden. Die Arbeiten von *van der Heijde et al.*, 2008 und *Breedveld et al.*, 2005 zeigen den Zusammenhang zwischen radiologisch nachgewiesenem Gelenkschaden und Verschlechterung der Gelenkfunktion^{3 4}. Ebenso konnten *Kavanaugh et al.*, 2004 die Korrelation zwischen radiologisch nachgewiesener Gelenkzerstörung und verminderter Lebensqualität, sowie die Korrelation zwischen Gelenkzerstörung und verringerter Teilhabe am beruflichen Leben belegen⁵.

Auf Seite 8 sollte der Begriff „interventions- und erkrankungsbedingter Aufwand“ präzisiert werden, da nicht klar ist, was hier gemeint ist: handelt es sich zum Beispiel um die Zahl der Arztbesuche, die Zahl der Arzneimittelanwendungen oder um die Zahl der Klinikeinweisungen?

Im Abschnitt 4.1.4 auf Seite 8 wird erklärt, dass ausschließlich RCTs in die Nutzenbewertung einfließen werden. Eine Beschränkung auf RCTs halten wir für unangemessen, da auch aus Non-RCTs wichtige Erkenntnisse über den Nutzen eines Arzneimittels gewonnen werden können. Das gilt insbesondere für Langzeitstudien. Als Beispiel sei hier die Arbeit von *Pincus* (1995) genannt⁶, in der beschrieben wird, dass neben RCTs auch zusätzlich Beobachtungsstudien in Betracht gezogen werden sollten. Zwar sind Non-RCTs wegen fehlender Placebokontrolle nicht völlig frei von Verzerrungen, doch liefern auch Non-RCTs relevante Evidenz.

Für Certolizumab Pegol konnte in der „open-label-extension“ der RAPID 1 Studie gezeigt werden, dass über weitere 50 Wochen eine nachhaltige klinische Wirksamkeit, und in diesem Fall ein patientenrelevanter Nutzen vorhanden ist⁷. Ein solches Ergebnis sollte in die Gesamtbewertung einer Substanz unter Berücksichtigung des niedrigeren Evidenzgrades (d.h. also mit niedrigerer Gewichtung) einfließen.



Ebenfalls auf Seite 8, im Abschnitt 4.1.4, wird als Einschlusskriterium eine Studiendauer von mindestens 24 Wochen genannt. Hier halten wir eine Studiendauer von mindestens 12 Wochen für sinnvoll, um auch kurze Studien mit dem Nachweis eines schnell erreichten Patientennutzens zu erfassen.

Zwar entspricht die Mindestdauer von 24 Wochen sowohl den Forderungen der US-amerikanischen Zulassungsbehörde FDA, als auch der europäischen Zulassungsbehörde EMA, jedoch sollten folgende Fakten Anlass für den Einschluss von Studien mit einer Mindestdauer von 12 Wochen sein:

- Eine retrospektive Analyse mehrerer RCTs von *M. Boers* (2010)⁸ zeigte, dass das symptombezogene Ansprechen der Therapie bereits nach drei Monaten beurteilt werden kann.
- Die Deutsche Gesellschaft für Rheumatologie (DGRh) empfiehlt, die Entscheidung über die Fortführung einer Therapie auf der Basis des Ansprechens nach 12 Wochen zu treffen⁹.
- Eine frühe Entscheidung über die Fortführung der Therapie vermeidet im Falle des nicht ausreichenden Ansprechens eine unnötige Belastung des Patienten. Ferner ist hier eine Relevanz im Hinblick auf die Wirtschaftlichkeit der Therapie aus GKV-Perspektive gegeben, welche Gegenstand einer nachfolgenden Kosten-Nutzen-Bewertung werden könnte.

Auf Seite 10 sind im Abschnitt 4.2.1 die Datenbanken für die durchzuführende Literaturrecherche aufgelistet. Das „*PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions*“ empfiehlt die Nennung der einzelnen Schlagworte der Recherche¹⁰. Im Sinne der Verfahrenstransparenz, halten wir die Einfügung einer solchen Liste in den Berichtsplan für notwendig.

Auf Seite 12 des vorläufigen Berichtsplans wird die Bewertung patientenrelevanter Endpunkte, die mit Hilfe von komplexen Skalen operationalisiert werden, beschrieben. Hier soll eine Relevanzbewertung auf Basis von Mittelwertdifferenzen und Responderanalysen vorgenommen werden. Es sollen ggf. auch standardisierte Mittelwertdifferenzen (SMD in Form von Hedge's g) mit einer Irrelevanzschwelle von 0,2 SMD zur Anwendung kommen.

Es ist aus unserer Sicht leider nicht nachvollziehbar, warum der Wert von 0,2 für kleine Werte von Hedge's g gewählt wurde. Dieser Wert sollte nicht ohne detaillierte Begründung willkürlich auf die Erhebungen von patientenrelevanten Endpunkten angewendet werden. Sinnvoll wäre vielmehr eine Konsensbildung mit allen Beteiligten.

Auf Seite 14 werden schließlich im Abschnitt 4.4.4 die zu berücksichtigenden Subgruppenmerkmale und Effektmodifikatoren beschrieben. Wir möchten hier anregen, noch die Faktoren Krankheitsdauer, Schwere der Erkrankung und Begleittherapie mit aufzunehmen.

Monheim, 26. Oktober 2010



-
- ¹ D.T. Felson et al., *Arthritis and Rheumatism*, Vol. 36, No. 6 (June 1993), 729-740
- ² American College of Rheumatology Committee to reevaluate improvement Criteria, *Arthritis & Rheumatism (Arthritis Care & Research)*, Vol. 57, No. 2, March 15, 2007, pp 193-202 Heruntergeladen von http://www.rheumatology.org/practice/clinical/classification/response_08.pdf, Zugriff am 25.10.2010, 11.40h.
- ³ D. van der Heijde et al., *Ann. Rheum. Dis.* 2008; 67: 1267-1270
- ⁴ F.C. Breedveld, et al., *Ann. Rheum. Dis.* 2005, 64: 52-55
- ⁵ A. Kavanaugh et al., *The Journal of Rheumatology* 2004; 31: 5, 849-855
- ⁶ T. Pincus, *British Journal of Rheumatology* 1995; 34: (suppl. 2) 59-73
- ⁷ R. Fleischmann, *Expert Opin. Biol. Ther.* (2010) 10(5):773-786
- ⁸ M. Boers, *Ann Rheum Dis* 2010; 69: 186-192
- ⁹ M. Schneider et al., *DGRh-Leitlinie: Interdisziplinäre Leitlinie Management der frühen rheumatoiden Arthritis*, 2. überarbeitete Auflage, 2007.
- ¹⁰ A. Liberati et al., *Journal of Clinical Epidemiology* 62 (2009) e1ee34

A 1.9 Verband Forschender Arzneimittelhersteller e. V.

Autoren:

Throm, Siegfried, Dr.

Stellungnahme

zum vorläufigen Berichtsplan

**Biotechnologisch hergestellte Arzneimittel
in der Zweitlinientherapie bei der
rheumatoiden Arthritis**

(Auftrag A 10-01, Version 1.0, Stand
20.09.2010)



Der vfa, dem alle Hersteller der von diesem Nutzungsbewertungsverfahren betroffenen Arzneimittel angehören, bedankt sich für die Möglichkeit der Stellungnahme und hat folgende produktübergreifende Anmerkungen zu diesem Entwurf:

Seite 2/4

Seite 10, Punkt 4.2.1 – Bibliographische Literaturrecherche

In diesem Kapitel sind die Quellen aufgeführt, in denen die systematische Literaturrecherche betrieben werden soll.

Das PRISMA-Statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions von 2009 empfiehlt in Item 8: Search: Present the full electronic search strategy for at least one major database, including any limits used, such that it could be repeated. Dazu gehören auch die Suchbegriffe (search terms).

Daher regen wir an, in den Berichtsentwurf zusätzlich zu den Datenbanken auch die Schlagwörter für diese Literaturrecherche anzugeben.

Seite 12, Punkt 4.4 – Informationssynthese und -analyse, Absätze 2 – 4

Gemäß diesen Absätzen soll für die Bewertung patientenrelevanter Endpunkte, die mit Hilfe von komplexen Skalen operationalisiert werden, eine Relevanzbewertung auf Basis von Mittelwertdifferenzen und Responderanalysen vorgenommen werden. Hierbei sollen ggf. auch standardisierte Mittelwertdifferenzen (SMD in Form von Hedges'g) mit einer Irrelevanzschwelle von 0,2 SMD zur Anwendung kommen.

Hierzu ist anzumerken, dass dieses Konzept aus dem Bereich der Sozial- und Verhaltenswissenschaften stammt und der Wert von 0,2 für kleine Werte von Cohen willkürlich gewählt wurde. Daher kann dieser Wert nicht ohne weiteres auf die entsprechenden Erhebungen von patientenrelevanten Endpunkten übertragen werden.

Seite 14, Punkt 4.4.4 – Subgruppenmerkmale und andere Effektmodifikatoren

Bisher ist vorgesehen, als mögliche Effektmodifikatoren Geschlecht, Alter und Art der Vorbehandlung in die Analysen einzubeziehen.

Hier sollten unbedingt noch folgende klinische Faktoren aufgenommen werden, die ebenfalls die Effekte beeinflussen:

- Schweregrad der rheumatoiden Arthritis
- Krankheitsdauer

Seite 3/4

Seite 6, Punkt 3.1 – Zeitlicher Verlauf des Projekts

Kommentierungsfrist für den Vorbericht

Weiterhin möchten wir bereits hier darauf hinweisen, dass der Leitfaden des IQWiG zu Stellungnahmen zu Vorberichten und das Methodenpapier des IQWiG „Allgemeine Methoden (Version 3.0)“ auf Seite 19 eine Mindestfrist von 4 Wochen vorsieht. Da für die Kommentierung des Vorberichts in diesem Verfahren 4 Wochen nicht ausreichen, halten wir eine Frist von 8 Wochen für angemessen.

Dies ist zum einen darin begründet, dass angesichts der Vielzahl von betroffenen Wirkstoffen, die zum Teil schon lange im Verkehr sind, ein recht umfangreicher Vorbericht zu erwarten sein dürfte, dessen Prüfung und Kommentierung erhebliche Zeit in Anspruch nimmt.

Dies gilt nicht nur für die Firmen, sondern auch für die weiteren in die Kommentierung einbezogenen Kreise.

25. Oktober 2010

Literaturverzeichnis

Seite 4/4

- The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration Alessandro Liberati et. al.; Journal of Clinical Epidemiology 62 (2009) e1-e34 (<http://download.journals.elsevierhealth.com/pdfs/journals/0895-4356/PIIS0895435609001802.pdf>)