

Änderung des IQWiG-Methodenpapiers: Patientensicht im AMNOG in Gefahr

Patientenberichtete Endpunkte haben einen hohen Stellenwert in der Nutzenbewertung. Sie liefern wertvolle Informationen über die Sicht von Patientinnen und Patienten auf die Erkrankung, die Nebenwirkungen der Therapie oder die Lebensqualität.

Das IQWiG hat nun zur strengeren Beurteilung solcher patientenbezogenen Auswertungen ein neues, generisches Richtmaß in einem „one-size-fits-all“-Ansatz festgesetzt. Die Messlatte für bedeutsame patientenberichtete Ergebnisse wird auf diesem Wege substanziell erhöht, wobei patientengerechte Besonderheiten nicht berücksichtigt werden. Der Entwicklungsansatz der Wissenschaft, die Bewertungsstandards durch sinnvolle Qualitätskriterien zu verbessern, wird außer Acht gelassen. In Folge stehen fast alle bisher akzeptierten Auswertungen patientenberichteter Endpunkte vor dem Aus. Der Nachweis von Verbesserungen, aber auch von Verschlechterungen, z. B. der Lebensqualität kann erschwert sein. Ein Bedeutungsverlust patientenberichteter Messungen und damit eine Schwächung der Patientenperspektive in der Nutzenbewertung ist zu befürchten.

Der vfa hält deshalb eine bedenkenlose Anwendung des festgesetzten Richtmaßes für nicht angebracht. Die Eignung des Richtmaßes sollte in G-BA-Bewertungen in jedem Einzelfall geprüft werden. Auch sind die Bestrebungen der Wissenschaft, bessere Standards zu definieren, zu unterstützen. Ein allgemein akzeptierter Katalog von geeigneten Qualitätskriterien zur Beurteilung solcher Auswertungen sollte entsprechend vorangebracht werden. Bis dahin sollten alle bisher akzeptierten Auswertungen aus Gründen der Verfahrenskonsistenz auch weiterhin herangezogen werden.

Patientenberichtete Endpunkte – Patientensicht im AMNOG

Patientenberichtete Endpunkte (PRO) zur Messung der Lebensqualität und Symptomatik haben aufgrund ihrer Patientenzentrierung einen besonderen Stellenwert in der Nutzenbewertung. Sie liefern wertvolle Informationen über die Sicht von Patientinnen und Patienten auf die Erkrankung, die Nebenwirkungen der Therapie oder die Lebensqualität. Sie ergänzen damit „harte Endpunkte“ wie bspw. Mortalität. Häufig vermitteln PRO so den entscheidenden Eindruck für den patientenrelevanten Zusatznutzen, insbesondere von Onkologika. Hier können bspw. Verschlechterungen der Lebensqualität einen möglichen Zusatznutzen im Überleben abschwächen oder sogar in Frage stellen. Andererseits können Verbesserungen in der Lebensqualität bzw. Symptomatik einen eigenständigen beträchtlichen Zusatznutzen begründen. PRO spielen aber natürlich auch in anderen Indikationen

eine wichtige Rolle, gerade bei chronischen Erkrankungen und Erkrankungen die stark das Empfinden des Menschen beeinflussen (psychische Erkrankungen). PRO sind wesentlich für die Patientenperspektive in der Nutzenbewertung, weshalb auch im AMNOG immer wieder bedeutsame Ergebnisse für solche Endpunkte eingefordert werden.

MID – essenziell für bedeutsame Ergebnisse der Patientensicht

Bei der Bewertung von patientenberichteten Messungen spielen sogenannte Responsekriterien bzw. MID („Minimal Important Differences“) eine tragende Rolle. Diese sagen etwas über die für Patientinnen und Patienten merkliche Veränderung bzw. die klinische Relevanz von therapeutischen Verbesserungen oder Verschlechterungen aus. Erst wenn dieser Wert erreicht wird, kann von „bedeutsamen“ Ergebnissen für Patientinnen und Patienten ausgegangen werden. Die Praxis hat

gezeigt, dass der Nachweis von bedeutsamen Ergebnissen ohne die Verwendung von MID kaum möglich ist, da eine alternative Auswertetechnik (d. h. standardisierte Mittelwertdifferenz [Hedges' g]) wenig empfindlich ist und nur extreme Veränderungen entdecken kann. MID werden methodisch mittels Validierungsstudien unter Beteiligung von Patientinnen und Patienten bestimmt und sind entsprechend des Charakters des Fragebogens in der jeweiligen Krankheitsentität individuell verschieden. Bisher wurden in AMNOG-Verfahren nur MID berücksichtigt, die in der wissenschaftlichen Literatur als etabliert bzw. validiert galten.

Diskussion um verbesserte Standards zur MID-Bewertung fortgeschritten

Die internationale wissenschaftliche Gemeinschaft diskutiert derzeit verbesserte Standards für die Bewertung von MID. So wurden bereits von mehreren Seiten eine Reihe von Qualitätskriterien vorgeschlagen. Empfohlen werden bspw. neben der Verwendung von „ankerbasierten Verfahren“ und „Längsschnittstudien“ eine „hinreichende Korrelation zwischen Anker und PRO“ [1,2,3,4]. Zuletzt wurde in diesem Jahr ein neues Instrument entwickelt, welches die Zuverlässigkeit von MID bestimmt. Dabei wurden bekannte Kriterien aufgegriffen und in einen substanziellen Kriterienkatalog erweitert [5]. Das Instrument kann als weiterer Baustein im wissenschaftlichen Weiterentwicklungsprozess der Standards zur Beurteilung von MID verstanden werden und ist nun Grundlage für Diskussionen. So haben andere Wissenschaftler bereits Nachbesserungsbedarf formuliert, bspw. bezüglich der erforderlichen Stärke der Korrelation [6]. Die Diskussion um verbesserte Standards zur MID-Bewertung mittels sinnvoller Kriterien ist so bereits fortgeschritten, aber noch nicht abgeschlossen.

Überraschende IQWiG-Festlegung mit anderem Ansatz

Trotz der anhaltenden Diskussion um mögliche Standards hat das IQWiG nun überraschend einen eigenen Bewertungsansatz für MID vorgelegt, der aus Sicht des vfa übereilt erscheint und in seiner Richtung hinterfragt werden muss. Denn er lässt den internationalen Entwicklungsansatz von möglichen Qualitätskriterien gänzlich außer Acht. Stattdessen soll durch die Festsetzung eines Richtmaßes

in einem „one-size-fits-all“-Ansatz eine strengere MID-Bewertung in „pragmatischer Weise“ bewerkstelligt werden. Das Richtmaß in Form eines generischen Schwellenwerts soll dabei eine schnelle Beurteilung der „Mindestgröße“ einer geeigneten MID erlauben. Gleichzeitig wird die Eignung fast aller bisherigen MID für die Nutzenbewertung grundsätzlich in Frage gestellt und die Notwendigkeit einer strengeren Handhabung betont. Anstatt möglicher Qualitätskriterien soll hierfür nun aber ein generisches Richtmaß zur Anwendung kommen.

Generisches Richtmaß nicht für alle Fälle gleichermaßen sinnvoll

Der Ansatz eines generischen Richtmaßes ist problematisch. Demnach sollen MID-Werte nur noch dann akzeptiert werden, wenn diese mindestens 15 % der Skalenspannweite des Erhebungsinstruments erreichen [7]. Dabei ist bereits die Auswahl dieses Schwellenwerts zu hinterfragen, der nicht unter Beteiligung von Patientinnen und Patienten ermittelt wurde, sondern sich als „mittlere Festlegung“ der vom IQWiG beobachteten MID-Spannweiten in der Literatur ergibt. Diese würden in vielen Fällen zwischen 10 % und 20 % liegen [7]. Eine vfa-Überprüfung aller bisher im AMNOG-Verfahren akzeptierten MID zeichnet dagegen ein anderes Bild [8]. Hier liegen die MID-Spannweiten zwischen 2 % und 40 % und erreichen im Mittel ca. 9 %. Die Variabilität erscheint also größer und der mittlere Wert niedriger als angenommen. Darüber hinaus ist ein generischer Schwellenwert nicht für alle Fälle gleichermaßen sinnvoll. So zeigt die vfa-Überprüfung deutlich, wie individuell verschieden einzelne MID für Patientinnen und Patienten mit einer bestimmten Erkrankung im speziellen Fragebogen sein können. Der festgelegte „one-size-fits-all“-Ansatz kann also bekannte Unterschiede der Patienten-sicht auf bedeutsame Ergebnisse nicht hinreichend berücksichtigen und muss deshalb aus wissenschaftlicher Sicht hinterfragt werden.

Fast alle bisher im AMNOG akzeptierten MID verlieren ihre Gültigkeit

Das festgesetzte Richtmaß ist regelhaft höher als die MID-Werte, die bislang in AMNOG-Verfahren akzeptiert wurden. Eine vfa-Überprüfung aller bisher im AMNOG-Verfahren akzeptierten MID zeigt, dass das Richtmaß in fast 90 % der Fälle teils deutlich (um das

1,1-fache bis 8,3-fache) höher liegt [8]. Durch die Anwendung des Richtmaßes würden damit fast alle bisher im AMNOG akzeptierten MID Ihre Gültigkeit verlieren. Etablierte patientenberichtete Auswertungen ständen vor dem Aus und die Messlatte für den Nachweis von therapeutischen Verbesserungen oder Verschlechterungen würde substantiell erhöht. Die teils deutlichen Unterschiede offenbaren dabei die individuellen, patientengerechten MID-Anforderungen, welche durch einen generischen Schwellenwert nicht hinreichend berücksichtigt werden können.

Patientenperspektive in der Nutzenbewertung geschwächt

Höhere Messlatten sind schwerer zu überwinden. Für patientenberichtete Auswertungen kann dies einen erschwerten methodischen Nachweis von Vor- und Nachteilen von Therapien bedeuten. Verbesserungen oder Verschlechterungen des patientenzentrierten Befindens zu Symptomatik und Lebensqualität könnten so weniger häufig entdeckt werden. Patientenberichtete Endpunkte würden zugunsten „harter“ Endpunkte an Bedeutung verlieren. Zudem kann der IQWiG-Ansatz bekannte Unterschiede der Patientensicht auf bedeutsame Ergebnisse nicht hinreichend abbilden. Die Patientenperspektive in der Nutzenbewertung wird so insgesamt geschwächt.

Der vfa empfiehlt daher:

1. Einzelfallprüfung statt „one-size-fits-all“

Der vfa hält eine bedenkenlose Anwendung des festgesetzten Richtmaßes für nicht angebracht. Die Eignung des Richtmaßes sollte in G-BA-Bewertungen in jedem Einzelfall geprüft werden. Denn eine MID-Bewertung im festgelegten „one-size-fits-all“-Ansatz kann bekannte Unterschiede der Patientensicht auf bedeutsame Ergebnisse nicht hinreichend berücksichtigen. Zudem wird der Entwicklungsansatz der Wissenschaft zur Verbesserung von Bewertungsstandards mittels sinnvoller Qualitätskriterien dabei außer Acht gelassen. Aus der Sicht des vfa sollte bei Änderungen der Methoden zur Beurteilung von MID die Wissenschaftlichkeit im Vordergrund stehen, nicht ein alleiniger Pragmatismus.

2. Allgemein akzeptierter Katalog von Bewertungskriterien

Die Bestrebungen der wissenschaftlichen Gemeinschaft um verbesserte Bewertungskriterien sollten nicht ignoriert, sondern unterstützt werden. Ziel sollte ein allgemein akzeptierter Katalog von Bewertungskriterien sein, der eine angemessene Beurteilung der Zuverlässigkeit von MID erlaubt. Dieser sollte auf Grundlage der bisherigen Empfehlungen und im weiteren gemeinsamen Dialog aus Wissenschaft, Institutionen und Industrie entwickelt werden. Methoden zur Beurteilung von MID sollten erst dann geändert werden, wenn diese Diskussionen eine ausreichende Einigkeit erreicht haben.

3. Verfahrenskonsistenz

Bis zur Festlegung einer verbesserten und allgemein akzeptierten Methode sollten die bisher geltenden Bewertungsmaßstäbe nicht geändert werden. Deshalb sollten alle bisher als etabliert bzw. validiert akzeptierten MID aus Gründen der Verfahrenskonsistenz auch weiterhin vom G-BA herangezogen werden.

Berlin, 05. November 2020

Quellenangaben

1. Coon, C. D., Cook K. F. Moving from significance to real-world meaning: methods for interpreting change in clinical outcome assessment scores. Qual Life Res 2018; 27:33–40 <https://link.springer.com/article/10.1007%2Fs11136-017-1616-3>
2. 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. [https://www.jclinepi.com/article/S0895-4356\(07\)00119-9/fulltext](https://www.jclinepi.com/article/S0895-4356(07)00119-9/fulltext)
3. FDA Guideline „Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims“ <https://www.fda.gov/media/77832/download>
4. Musoro ZJ, Hamel J, Ediebah DE on Behalf of the EORTC Quality of Life Group, et al. Establishing anchor-based minimally important differences (MID) with the EORTC quality-of-life measures: a meta-analysis protocol BMJ Open 2018;8: e019117. doi:10.1136/bmjopen-2017-019117
5. Devji T, Carrasco-Labra A, Qasim A, Phillips M, Johnston BC, Devasenapathy N, Zeraatkar D, Bhatt M, Jin X, Brignardello-Petersen R, Urquhart O, Foroutan F, Schandelmaier S, Pardo-Hernandez H, Vernooij RW, Huang H, Rizwan Y, Siemieniuk R, Lytvyn L, Patrick DL, Ebrahim S, Furukawa T, Nesrallah G, Schünemann HJ, Bhandari M, Thabane L, Guyatt GH. Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study. BMJ 2020 Jun 4;369:m1714. doi: 10.1136/bmj.m1714 <https://www.bmj.com/content/369/bmj.m1714>
6. John D Peipert Assistant Professor David Cella (Northwestern University Feinberg School of Medicine) Rapid Response to “Evaluating the credibility of anchor based estimates of minimal important differences for patient reported outcomes: instrument development and reliability study” (doi: <https://doi.org/10.1136/bmj.m1714>) 06. September 2020 <https://www.bmj.com/content/369/bmj.m1714/rapid-responses>
7. IQWiG Allgemeine Methoden Entwurf für Version 6.0 vom 05.12.2019
8. vfa-Analyse (2020) Übersicht von bisher in AMNOG-Verfahren akzeptierten MID im Vergleich zum Richtmaß 15% der Skalenspannweite. (Stand: 15.2.2019)