

**Arzneimittelkommission
der deutschen Ärzteschaft**
Fachausschuss der Bundesärztekammer



**Stellungnahme der Arzneimittelkommission der
deutschen Ärzteschaft (AkdÄ) zur**

**„Aktualisierung einiger Abschnitte der
allgemeinen Methoden Version 4.0 sowie neuer
Abschnitte zur Erstellung der Allgemeinen
Methoden Version 4.1“
des Instituts für Qualität und Wirtschaftlichkeit
im Gesundheitswesen (IQWiG)
Entwurf vom 18.04.2013**

Berlin, den 22.05.2013

www.akdae.de

Ad 3.1.4 Endpunktbezogene Bewertung

In dem neugefassten Abschnitt 3.1.4 „Endpunktbezogene Bewertung“ ist das Vorgehen bei der Ableitung der Belege für die Situationen präzisiert und erweitert worden, in denen wegen Heterogenität der Ergebnisse der Einzelstudien kein gemeinsamer Effektschätzer berechnet werden kann. In diesen Situationen kann es in speziellen Fällen sinnvoll und gerechtfertigt sein, Aussagen zur Beleglage von Effekten zu machen, auch wenn diese nicht quantifizierbar sind – als (extremes) Beispiel: alle Studien des Pools zeigen signifikante Effekte der gleichen Richtung, die Effektschätzer variieren aber so stark, dass Heterogenität vorliegt und keine quantitative Metaanalyse möglich ist.

Nach dem Methodenpapier 4.0 können bisher in diesen Situationen gegebenenfalls auch Belege abgeleitet werden, wenn bestimmte Kriterien erfüllt und so genannte „gleichgerichtete Effekte“ nachweisbar sind. Dem Nachweis „gleichgerichteter Effekte“ wird bei der Bestimmung der Beleglage dann gleiche Bedeutung beigemessen wie einem signifikanten Ergebnis einer Metaanalyse bei homogenem Studienpool. Die Überarbeitung des Methodenpapiers sieht vor, die Kriterien für „gleichgerichtete Effekte“ zu präzisieren und zudem die Unterkategorien „deutlich gleichgerichtet“ und „mäßig gleichgerichtet“ einzuführen.

Zum Nachweis „gleichgerichteter Effekte“ soll jetzt zunächst das so genannte Prädiktionsintervall betrachtet werden. Das Modell des Prädiktionsintervalls wurde entwickelt, um Ergebnisse von „Random-Effects“-Metaanalysen für praktische Umsetzungen und Entscheidungen besser interpretierbar zu gestalten, vor allem bei Vorliegen von Heterogenität (1;2). Es kann vereinfacht als der Größenbereich bezeichnet werden, in dem ein Effekt (mit in der Regel 95-prozentiger Wahrscheinlichkeit) erwartet werden kann, wenn die Ergebnisse einer „Random-Effects“-Metaanalyse auf eine Einzelsituation übertragen werden. Ein Prädiktionsintervall kann auch bei heterogener Studienlage berechnet werden und dann bei der Interpretation der Datenlage hilfreich sein.

Wenn wegen Heterogenität ein Summenschätzer nicht berechnet werden kann, das Prädiktionsintervall den Nulleffekt aber nicht überdeckt, soll nach dem Überarbeitungsentwurf des Methodenpapiers von „deutlich gleichgerichteten Effekten“ ausgegangen werden, die wie signifikante Ergebnisse für den Summenschätzer bei homogener Studienlage behandelt werden. Überdeckt das

Prädiktionsintervall dagegen den Nulleffekt, sollen nur „mäßig gleichgerichtete Effekte“ konstatiert werden, und auch nur dann, wenn weitere Kriterien erfüllt sind. Diese weiteren Kriterien sind weitgehend identisch mit den Kriterien, die bisher für die Feststellung „gleichgerichteter Effekte“ ausreichen (siehe IQWiG Methodenpapier 4.0, Seite 38).

Im Endergebnis werden mit Einführung des Prädiktionsintervalls als Methode höhere Anforderungen gestellt an (dann jedoch auch „deutlich ...“) „gleichgerichtete Effekte“, da Prädiktionsintervalle beispielsweise in der Regel breiter sind als Konfidenzintervalle (2). Die Verwendung von Prädiktionsintervallen kann dazu führen, dass Ergebnisse von Metaanalysen konservativer interpretiert werden (in Richtung eines Nulleffektes), wie auch kürzlich im Rahmen einer empirischen Überprüfung gezeigt worden ist (3). Sie erlauben zudem ein standardisiertes, quantifizierendes und damit objektiveres Vorgehen bei Heterogenität und erscheinen deshalb eine Verbesserung. „Deutlich gleichgerichtete Effekte“ führen gemäß der geplanten Überarbeitung, wie in der Tabelle 2 erkennbar, als Konsequenz zu gleichen Schlussfolgerungen für die Beleglage wie signifikante Ergebnisse für den Summenschätzer bei homogener Studienlage. Nur „mäßig gleichgerichtete Effekte“ würden im Gegensatz dazu – und im Gegensatz zum Verfahren gemäß Methodenpapier 4.0 bei „gleichgerichteten Effekten“ – im Endergebnis zu einer Herabstufung der Beleglage führen: von Beleg zu Hinweis bzw. von Hinweis zu Anhaltspunkt. Auch hier bedeuten die vorgesehenen Änderungen strengere Anforderungen an eine höhere Beleglage.

Da Prädiktionsintervalle erst ab mindestens vier Studien valide berechnet werden können (1), werden bei einem Pool von zwei oder drei Studien gesonderte Kriterien für die Einteilung als „mäßig gleichgerichtete Effekte“ oder „deutlich gleichgerichtete Effekte“ aufgestellt (Seite 16, Punkte 1. und 2.). Diese Kriterien erscheinen plausibel und nachvollziehbar; es ist jedoch unklar, durch welche Überlegungen und empirischen Erfahrungen sie genau begründet sind.

In diesem Punkt unterstützt die AkdÄ den Vorschlag des IQWiG, schlägt jedoch vor, die Kriterien für die Einteilung bei Vorliegen eines Pools von zwei und drei Studien differenzierter darzustellen und mit empirischen Erfahrungen zu unterstützen.

Die gegenüber der entsprechenden Tabelle im Methodenpapier 4.0 jetzt erweiterte Tabelle 2 für die Anforderungen an die Beleglage ergibt sich allein durch die

differenziertere Vorgehensweise für Studien mit „gleichgerichteten Effekten“ bei heterogener Studienlage. Das Verfahren zur Ermittlung der Beleglage ist damit im Überarbeitungsentwurf noch differenzierter und komplexer geworden; inhaltlich erscheint er nachvollziehbar und begründet. Die weitere Differenzierung gerade für Situationen mit heterogenen Studien, die aber gleichgerichtete Effekte aufweisen, ist vor dem Hintergrund der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV) notwendig geworden, die als eine der Kategorien einen belegten, aber „nicht quantifizierbaren Zusatznutzen“ vorsieht.

Die AkdÄ stimmt der weiteren Differenzierung zu.

Das Schema der Anforderungen an die Beleglage (Tabelle 2) zielt weiterhin nicht auf die Situationen ab, in denen auf Basis der Ergebnisse einer einzigen Studie Nutzenbelege abgeleitet werden können. Wie im Methodenpapier 4.0 findet sich hierzu weiterhin nur als einziger knapper Satz „Soll aus lediglich einer Studie im Ausnahmefall ein Nutzenbeleg abgeleitet werden, so sind an eine solche Studie und ihre Ergebnisse besondere Anforderungen zu stellen [157].“ Die Literaturstelle [157] verweist auf ein Dokument der EMA aus 2011: „Points to consider on application with: 1. meta-analyses; 2. one pivotal study“. Dort werden unter „III.2 Prerequisites for one pivotal study application“ die – explizit besonderen – Anforderungen genannt, die bei Einreichung lediglich einer Studie für eine Zulassung erfüllt sein müssen. Unter den Forderungen sind erwähnenswert: externe Validität und Übertragbarkeit der Ergebnisse; klinische Relevanz der Effektgröße; Signifikanzniveau deutlich unter 5 % mit engen Konfidenzintervallen; innere Konsistenz über Subgruppen; keine Zentreffekte.

Die Situationen, dass eine einzelne Studie für die Nutzenbewertung eines Arzneimittels zur Verfügung steht, sind, wie die bisherigen Verfahren gezeigt haben, nicht selten (z. B. Ticagrelor, Apixaban bei VHF, demnächst Rivaroxaban und Dabigatran bei VHF, viele onkologische Wirkstoffe). Es ist nicht nachvollziehbar, warum entsprechende Anforderungen bisher nicht weiter konkretisiert und operationalisiert worden sind und dies bei der Überarbeitung ebenfalls nicht vorgesehen ist. Zudem beziehen sich die Anforderungen der EMA in dem Dokument auf die Zulassung eines Wirkstoffs und es ist unklar, ob und wie sie für die Nutzenbewertung anwendbar sind.

Für Situationen, in denen für die Nutzenbewertung nur eine einzige Studie vorliegt, schlägt die AkdÄ vor, eine von den Kriterien der Zulassung unabhängige Operationalisierung zu erarbeiten.

Ad 3.3.3 Nutzenbewertung von Arzneimitteln gemäß § 35a SGB V

(einschließlich Anhang: Rationale der Methodik zur Feststellung des Ausmaßes des Zusatznutzens)

Im Zentrum dieses Abschnittes einschließlich des entsprechenden Anhangs steht die Operationalisierung zur Feststellung des Ausmaßes des Zusatznutzens. Ziel dieser Operationalisierung ist die Umsetzung von Rechtsvorgaben, die in § 5 Abs. 7 der AM-NutzenV formuliert sind. Diese Verordnung gibt sowohl für das Ausmaß (erheblich; beträchtlich; gering; nicht quantifizierbar; kein; geringerer) als auch für Qualität und Art des Zusatznutzens (z. B. Überlebensdauer; schwerwiegende Symptome; Lebensqualität) Kategorien vor, auch wenn diese nicht abschließend formuliert oder konkretisiert sind. Die Aktualisierung des Methodenpapiers ist auf der einen Seite bei der Operationalisierung des Ausmaßes des Zusatznutzens an diese Vorgaben gebunden, muss sie andererseits für die Bewertung im konkreten Fall umsetzbar machen.

Weiter ist von Bedeutung, dass der Gesetzgeber für die Preisfindung zur Erstattung von (zumindest neu zugelassenen) Arzneimitteln im GKV-Bereich primär nicht Kosten-Nutzen- oder Kosten-Nutzwert-Analysen vorsieht, sondern eine vorgeschaltete Nutzenanalyse mit nachfolgender Festsetzung bzw. Verhandlung des Erstattungsbetrags. Dieses Verfahren ist zunächst zu begrüßen; für seine Umsetzung existieren aber international vergleichsweise nur wenige verwertbare Erfahrungen.

Die im Überarbeitungsentwurf dargestellte Operationalisierung zur Feststellung des Ausmaßes des Zusatznutzens ist *de facto* nicht neu; die Kernelemente sind bereits in der Dossierbewertung von Ticagrelor durch das IQWiG beschrieben (Anhang A Dossierbewertung A11-02 vom 29.09.2011). Dieses Verfahren zur Operationalisierung wendet das IQWiG seitdem bei Nutzenbewertungen von Arzneimitteln nach § 35a SGB V an.

Von den in der AM-NutzenV vorgegebenen Kategorien für das Ausmaß des Zusatznutzens sind für die Operationalisierung nur die Kategorien „erheblich“, „beträchtlich“ und „gering“ von Bedeutung. Nur patientenrelevante Nutzen- und Schadenereignisse (wenn man von den seltenen Ausnahmen solcher Surrogate absieht, die ausreichend validiert sind und gleichzeitig eine Quantifizierung patientenrelevanter Ereignisse zulassen) werden betrachtet und hierarchisiert. Die Ausgestaltung der Kategorien und Einordnung einzelner Nutzen- oder Schadenereignisse in die Kategorien ist im Detail arbiträr, grundsätzlich aber nachvollziehbar. Weil die Einteilung im Detail arbiträr ist, erscheint wichtig, dass sie transparent erfolgt, um gegebenenfalls andere Einteilungen und Bewertungen von Nutzen- oder Schadenereignissen zu ermöglichen. Letzteres erscheint in ausreichendem Maße gegeben.

Weiterhin werden für jede der Ereigniskategorien spezifische Grenzen zum Ausmaß von Effekten festgelegt, die eine Kategorisierung als „erheblicher“, „beträchtlicher“ oder „geringer“ Zusatznutzen erlauben. Die Kategorisierung des Ausmaßes des Zusatznutzens werden exemplarisch für relative Risiken und Hazard Ratios entwickelt (und versucht, andere Effektmaße auf diese zurückzuführen), wobei für die Patienten positive Veränderungen von Ereignisraten als relative Risiken oder Hazard Ratios unter 1 dargestellt werden. Die relativen Risiken (genauer die oberen Grenzen des 95 % Konfidenzintervalls), die ein Arzneimittel für einen beispielsweise beträchtlichen Zusatznutzen erreichen muss, sind bei Ereignissen, die für die Patienten gravierend sind, näher an der 1 gelegen (Nulleffekt) als bei nicht schwerwiegenden Ereignissen. Auch die Grenzsetzung in den einzelnen Ereigniskategorien für einen „erheblichen“, „beträchtlichen“ oder „geringen“ Zusatznutzen ist arbiträr; andere Grenzen könnten diskutiert werden. Soweit bekannt, liegen empirische Erfahrungen mit dieser Vorgehensweise auch aus anderen Ländern kaum vor; dies erscheint plausibel und begründet.

Es wird in dem Methodenpapier weiterhin von der Grundforderung ausgegangen, dass hinsichtlich der Kategorie Mortalität eine Halbierung des Risikos gegenüber einer Vergleichstherapie erreicht werden muss, um als erheblicher Zusatznutzen bezeichnet werden zu können (4). Dabei handelt es sich aber um eine arbiträre Grenzsetzung, die nicht weiter auf Konsistenz und Anwendbarkeit geprüft ist. Aus dieser Grundannahme entwickelt das IQWiG mittels verschiedener Rasterungen für die einzelnen Kategorisierungen Grenzen (Schwellenwerte) für die übrigen

Kategorien zum Ausmaß des Zusatznutzens bei der Zielgröße Überleben und auch die Grenzen für das Ausmaß „erheblicher“, „beträchtlicher“ oder „geringer“ Zusatznutzen für andere Kategorien der Zielgrößen. In Simulationsverfahren wurde dann gezeigt, dass die in dem Überarbeitungsentwurf verwendeten Schwellenwerte zu Kategorien für das Ausmaß des Zusatznutzens führen, die (in praktisch relevanten Bereichen) unabhängig vom Basisrisiko der untersuchten Populationen sind.

Die verwendeten Grenzziehungen selbst – von den dann zur Ausgestaltung verwendeten Methoden abgesehen – sind kein methodisches Problem, sondern die Wertungen, die in dem Geltungsraum von Entscheidungen, die daraus abgeleitet werden, konsentiert sein müssen. Wenn sie auch grundsätzlich plausibel erscheinen, sind jetzt allerdings schon dadurch Tatsachen geschaffen worden, dass sie – zumindest vom IQWiG für seine Dossierbewertungen – bei der frühen Nutzenbewertung Anwendung gefunden haben. Für zukünftige Bewertungen werden diese Grenzziehungen aufgrund des Gebots der Gleichbehandlung schwer rückgängig zu machen sein.

Für die Formulierung des Ausmaßes des Zusatznutzens wird auf das relative Risiko als Effektmaß zurückgegriffen und nicht auf absolute Risikodifferenzen. Die Vorteile des relativen Risikos gegenüber Risikodifferenzen werden im Entwurf der Überarbeitung ausführlich beschrieben (beispielsweise Stabilität des relativen Risikos gegenüber dem Basisrisiko; Abhängigkeit der Risikodifferenzen von betrachteten Zeitintervallen). Wenn es gilt, Methoden (hier speziell Wirkstoffe) zu bewerten, ist dies unstrittig. Allerdings sollte auch bewertet oder zumindest ausreichend transparent dargestellt werden, welche absoluten Ausmaße an Zusatznutzen sich hinter den einzelnen Kategorien „erheblicher“, „beträchtlicher“ oder „geringer“ Zusatznutzen für die deutsche Versorgungssituation verbergen. Als Beispiel würde eine 50-prozentige Reduktion der Mortalität durch ein Arzneimittel (erheblicher Zusatznutzen) genauso eine Reduktion von 10 % auf 5 % wie von 1 % auf 0,5 % bedeuten können. Die entsprechenden Risikodifferenzen sind aber – trotz gleicher Kategorisierung des Zusatznutzens nach der Methode des IQWiG – unter Versorgungsaspekten unterschiedlich zu bewerten. Der Vorteil der relativen Risiken wie beispielsweise Übertragbarkeit auf ganze oder Subpopulationen mit anderem Basisrisiko soll nicht negiert werden.

Die AkdÄ regt an, zu überlegen, ob bei den endpunktbezogenen Ergebnissen abschließend neben den Ausmaßkategorien („erheblicher“, „beträchtlicher“ oder

„geringer“ Zusatznutzen) auch die relativen Risiken (mit 95 % Konfidenzintervall) und die absoluten Risikodifferenzen (mit 95 % Konfidenzintervall) für die untersuchte Population und Zeitdauer angegeben werden sollten. Dies würde die Transparenz maßgeblich erhöhen, ohne eine Übertragung auf Sub- oder andere Populationen zu verhindern.

Des Weiteren schlägt die AkdÄ vor, zu erwägen, ob bei der Ergebnisdarstellung für den Zusatznutzen bei Arzneimitteln auf eine abschließende zusammenfassende Bewertung nicht verzichtet werden kann und nur das Ausmaß des Zusatznutzens für die einzelnen Endpunkte präsentiert werden sollte. Die abschließende Bewertung unter Abwägung der Nutzen- und Schadensaspekte ist (zumindest bisher) kein methodisches Verfahren, sondern eine Werteabwägung, für die ein Konsensprozess unter den Betroffenen notwendig ist.

Die AkdÄ regt an, eventuell auf die abschließende zusammenfassende Bewertung zu verzichten, zumindest aber die endpunktbezogenen Ergebnisse, wie oben beschrieben, transparenter und mit den absoluten Risikodifferenzen (einschließlich der bezüglichen Rahmenbedingungen) darzustellen.

Ad 7.3.8 Meta-Analysen

Der Überarbeitungsentwurf unterscheidet sich vom entsprechenden Abschnitt in dem Methodenpapier 4.0 lediglich dadurch, dass 2 Absätze ergänzt werden sollen, die sich mit der geplanten Verwendung der Prädiktionsintervalle befassen. Demnach ist vorgesehen, Prädiktionsintervalle grafisch künftig als zusätzliche waagerechte Rechtecke in Forest-Plots aufzunehmen, um die Heterogenität der Studien zu veranschaulichen. Die Prädiktionsintervalle sollen dabei nicht zur Beurteilung der Signifikanz von Effekten herangezogen werden.

Die AkdÄ stimmt dieser Ergänzung zu.

Literatur

1. Higgins JP, Thompson SG, Spiegelhalter DJ: A re-evaluation of random-effects meta-analysis. J R Stat Soc Ser A Stat Soc 2009; 172: 137-159.

2. Riley RD, Higgins JP, Deeks JJ: Interpretation of random effects meta-analyses. *BMJ* 2011; 342: d549.
3. Graham PL, Moran JL: Robust meta-analytic conclusions mandate the provision of prediction intervals in meta-analysis summaries. *J Clin Epidemiol* 2012; 65: 503-510.
4. Djulbegovic B, Kumar A, Soares HP et al.: Treatment success in cancer: new cancer treatment successes identified in phase 3 randomized controlled trials conducted by the National Cancer Institute-sponsored cooperative oncology groups, 1955 to 2006. *Arch Intern Med* 2008; 168: 632-642.