

Bundesverband
Medizintechnologie e.V.
Reinhardtstraße 29b
10117 Berlin
Tel. +49 (0)30 246 255 - 0
Fax +49 (0)30 246 255 - 99
info@bvmed.de
www.bvmed.de

Berlin, 31. Januar 2020
Win
☎ +49 (0)30 246 255-26

Stellungnahme

zum

Entwurf des allgemeinen Methodenpapierses Version 6.0 des IQWiG

Die vorliegende Stellungnahme gibt die Auffassung der Mitglieder des Fachbereiches Market Access der AG Methodenbewertung des Bundesverbands Medizintechnologie wieder.

I. Grundsätzliche Anmerkung

Die Methoden des IQWiG zur Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten muss sich an die Dynamik und Besonderheiten der Technologien kontinuierlich anpassen. Die derzeitige Ausgestaltung des Methodenpapiers ist im vorliegenden Entwurf nicht vollumfänglich ausreichend, um die Besonderheiten der Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten in der Arbeitsweise des IQWiG zu berücksichtigen. Wir geben mit dieser Stellungnahme Hinweise zur sachgerechten Anpassung der methodischen Werkzeuge mit dem Angebot des weiteren Dialoges.

Bei kritischen Fragestellungen, die derzeit auf der alleinigen Expertise des IQWiGs beruhen (wie zum Beispiel die Einschätzung der qualitativen Ergebnissicherheit) empfehlen wir die Einbeziehung von Fachgesellschaften. So kann eine objektivere Entscheidung beruhend auf der bestmöglichen Fachexpertise getroffen werden.

BVMed bietet weiterhin Expertise für das Kuratorium an

Die Besonderheiten der Bewertung neuer Untersuchungs- und Behandlungsmethoden mit Medizinprodukten, insbesondere hoher Risikoklassen, sollen, im Gegensatz zu den Methoden der Nutzenbewertung von Arzneimitteln, ausreichend in der Arbeitsweise des IQWiG berücksichtigt werden. Wir halten es weiterhin für angezeigt, dass der BVMed zukünftig für die Medizinprodukteindustrie aktiv beratend im Kuratorium der Stiftung vertreten ist, um hier die Expertise der Medizinprodukteindustrie zur Verfügung zu stellen.

II. Anmerkungen zu den einzelnen Kapiteln

1.2 Evidenzbasierte Medizin

Zu 1.2.5 Die Verbindung zwischen Ergebnissicherheit und Alltagsnähe

Änderungsvorschlag:

Großen Wert auf die Abschätzung der Ergebnissicherheit zu legen, wird oft kritisiert. Ein Argument ist, dass Studien mit hoher Ergebnissicherheit, insbesondere randomisierte kontrollierte Studien, zwar eine hohe interne Validität besäßen, jedoch oft die Versorgung von Patientinnen und Patienten im Alltag nicht abbildeten und deshalb nicht „übertragbar“ seien, also nur eine geringe externe Validität vorhanden sei. Dabei ist zu prüfen, wie gut die in Studien untersuchte Patientenpopulation, die eingesetzten Interventionen und die Zielkriterien dem Versorgungsalltag entsprechen. Diese Kritik ist dann oft mit der Forderung verbunden, andere Studientypen ohne Randomisierung hinzuzunehmen, um Alltagsbedingungen besser einfließen zu lassen.

Diese Kritik vermischt jedoch Argumentationsebenen, die klar voneinander zu trennen sind. Folgende Aspekte müssen Beachtung finden:

- 1) Grundlage einer Nutzenbewertung ist der Nachweis von Kausalität. Unverzichtbare Bedingung für den Nachweis von Kausalität ist ein vergleichendes Experiment, das so angelegt sein muss, dass ein Unterschied zwischen Interventionsgruppen – ein Effekt – nur auf eine einzige Einflussgröße – die geprüfte Intervention – zurückgeführt werden kann. Dieses Ziel macht für klinische Studien erhebliche Anstrengungen nötig, weil es zahlreiche unerwünschte Einflüsse gibt, die einen Effekt vortäuschen oder auch verbergen können (Verzerrung). Der stärkste dieser verzerrenden Einflüsse sind ungleiche Ausgangsbedingungen zwischen den Vergleichsgruppen. Die Randomisierung zusammen mit einer sorgfältigen Verdeckung der Zuteilung (Allocation Concealment) ist das beste gegenwärtig verfügbare Instrument, um diese Verzerrung zu minimieren. Die zufällige Zuteilung der Teilnehmer zu den Gruppen stellt sicher, dass es keine systematischen

Unterschiede zwischen den Gruppen gibt, und zwar weder bezüglich bekannter Eigenschaften (z. B. Alter, Geschlecht, Schwere der Erkrankung), noch bezüglich unbekannter Faktoren. Damit ist in randomisierten kontrollierten Studien (Randomized controlled Trials, RCTs) eine grundlegende Voraussetzung für einen Kausalitätsnachweis gegeben. Allerdings ist auch eine Randomisierung allein kein Garant für eine hohe Ergebnissicherheit. Dazu sind z. B. auch eine unverzerrte Erhebung, Zusammenfassung und Veröffentlichung der Ergebnisse nötig.

- 2) Andere Studientypen als RCTs sind in der Regel für einen Kausalitätsnachweis nicht geeignet. In nicht randomisierten vergleichenden Studien kann grundsätzlich nicht von einer Strukturgleichheit der Gruppen ausgegangen werden. Sie liefern daher immer ein potenziell verzerrtes Ergebnis und können zumeist die maßgebliche Frage nicht hinreichend sicher beantworten, ob ein Unterschied auf der geprüften Intervention beruht. Daher bedarf die Verwendung nicht randomisierter Studien, als Nachweis der Kausalität, einer Intervention einer besonderen Begründung bzw. besonderer Voraussetzungen und spezieller Qualitätsanforderungen.
- 3) Es trifft zu, dass viele randomisierte Studien Aspekte des Alltags der Patientenversorgung nicht abbilden, z. B. Patientinnen und Patienten mit Begleiterkrankungen ausschließen, die im Alltag häufig sind. Dies ist jedoch keine Konsequenz der Technik der Randomisierung, sondern die Folge anderer Faktoren wie z. B. der Festlegung enger Ein- und Ausschlusskriterien der Studie, der Auswahl der Interventionen oder der Zielkriterien. Zudem werden Patientinnen und Patienten in randomisierten Studien oft anders (intensiver und enger) betreut als im Alltag. Das alles sind aber bewusste Entscheidungen derjenigen, die in einer Studie eine bestimmte Frage beantworten wollen. Durch einen Verzicht auf die Randomisierung werden diese Entscheidungen nicht geändert. Auch in nicht randomisierten Studien findet eine Selektion der Teilnehmer durch Ein- und Ausschlusskriterien und ggf. andere Designcharakteristika statt, sodass externe Validität ebenso wenig per se gegeben ist wie bei RCTs.
- 4) Aber selbst wenn sich Patientengruppen einer RCT vom Versorgungsalltag unterscheiden, muss das die externe Validität des Ergebnisses nicht infrage stellen. Entscheidend ist vielmehr die Frage, ob zu erwarten ist, dass der in einer Population festgestellte Therapieeffekt in einer anderen Population anders ist.
- 5) Es hängt vom Einzelfall ab, wie sich die Intensität der Betreuung in einer Studie auswirken könnte. Zum Beispiel ist denkbar, dass ein Nutzen tatsächlich nur in den Händen besonders qualifizierter Ärztinnen und Ärzte vorhanden ist, weil unter Alltagsbedingungen ansonsten zu häufig Komplikationen auftreten. Alternativ bieten sich auch in diesen Fällen Studiendesigns deskriptiver Beobachtungsstudien in Form von prognostischen Korrelations- und Querschnittstudien aus hochstandardisierten Real-World-Evidenz Datenbanken an, die Unterschiede auf der Ebene der Arztqualifikation und Erfahrung, auf Grundlage der großen Datenmengen statistisch ermittelbar machen können. Es ist aber auch möglich, dass eine enge Betreuung Unterschiede zwischen Gruppen eher verringert. Zum Beispiel könnten Unterschiede in der Therapieadhärenz in Studien kleiner ausfallen, in denen Patientinnen und Patienten grundsätzlich sehr eng betreut werden.
- 6) Die Festlegung der Studienbedingungen liegt jedoch in der Hand der Initiatorinnen und Initiatoren einer klinischen Studie. Sie können die Fragestellung und Endpunkte definieren, die sie als so wichtig einschätzen, dass sie in der Studie beobachtet werden sollen. Wenn eine Herstellerin oder ein Hersteller eines Arzneimittels z. B. Therapieadhärenz für einen wichtigen Aspekt des Nutzens seines Produkts hält, wäre die nahe liegende Konsequenz, dann Studien aufzulegen, die diesen Aspekt mit möglichst großer Ergebnissicherheit und Nähe zu Alltagsbedingungen messen können und gleichzeitig die Patientenrelevanz demonstrieren.

Diese Ausführungen zeigen, dass Ergebnissicherheit und Alltagsnähe (oder interne und externe Validität) nicht fest miteinander verknüpft sind; Hohe Ergebnissicherheit und Alltagsnähe schließen einander nicht aus, sondern erfordern nur eine geeignete Kombination aus Studientyp, Design und Durchführung.

Auch wenn die Kritik mangelnder Alltagsnähe tatsächlich für viele Studien zutreffen mag, wäre nichts dadurch gewonnen, wenn man eine hohe Ergebnissicherheit für eine größere Alltagsnähe aufgeben würde. Man würde einen Mangel durch Inkaufnahme eines zweiten, noch schwerwiegenderen Mangels zu kompensieren versuchen.

Es sind Studien wünschenswert und machbar, die Alltagsnähe und hohe Ergebnissicherheit vereinigen. RCTs, in denen weder über den Behandlungsalltag hinausgehende Anforderungen an Patientinnen und Patienten gestellt noch definierte Studienvisiten vorgegeben werden, sind sehr wohl möglich. Sie werden international als Real-World Trials, Practical Trials oder Pragmatic Trials diskutiert. Allerdings können solche sogenannten pragmatischen Studien ihrerseits ebenfalls zu Interpretationsproblemen führen. Werden z. B. sehr breite Einschlusskriterien gewählt, stellt sich die Frage, ob die (Gesamt-)Studienergebnisse auf die gesamte Studienpopulation anwendbar sind, was letztlich über adäquate Subgruppenanalysen zumindest ansatzweise zu beantworten wäre. Der entscheidende Vorteil deskriptiver Beobachtungsstudien in Form von prognostischen Korrelations- und Querschnittstudien aus hoch standardisierten Real-World-Evidenz Datenbanken ist, dass klinische Effekte ermittelbar sind die sich in kohortenspezifischen Trends zeigen, auf Basis derer man Potentiale für einen Nutzen oder Schaden ableiten kann.

Begründung:

Mit dem Implantateregister - Errichtungsgesetz hat der Gesetzgeber eine Grundlage geschaffen, die neben dem Schutz der Gesundheit und Sicherheit von Patientinnen und Patienten, von Anwendern und von Dritten sowie der Abwehr von Risiken durch Implantate, explizit die Informationsgewinnung über die Qualität, und die Qualitätssicherung der Implantate und der medizinischen Versorgung mit Implantaten in den verantwortlichen Gesundheitseinrichtungen verfolgt. Dieser Zweck führt in der Berücksichtigung von Datenquellen und den daraus abzuleitenden Studiendesigns zu den deskriptiven Beobachtungsstudien in Form von prognostischen Korrelations- und Querschnittstudien. Insbesondere in der Bewertung von Nutzen und Schaden nicht-Medikamentöser Methoden und Interventionen unter der Anwendung fortschrittlicher Medizinprodukte darf die Berücksichtigung von Studienergebnissen aus diesem Kontext nicht ausbleiben und ist entsprechend in das Methodenpapier aufzunehmen.

3.1 Patientenrelevanter medizinischen Nutzen und Schaden

Zu 3.1.4 Endpunktbezogene Bewertung

B) Ableitung der Beleglage und Aussagesicherheit

Änderungsvorschlag:

Aus lediglich 1 Studie kann im Ausnahmefall für einen Endpunkt ein Beleg für eine spezifische (Teil-)Population abgeleitet werden, wenn ein Studienbericht gemäß den ICH-Richtlinien vorliegt, die auch sonst für einen Beleg geforderten Bedingungen erfüllt sind und darüber hinaus die Studie folgende besondere Anforderungen erfüllt:

- Die Studie ist multizentrisch, in jeden Studienarm wurden der wissenschaftlichen Fragestellung entsprechend Patientinnen oder Patienten eingeschlossen.
- Das Ergebnis ist innerhalb der Studie konsistent: Für die interessierende (Teil-)Population liegen Auswertungen verschiedener weiterer Teilpopulationen (insbesondere Teilmengen von Studienzentren) vor, die jeweils bewertbare und ausreichend homogene Effektschätzer ergeben. Diese Konsistenzbewertung ist bei binären Daten nur möglich, falls eine gewisse Mindestanzahl von Ereignissen vorliegt.

- Die im vorherigen Punkt adressierten Auswertungen für Teilpopulationen liegen für alle relevanten Endpunkte vor, d. h. diese Auswertungen sind nicht auf einzelne ausgewählte Endpunkte beschränkt.

Es ist möglich, dass sich die Beleglage beim Vorliegen von lediglich einer Studie, die für sich genommen nur einen Hinweis oder Anhaltspunkt ergibt, durch zusätzliche indirekte Vergleiche ändert. Hierfür sind allerdings hohe methodische Anforderungen an die indirekten Vergleiche zu stellen (siehe Abschnitt 9.3.9). Außerdem ist es bei entsprechend homogener Datenlage möglich, dass sich durch Ergänzung indirekter Vergleiche die Präzision der Effektschätzung erhöht, was bei der Feststellung des Ausmaßes des Zusatznutzens eine wichtige Rolle spielt (siehe Abschnitt 3.3.3).

Begründung:

Der Umfang der Stichprobengröße steht in direkter Verbindung zur wissenschaftlichen Fragestellung und den daraus abzuleitenden statistischen Testverfahren. Entsprechende Kalkulationsmethoden sind in der Humanwissenschaft validiert anzuwenden und transparent zu machen. (Röhrig B et al. Fallzahlplanung in klinischen Studien, *Deutsches Ärzteblatt Int* 2010; 107(31–32): 552–6 DOI: 10.3238/arztebl.2010.0552) Das Voraussetzen von Stichprobengrößen und Zentrenzahlen ist unwissenschaftlich und methodisch nicht haltbar. In der biostatistischen Wissenschaftsgemeinde ist die Debatte um das Signifikanzniveau eine derzeit im Schwerpunkt der „scientific method“ geführte Auseinandersetzung, die jedoch noch keinen einheitlichen Konsens gefunden hat (Ioannidis JPA. The Proposal to Lower P Value Thresholds to .005. *JAMA*. 2018; 319(14):1429–1430. doi:10.1001/jama.2018.1536). Mit der Aufnahme derart sensitiver Signifikanzniveaus setzt das Methodenpapier einen Standard, der in der wissenschaftstheoretischen Herangehensweise der Humanwissenschaften in diesem Maß noch keinen Konsens hat. Das IQWiG sollte in seinem Methodenpapier die etablierten Standards berücksichtigen. Die derzeitige Diskussion bezüglich des Sachverhaltes sollte durch das Methodenpapier nicht weiteren kontroversen Zuspruch bekommen, bis ein Sensitivitätsniveau für den p-Wert konsentiert ist. (0.05 or 0.005? P-value Wars Continue. The p-value is under fire yet again, but this time with some quick-and-dirty solutions (and some long-and-onerous ones too) to the problems created by relying on this quick-and-dirty test. <https://sciencebasedmedicine.org/0-05-or-0-005-p-value-wars-continue/>) Dem gegenübergestellt ist die Gewichtung des p-Wert alleine unvollständig. Der p-Wert bzw. die Nullhypothese muss in der Prüfgröße getestet werden, welches über Einseitige, oder im Fall eines ungerichteten Tests, über zweiseitige Signifikanztests durchführbar ist. Erst hierüber kann innerhalb der statistischen Ergebnisse entsprechende Validität abgeleitet werden, die eine haltbare Bewertung von Nutzen oder Schaden im Falle einer 1-Studiensituation zulässt.

3.2 Spezielle Aspekte der Nutzenbewertung

Zu 3.2.1 Auswirkung nicht publizierter Studienergebnisse auf das Fazit

Änderungsvorschlag:

Die folgende Tabelle 4 beschreibt, in welchen Konstellationen ein Risiko für eine Verzerrung von Bewertungsergebnissen besteht und welche Konsequenzen sich daraus für das Fazit der Nutzenbewertung ergeben. Die Szenarien 1 bis 3 beziehen sich auf Nutzenbewertungen mit Datenübermittlung durch Dritte, Szenarien 4 bis 6 auf solche ohne zunächst vorgesehene Datenübermittlung.

Tabelle 4: Szenarien der Datenvollständigkeit und Konsequenzen für das Fazit einer Nutzenbewertung

Szenario	Datenübermittlung durch Dritte (z. B. Daten von Herstellerinnen / Herstellern)	Nachweis, dass insgesamt Daten fehlen	Verzerrung durch Publikationsbias	Bewertung / Einfluss auf das Fazit
1	vollständig	nein oder in vernachlässigbarem Umfang	unwahrscheinlich	keine Einschränkung des Fazits
2	vollständig	Ja	möglich	Einschränkung der Aussagesicherheit
3	unvollständig	ja oder unklar	wahrscheinlich	kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen bzw. Schaden
4	keine Datenübermittlung vorgesehen	nein oder in vernachlässigbarem Umfang	unwahrscheinlich	keine Einschränkung des Fazits
5	keine Datenübermittlung vorgesehen	Ja	möglich	Einschränkung der Aussagesicherheit
6	keine Datenübermittlung vorgesehen	Ja oder unklar	wahrscheinlich	kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen bzw. Schaden

a: Bei Nutzenbewertungen mit eigenem Stellungnahmeverfahren kann das Fazit der Bewertung auch unter Vorbehalt gestellt werden.

b: Sofern im Ausnahmefall vom Datenübermittelnden nachgewiesen wird, dass die Menge der fehlenden Daten sicher irrelevant ist, wird das Fazit nicht eingeschränkt.

Ab Szenario 3: Sofern Unklarheit über die Vollständigkeit der Daten besteht, fordert das IQWiG den Hersteller oder die Herstellerin zur Datenübermittlung auf, kommt dieser oder diese der Aufforderung nicht nach, erfolgt die kategorisierte Einschätzung.

War in einer Nutzenbewertung mit Datenübermittlung durch Dritte diese Datenübermittlung vollständig und liegt kein Nachweis vor, dass Daten in einem mehr als vernachlässigbaren Umfang fehlen (Szenario 1), wird eine Verzerrung als unwahrscheinlich eingestuft. Die Schlussfolgerungen aus der Bewertung der Daten werden ohne Einschränkung in das Fazit übernommen. Wird bei einer vollständigen Datenübermittlung durch die Herstellerin oder den Hersteller durch andere Recherchen der Nachweis erbracht, dass Daten in einem mäßigen oder erheblichen Umfang (z. B. aus Investigator-initiierten Studien oder Studien anderer Hersteller) fehlen, auf die die Herstellerin oder der Hersteller keinen Zugriff hat, liegt zwar keine selektive Datenübermittlung durch die Herstellerin oder den Hersteller vor, ist aber eine Verzerrung durch die anderweitig fehlenden Daten möglich (Szenario 2). Eine Nutzensaussage im Fazit (im Sinne von Beleg) wird deshalb in ihrer Aussagesicherheit entsprechend abgeschwächt und als Hinweis oder Anhaltspunkt für Nutzen oder Schaden interpretiert.

Führte eine unvollständige Datenübermittlung durch Dritte zum Fehlen von Daten in einem erheblichen, mäßigen oder unklaren Umfang (Szenario 3), wird von einer selektiven Datenübermittlung ausgegangen. Weitere relevante unpublizierte Daten, die durch die Rechenschritte nicht zugänglich sind, können existieren. Eine weitere Auswertung der verfügbaren limitierten Daten und eine daraus abgeleitete Aussage zum Nutzen oder Schaden sind in dieser Situation wahrscheinlich hochgradig verzerrt und stellen damit keine valide Bewertungsgrundlage dar. In dieser Situation kann deshalb kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen oder Schaden der zu bewertenden Intervention festgestellt werden, unabhängig von den beobachteten Ergebnissen in den vorgelegten Daten.

Lässt sich in einer Nutzenbewertung ohne vorgesehene Datenübermittlung durch Dritte ein Fehlen von Daten nicht erkennen (Szenario 4), ergibt sich keine Einschränkung für das Fazit. Lassen sich jedoch fehlende Daten nachweisen (z. B. durch Abgleich zwischen registrierten und publizierten Studien), so ist eine Verzerrung durch Publikationsbias möglich (Szenario 5) oder sogar wahrscheinlich (Szenario 6). Die Abgrenzung zwischen beiden Szenarien erfolgt anhand der eingeschätzten Höhe des Anteils fehlender Daten. Bei möglichem Publikationsbias wird das Fazit in seiner Aussagesicherheit herabgestuft. Alternativ kann das Fazit unter Vorbehalt gestellt werden, wenn im Rahmen eines vom IQWiG durchgeführten Stellungnahmeverfahrens zu einer Nutzenbewertung die Möglichkeit besteht, dass fehlende Daten nachgeliefert werden. Bei wahrscheinlichem Publikationsbias wird nach erfolgloser Datenabfrage beim Hersteller oder der Herstellerin, kein Beleg (und kein Hinweis oder Anhaltspunkt) für einen Nutzen bzw. Schaden abgeleitet. In diesem Fall werden lediglich die vorhandenen und die fehlenden Informationen dargestellt.

Begründung:

Es ist in der Humanwissenschaft evident, dass nur etwa 36% der Abstract-Einreichungen mit vorläufigen Studienergebnisse zu einer Publikation führen. (Sprague S. et al. Barriers to full-text publication following presentation of abstract at annual orthopaedic meetings. *J Bone Joint Surg Am* 2003;85(1):158-163) Hinsichtlich einer behaupteten Ergebnisverzerrung der Outcomes von bisher nicht veröffentlichten Daten aus Studien, spiegelt die belastbare Datenlage solcher Data Bias, insbesondere bei den Outcomes, eine vollkommen andere Realität wider. In nur etwa 2% der Fälle von preminilary abstracts, die zur Publikation kamen, lagen Outcome Bias vor. (Guryel E. et al. From presentation to publication: the natural history of orthopaedics abstracts in the United Kingdom. *Postgrad Med J* 2006;82(963):70-72) Es liegen also hinreichende wissenschaftliche Erkenntnisse zur entsprechenden Würdigung solcher Daten vor. Zudem muss dem Hersteller oder der Herstellerin die Möglichkeit der Datenübermittlung gegeben werden, insofern der Bedarf dazu deutlich wird.

Zu 3.2.4 Patientenberichtete Endpunkte

Änderungsvorschlag:

Die im Fließtext dieses Abschnitts zitierte Referenz 251 ist zu ersetzen, denn die inhaltliche Ableitung aus der Referenz ist unzutreffend. Es liegt dementsprechend ein Zitationsfehler vor.

Begründung:

Die Referenz 251 verweist auf die Fundstelle der FDA "*Developing Medical Imaging Drug and Biological Products, Part 2: Clinical Indications (I)*". Die zutreffende Referenz für das Interpretieren von geeigneten Studien und deren Ergebnisse ist aber ausdrücklich die Fundstelle der FDA "*Developing Medical Imaging Drug and Biological Products, Part 3: Design, Analysis, and Interpretation of Clinical Studies (I)*". Bei allen Fundstellen dieser Art muss erwähnt werden, dass es keine Guidelines sondern Guidances sind. Bei allen diesen Guidances dieser Art vermerkt die veröffentlichende FDA zudem "Contains Nonbinding Recommendations". Die inhaltliche Gewichtung der wissenschaftstheoretischen Ableitung des Abschnittes ist dementsprechend zu übertragen.

Zu 3.4 Nicht medikamentöse therapeutische Interventionen

Vorbemerkung:

Die aufgezählten Felder der Anwendung der Methodik bedürfen einer gesonderten Betrachtung, wie beispielsweise das IDEAL –Statement für chirurgische Verfahren (Literaturhinweis, <https://doi.org/10.1016/j.zefq.2015.05.013>) es bereits aufgezeigt hat. Es muss gleiche Bedingungen für alle Interventionen geben, d. h. alle Felder der Medizin müssen einen Nutzen nachweisen, aber die Anforderung an evidenzbasierte Medizin kann sich nicht ausschließlich auf RCTs stützen, so dass für spezielle Felder (Chirurgie, etc.) auch entsprechend angepasste Evidenzgenerierung akzeptiert werden sollte.

Die pauschale Benennung der Nutznachweise für E-Health-Anwendungen ist nicht ausreichend, da bereits durch das Digitale-Versorgung-Gesetz (Drucksache 19/14867) für die dort genannten digitalen Gesundheitsleistungen für Medizinprodukte mit niedrigen Risikoklassen neben den Nutznachweisen auch positive Versorgungseffekte genannt werden, die in dem Entwurfspapier bisher nicht berücksichtigt werden. Zudem fehlen für die Bewertung von weiteren E-Health-Anwendungen derzeit der gesetzliche Auftrag bzw. die Anforderungen in der Verfahrensordnung durch den G-BA. Die besonderen Herausforderungen für die Konzeption und Durchführung von klinischen Studien bei Untersuchungs- und Behandlungsmethoden mit Medizinprodukten sind erstmals die Herausforderungen und Schwierigkeiten in der Entwurfsversion erwähnt worden, jedoch ist nicht erkennbar, inwiefern allein diese Benennung dazu beiträgt, eine sachgerechte Interpretation der genannten Schwierigkeiten zu gewährleisten. Es ist derzeit nicht erkenntlich, wie die genannten einzelnen Punkte (Lernkurve, Schrittinnovationen und Verblindungsmöglichkeiten/ subjektive Endpunkte) in welcher Form konkret berücksichtigt werden. Hier bedarf es weiterer Spezifikationen unter Berücksichtigung internationaler Erfahrungen und Anwendungen.

Zu 9.3.3 Beurteilung klinischer Relevanz

Punkte 1, 2 und 3: Spezifikation und Ausführungen zum Responsekriterium 15%

Änderungsvorschlag:

Wir halten eine starre Festlegung des Responsekriteriums auf 15 % zwar für pragmatisch aber nicht sinnvoll. Wir schlagen stattdessen die Verwendung einer angemessenen und begründeten, für die Therapie und den jeweiligen Outcome Parametern, definierten MID vor.

Begründung:

Wie das IQWiG selbst schreibt: "Eine Sichtung von systematischen Übersichtsarbeiten zu MIDs [z. B. [523,655]] hat ergeben, dass sich MIDs in vielen Fällen zwischen 10 % und 20 % der Spannweite der Scores eines Erhebungsinstruments bewegen."

Mit einer starren Festlegung des Responsekriteriums auf 15 % wird wissentlich in Kauf genommen, dass damit falsch zu hoch bzw. falsch zu niedrig kalkulierte Responderanalysen erfolgen, die die Evidenz für den Nutzen einer medizinischen Intervention einschränken können. In einem Worst Case Szenario wird damit eine wirkungsvolle Therapie aufgrund eines fehlerhaften Responsekriteriums als nicht besser als bestehende Therapien eingestuft und damit den Patienten ggf. eine wirksame Therapieoption vorenthalten.

Zu 9.3.12 Umgang unvollständigen Daten

Textabschnitt aus dem IQWiG Methodenentwurf 6.0:

„Ergebnisse fließen daher in der Regel nicht in die Nutzenbewertung ein, wenn diese auf weniger als 70% der in die Auswertung einzuschließenden Patientinnen und Patienten basieren, das heißt, wenn der Anteil der Patientinnen und Patienten, die nicht in der Auswertung berücksichtigt werden, größer als 30% ist.“

Änderungsvorschlag:

Wir schlagen die folgende Ergänzung vor: "Kann nachgewiesen werden, dass die jeweiligen Ergebnisse nicht empfindlich gegenüber fehlenden Daten sind, fließen die Ergebnisse in die Nutzenbewertung ein."

Begründung:

Wir halten eine generelle Festlegung, dass Ergebnisse in der Regel nicht in die Nutzenbewertung einfließen, wenn diese auf weniger als 70 % der in die Auswertung einzuschließenden Patientinnen und Patienten basieren, für nicht sinnvoll. Dies gilt insbesondere für Patienten mit terminalen Erkrankungen. Als Beispiel: eine Studie zur Lebensqualität bei Krebspatienten im terminalen Stadium kann trotz einer sehr hohen Todesrate, dennoch wertvolle Ergebnisse liefern. Eine Standard Kaplan-Meier Analyse zur Mortalität, zum Beispiel nach 24 Monaten, ist auch valide mit deutlich mehr als 30% der Patienten, zensiert nach 24 Monaten.

Abschließende Anmerkung:

In der Prüfung nicht-Medikamentöser Methoden und Interventionen unter der Anwendung fortschrittlicher Medizinprodukte, muss die Bewertung von Nutzen und Schaden den teils erschwerten Umständen einer klinischen Prüfung Rechnung getragen werden. Bei allen durchgeführten klinischen Prüfungen nicht-Medikamentöser Methoden und Interventionen unter der Anwendung fortschrittlicher Medizinprodukte im regulatorischen Umfang oder vor einer Nutzenbewertung im Zuge der Verfahren nach §§135, 137c, 137e und 137h SGB V, wurden bereits Ethikprüfungen nach humanwissenschaftlichen Standards durchgeführt. Gleiches gilt für die Durchführung von prospektiv durchgeführten Beobachtungsstudien mit Kontrollgruppen vor Verfahren der o.g. Rechtsgrundlagen, woraus in der Berücksichtigung dessen das Methodenpapier keine höhergestellten Anforderungen ex Post stellen kann. Das Methodenpapier kann lediglich Maßgabe für die Ableitung eines Nutzens oder Schadens sein, die sich alleine auf die Inhalte der Studien innerhalb der vorliegenden Designs stützt, und sollte keinen Standard als Voraussetzung für eine patientenrelevante Gewichtung der Endpunkte anlegen. Die Würdigung humanwissenschaftlicher Studienergebnisse bei nicht-Medikamentösen Methoden und Interventionen unter der Anwendung fortschrittlicher Medizinprodukte, muss daher an der verfügbaren Evidenz, und nicht an einer ex Ante behauptet vorliegenden Evidenz mit bestimmten Designfaktoren, beurteilt werden. Wissenschaftstheoretische Ansätze folgen oft dem Umstand des Möglichen und nicht der Maßgabe des Vermeintlichen, alles andere käme einer unangemessenen wissenschaftlichen Zäsur gleich und wäre als solche zu bewerten.

Mit freundlichen Grüßen

BVMed – Bundesverband
Medizintechnologie e. V.



Dr. Marc-Pierre Möll
Geschäftsführer



Olaf Winkler
Leiter Referat Industrieller Gesundheitsmarkt