



Schriftenreihe

INTERDISZIPLINÄRE PLATTFORM ZUR NUTZENBEWERTUNG

Welche Endpunkte sind patientenrelevant?

Ausgabe 20
März 2025
ISSN 2364-916X

HEFTE DER INTERDISZIPLINÄREN PLATTFORM ZUR NUTZENBEWERTUNG

- HEFT 1 Vier Jahre AMNOG – Diskurs und Impulse
- HEFT 2 Klinische Studien – welche Endpunkte zählen?
- HEFT 3 Adaptive Pathways – Chancen und Risiken
- HEFT 4 AMNOG 2.0 – Informationsprobleme
- HEFT 5 Lücken in der Evidenz – was leisten Registerdaten?
- HEFT 6 Arztinformation via Software: Wege und Ziele
- HEFT 7 Arztinformation via Software – Orientierung oder Steuerung?
- HEFT 8 Europäische Nutzenbewertung – Chancen und Risiken
- HEFT 9 Kontextuelle Evidenz – Wege zur gezielten Therapie
- HEFT 10 Welchen (Zusatz-) Nutzen haben Registerdaten?
- HEFT 11 Europäisches HTA-Verfahren: Fortschritte und Fallstricke
- HEFT 12 Digitale Gesundheitsdaten: Nutzen, Kosten, Governance
- HEFT 13 Patienten und Fachgesellschaften: Zusätzliche Expertise fürs AMNOG
- HEFT 14 Leitlinien – ihre Rolle in AMNOG und Versorgung
- HEFT 15 Weiterentwicklung des AMNOG mit Augenmaß und Evidenz
- HEFT 16 AMNOG: Finanzstabilisierung – neue Behandlungsparadigmen
- HEFT 17 Auswirkungen von EU HTA auf das AMNOG-Verfahren
- HEFT 18 AMNOG 2.0: Auf dem Weg zu einem effizienten System
- HEFT 19 Zusammenspiel von HTA und Zulassung
- HEFT 20 Welche Endpunkte sind patientenrelevant?

ALLE HEFTE SIND ABRUFBAR UNTER:

[HTTPS://WWW.AERZTEZEITUNG.DE/KOOPERATIONEN/PLATTFORM-ZUR-NUTZENBEWERTUNG](https://www.aerztezeitung.de/kooperationen/plattform-zur-nutzenbewertung)

Inhalt

EDITORIAL

Relevanz klinischer Endpunkte für Patienten: Standortbestimmung und Ausblick	6
---	---

STEFAN SCHWARTZE

Nutzenbewertung – welche Zielgrößen sind relevant für Patienten?	8
---	---

MARTIN DANNER UND EVA STUMPE

Patientenrelevante Endpunkte: Hauptfragen aus Patientensicht	10
---	----

BERNHARD WÖRMANN

Primäre Endpunkte im AMNOG: Standortbestimmung und Ausblick am Beispiel Onkologie	12
--	----

SEBASTIAN KLUCKERT

Patientenrelevanz: Ein Kommentar aus juristischer Sicht	22
--	----

DANIELA PREUKSCHAT UND SEBASTIAN MELLER

Studienendpunkte in der frühen Nutzenbewertung aus Sicht des IQWiG	26
---	----

JUTTA WENDEL-SCHRIEF

Innovation in Studienendpunkten – Studienendpunkte für Innovationen	38
--	----

ANDREAS RHODE

Patientenrelevante Endpunkte: Fragen aus Sicht der Medizinischen Dienste	46
---	----

PETER FALKAI UND JÖRG RUOF

Erhebung von patientenberichteten Endpunkten: Einblicke aus der Psychiatrie	48
--	----

ANDREJ RASCH

Endpunkte – Herausforderungen aus der Sicht des vfa	52
--	----

JOHANNA SEEGER

Patientenrelevante Endpunkte im Kontext des europäischen HTA-Prozesses	60
---	----

TOMASZ BURZYKOWSKI

Validierung von Surrogatendpunkten	64
---	----

FLORIAN STAECK

Was ist ein patientenrelevanter Zusatznutzen? Gesucht wird ein strukturierter Diskurs für ein Konsensverfahren	74
---	----

Ziele der Plattform

Seit der Einführung des AMNOG im Jahr 2011 verfügt Deutschland über ein inzwischen etabliertes und weitgehend anerkanntes ‚lernendes System‘ zur Bewertung des patientenrelevanten Zusatznutzens (Health Technology Assessment, HTA). Die Bewertung des Zusatznutzens durch den Gemeinsamen Bundesausschuss (G-BA) ist das Ergebnis einer Expertenarbeit auf der Basis eines Gesetzes (AMNOG) und von Verfahrens- und Methodenvorschriften (z. B. IQWiG-Methoden).

Die handelnden Akteure auf Seiten des G-BA und der Krankenkassen sind als Wissenschaftler, als Klinik- und Vertragsärzte, als Medizinischer Dienst der Krankenkassen (MDK) und Mitarbeiter der Kassenverwaltungen, aber auch als Patientenvertreter qualifiziert, jedoch interessengeleitet. Ebenso qualifiziert und interessengeleitet werden von den pharmazeutischen Unternehmen Nutzendossiers beim G-BA eingereicht, die als Grundlage für die Bewertung des Zusatznutzens dienen.

Da von der Bewertung des Zusatznutzens die Arzneimittelversorgung der Bevölkerung maßgeblich beeinflusst wird, macht es Sinn, den Prozess aufmerksam und kritisch zu begleiten, um mögliche Verwerfungen aufzuzeigen und Dysbalancen entgegenzuwirken. Die interdisziplinäre Plattform zur Nutzenbewertung hat es sich zur Aufgabe gemacht, im kleinen Expertenkreis das Verfahren der Nutzenbewertung mit den folgenden Zielen zu begleiten:

- Verfahren von Bewertungen des Zusatznutzens auch im Verhältnis zur Zulassung zu diskutieren,
- darauf hinzuwirken, dass internationale Standards der evidenzbasierten Medizin und der Gesundheitsökonomie eingehalten, angewendet und weiterentwickelt werden,
- ob und inwieweit patientenrelevanter Zusatznutzen, insbesondere in den Bereichen Mortalität, Morbidität und Lebensqualität, anerkannt wird und welche methodischen Probleme dabei auftreten,

- mögliche Fehlentwicklungen insbesondere bezüglich der Versorgung der Patienten mit neuen Wirkstoffen zu identifizieren,
- einen konstruktiven Dialog mit allen Akteuren im Verfahren der Nutzenbewertung, z. B. auch bei der Weiterentwicklung der gesetzlichen Rahmenbedingungen zum AMNOG, zu ermöglichen.

Zudem verstärkt sich durch den 2018 von der Europäischen Kommission eingebrachten Verordnungsentwurf die europäische Perspektive beim Health Technology Assessment innovativer Arzneimittel. Die Begleitung des Spannungsfeldes zwischen der etablierten nationalen Bewertung und der angestrebten europäischen HTA-Harmonisierung ist ebenfalls ein zentrales Anliegen der Plattform.

Die interdisziplinäre Plattform möchte einen Beitrag dazu leisten, dass neue Wirkstoffe nachvollziehbar und fair bewertet werden. Der Beirat hält eine interdisziplinäre Diskussion über die Bewertungsergebnisse und die angewandten Methoden der Nutzenbewertung für unerlässlich. Darüber hinaus sieht er in dem Nutzenbewertungsprozess eine gute Möglichkeit, die verordnenden Ärzte früher als bislang üblich, über den zu erwartenden Zusatznutzen für Patienten zu neuen Arzneimitteln zu informieren.

Die Interdisziplinäre Plattform ist aus einem Diskussionsprozess zwischen Klinikern und Experten heraus entstanden. Der gemeinsame Wunsch, den Sachverstand in Form interdisziplinärer Tagungen zu bündeln, wird durch ein offenes Sponsorenkonsortium unterstützt. Diesem gehören AbbVie Deutschland GmbH & Co. KG, DAK Gesundheit, MSD Sharp & Dohme GmbH, Novo Nordisk Pharma GmbH, Roche Pharma AG und der Verband Forschender Arzneimittelhersteller e. V. an.

Der Beirat der Interdisziplinären Plattform zur Nutzenbewertung

Relevanz klinischer Endpunkte für Patienten: Standortbestimmung und Ausblick

Prof. Dr. Jörg Ruof

Liebe Leserinnen und Leser, ja, es stimmt: die Frage nach der Relevanz einzelner Endpunkte von klinischen Studien für Patienten ist so alt wie das AMNOG selbst – und trotzdem ist diese Frage noch immer nicht gelöst. So hatte auch das im Januar 2016 publizierte zweite Heft der vorliegenden Publikationsreihe bereits den Fokus auf: „Klinische Studien – welche Endpunkte zählen“. In seinem damaligen Eröffnungsreferat forderte Herr Thomas Kaiser – als damaliger Ressortleiter Arzneimittelbewertung im IQWiG – ein Umdenken: „Nicht das ist relevant, was untersucht wurde, sondern es sollte das Relevante untersucht werden.“

Der Begriff „patientenrelevante Endpunkte“ ist untergesetzlich verankert und wurde erstmals in der das AMNOG flankierenden Arzneimittelnutzenbewertungsverordnung Ende 2010 eingeführt. Dabei fällt auf, dass Paragraph 2 dieser Verordnung, der mit „Begriffsbestimmungen“ überschrieben ist, keine Definition von patientenrelevanten Endpunkten enthält. Diese wird implizit erst in Paragraph 5, der die Ausführungen zum Zusatznutzen enthält, eingeführt, und zwar einerseits in Bezug auf die Kategorien Mortalität, Morbidität und Lebensqualität (Paragraph 5 Abs. 2) und andererseits in Zusammenhang mit der jeweiligen Quantifizierung und Kategorisierung des Zusatznutzens z.B. als Heilung, Verlängerung der Überlebensdauer, Freiheit von Symptomen oder die Vermeidung von Nebenwirkungen (Paragraph 5 Abs. 7).

Einigkeit besteht dabei im Prinzip bei allen Systembeteiligten, dass Endpunkte von klinischen Studien und damit die den AMNOG-Verfahren zugrundeliegenden Endpunkte relevant für die betroffenen Patienten sein sollten. Weiterhin besteht bezüglich einer Vielzahl in den Verfahren verwendeter Endpunkte wie zum Beispiel Gesamtüberleben, Schmerz oder schwere Arzneimittelnebenwirkungen be-

reits Einigkeit zwischen allen Parteien, dass diese als relevant für Patienten einzuschätzen sind.

Der fortbestehende Dissens ergibt sich primär im Kontext der Diskussion um Surrogatparameter bzw. Surrogatendpunkte. Diese werden weder im AMNOG-Gesetzestext noch in der Arzneimittelnutzenbewertungsverordnung genannt oder definiert; die entsprechende Diskussion entwickelte sich erst im Kontext der ersten AMNOG-Verfahren und wurde in den methodischen Leitlinien zum Beispiel des IQWiG verankert. Der Fokus wird dabei in erster Linie auf Endpunkte gelegt, die zuverlässig und direkt konkrete Änderungen des Gesundheitszustandes abbilden. Surrogatendpunkte können als Ersatz eingestuft werden, die die eigentlichen patientenrelevanten Endpunkte früher und einfacher abbilden. Da die Aussagekraft der Surrogatendpunkte jedoch nicht immer verlässlich ist, werden diese im AMNOG nur nach erfolgreicher statistischer Validierung herangezogen.

Unklar bleibt hierbei nach wie vor – zumindest seitens des Ordnungsgebers – wie festgestellt werden kann bzw. auch wem die Definitionshoheit zusteht, bestimmte Endpunkte bei der Messung von Zusatznutzen oder Nebenwirkungen als primär patientenrelevant oder eben als Surrogat zu bezeichnen. Beispiele sind im Bereich der Onkologie progressionsfreies Überleben oder die Ansprechrate, bei Stoffwechselerkrankungen Laborwerte wie HbA_{1c} oder Enzymaktivitäten, bei Infektionen die SVR (Sustained Virologic Response), bei Nierenerkrankungen die eGFR oder Kreatinin Clearance.

Die jüngste Tagung der Plattform Nutzenbewertung und das vorliegende Berichtsheft bilden aus unterschiedlichen Perspektiven die ganze Bandbreite der zugehörigen Diskussion ab. Bei der Lektüre der Beiträge ergeben sich aus meiner Sicht drei zentrale Einsichten:

- Die Debatte um die Relevanz von Endpunkten für Patienten ist wichtig und wird – sowohl im Kontext des nationalen als auch des sich entwickelnden europäischen HTA-Verfahrens - weiterzuführen sein.
- Es ist nicht zu erwarten und wäre zudem vermutlich kaum sachdienlich, wenn die komplexen zugrundeliegenden Fragestellungen kategoriell politisch oder juristisch entschieden würden. Das heißt, der interdisziplinäre Dialog unter Beteiligung von Experten wie Patienten, Klinikern, Regulatoren, HTA-Institutionen, Methodikern oder Industrievertretern ist unerlässlich, um die Qualität und die spätere breite Akzeptanz der Beurteilung national und auch europäisch sicherzustellen.
- Die Einschätzung bestimmter Zielgrößen als relevant für Patienten hat Auswirkungen auf die klinische Versorgung, die weit über die reinen AMNOG-Verfahren hinausreichen. Dabei handelt es sich nicht nur um technische Detaildiskussionen zur Anwendbarkeit bestimmter statistisch-methodischer Instrumentarien im Kontext von Surrogat-Validierungen, sondern um eine klinisch-qualitativ, patienten-zentriert zu führende Diskussion, was in Forschung, Zulassung, HTA-Bewertung und klinischer Versorgung einen relevanten Nutzen für Patienten darstellt.

Ein herzlicher Dank geht an die Referenten und Autoren des Berichtsheftes, die ihre Beiträge in der Regel „on top“ zu ihren sonstigen Aufgaben verfasst haben – ein großartiges Engagement, das man nicht hoch genug einschätzen kann.

Ihnen, liebe Leserinnen und Leser, wünsche ich eine gewinnbringende Lektüre der Beiträge.

Jörg Ruof

Kontakt: joerg.ruof@r-connect.org

Nutzenbewertung – welche Zielgrößen sind relevant für Patienten?

Stefan Schwartz, MdB | Patientenbeauftragter der Bundesregierung

Als Patientenbeauftragter ist es mir ein Anliegen, die Qualität in der medizinischen Behandlung zu verbessern und die gesundheitliche Versorgung patientenorientierter auszurichten, damit unser Gesundheitssystem den Bedarfen und Bedürfnissen der Patientinnen und Patienten besser entspricht.

Um dies zu erreichen, ist es richtig und notwendig, patientenrelevanten Endpunkten bei der Bewertung medizinischer Leistungen einen hohen Stellenwert einzuräumen – neben der verfügbaren Evidenz als höchstem Stellenwert.

Was ist aber ein patientenrelevanter Zusatznutzen und wie wird dieser bestimmt und konsentiert? Wie wird mit Bewertungsunterschieden der verschiedenen Entscheider umgegangen?

Relevante Endpunkte sind natürlich die Verbesserung des Gesundheitszustands, die Verkürzung der Krankheitsdauer, die Verlängerung des Überlebens, die Verringerung von Nebenwirkungen oder eine Verbesserung der Lebensqualität.

Was ist aber mit den Surrogat-, also den Ersatz-Endpunkten wie etwa den Blutwerten? Bei diesen bestehen unterschiedliche Bewertungen und Einschätzungen der Entscheider – dem IQWiG, dem G-BA und auch der EMA. Seit nunmehr über zehn Jahren AMNOG hat sich dies nicht geändert. Es wäre wünschenswert und an der Zeit, ein Verfahren zu etablieren, an dessen Ende eine für alle verbindliche Einigung steht.

Mir ist aber auch wichtig, die Erfahrungen und Perspektiven der Patientinnen und Patienten stärker zu erfassen. Neben anderen Datenquellen sollten auch Instrumente zur systematischen Erfassung von patientenberichteten Behandlungsergebnissen genutzt werden. Schließlich sind es die Patientinnen und Patienten, die den Behandlungsprozess von Anfang an erleben.

Ihre Erfahrungen hinsichtlich ihrer Behandlung und der Frage, ob sie ihre Therapie als zielführend und erfolgreich empfunden haben, werden bislang kaum abgefragt; geschweige denn systematisch ausgewertet und für die Versorgung oder die Forschung genutzt.

Die regelmäßige Erhebung der Patientensicht auf den Therapieverlauf und ihre Erfahrungen hinsichtlich der Behandlungsergebnisse können einen wichtigen Beitrag leisten, die Wirksamkeit der medizinischen Intervention und die Adhärenz zu steigern.

Langfristig erhoffe ich mir, dass neben den patientenrelevanten Endpunkten auch die Erfahrungen und Perspektiven der Patientinnen und Patienten herangezogen werden und in die Bewertungen mit einfließen.



Stefan Schwartz, Industriemechaniker mit Fachabitur und langjähriger Berufserfahrung; seit 1994 Mitglied der SPD; seit 2006 Vorsitzender der SPD im Kreis Herford; seit 2014 Vorsitzender der SPD Region Ostwestfalen. Von 1999 bis 2009 Kreistagsabgeordneter; von 2004 bis 2009 stellvertretender Vorsitzender der SPD-Kreistagsfraktion. Seit Oktober 2009 Mitglied des Deutschen Bundestages. Von 2009 bis 2022 Mitglied im Petitionsausschuss; von 2013 bis 2022 Sprecher der Arbeitsgruppe Petitionen der SPD-Bundestagsfraktion. Seit 12. Januar 2022 Patientenbeauftragter der Bundesregierung.

Patientenrelevante Endpunkte: Hauptfragen aus Patientensicht

Dr. Martin Danner, BAG Selbsthilfe | Eva Stumpe, Deutsche Gesellschaft für Muskelkranke

Einer der wichtigsten Fortschritte des AMNOG war es, Innovationen im Arzneimittelbereich danach zu bemessen, ob ein patientenrelevanter Zusatznutzen gezeigt werden kann. Der Akzent liegt dabei auf der Patientenrelevanz. Nicht irgendein Zusatznutzen, sondern nur ein solcher, der für die Patientinnen und Patienten spürbar ist, wird als relevant angesehen.

Damit hat das AMNOG die Patientenperspektive deutlich in den Mittelpunkt der Betrachtungen gerückt. Andererseits wurden andere Endpunkte, insbesondere die sogenannten Surrogatparameter, in ihrer Bedeutung zurückgestuft.

Diese Rangordnung der Endpunkte ruft aber immer wieder Diskussionen hervor und führt auch zu Kontroversen. So lässt sich trefflich darüber streiten, ob die Viruslast bei Hepatitis oder das Wachsen eines Tumors für sich genommen patientenrelevante Sachverhalte sind, wenngleich ein „Spüren“ im alltäglichen Sinn damit nicht verbunden ist. Andere Aspekte der Patientenrelevanz scheinen dauerhaft ein Schattendasein zu führen. Dies betrifft insbesondere die sogenannten Patient Reported Outcomes (PRO).

Im Kontext der Einführung des europäischen Nutzenbewertungsverfahrens für Arzneimittel erhält die Diskussion zur Relevanz bestimmter Endpunkte eine noch größere Bedeutung. Absehbar ist, dass die unterschiedlichen Diskurse in den europäischen Ländern zu dieser Frage hier aufeinanderprallen und folglich in eine neue Ordnung gebracht werden müssen.

Was wird daher künftig relevant sein, um den Unterschied des Nutzens aufzuzeigen? Wie wird sich die Diskussion zur Patientenrelevanz von Endpunkten in Deutschland weiterentwickeln? Gerade die bei ATMP charakteristische Situation von Langzeitbetrachtungen wirft überdies neue Fragen zur Patientenrelevanz von Endpunkten auf.

Aus Patientensicht stellen sich aktuell insbesondere folgende Fragen:

- 1) Nach dem Methodenpapier des IQWiG definiert sich die Patientenrelevanz eines Endpunktes ganz klar aus der Patientenperspektive. Leider gelingt es aber nur selten, sogenannte PRO zum Gegenstand des Nutzenvergleichs zu machen. Warum fehlt es oftmals an validierten Messinstrumenten und wie kann man auf diesem Gebiet vorankommen?
- 2) In jüngster Zeit wird die Patientenrelevanz von Endpunkten teilweise methodisch dadurch in Abrede gestellt, dass bestimmte Effektunterschiede nicht als ausreichend angesehen werden. Dies wird sogar bei natürlichen Skalen diskutiert. Eine Erörterung der Körpergröße bei Wachstumsstörungen soll hiernach nicht per se als patientenrelevant anzusehen sein. Ist dies adäquat?
- 3) Warum werden Unterschiede in der Darreichungsform nicht als patientenrelevant angesehen, wo sie doch das Leben der Patient*innen massiv beeinflussen können?

Vertieft werden sollen diese Fragen durch folgende zusätzliche Aspekte:

- 4) Welche Rolle spielen Real-World-Data (RWD), die die gelebte Perspektive von Patient:innen außerhalb des klinischen Settings darstellen bei der Nutzenbewertung und der Entscheidung, ob ein Endpunkt tatsächlich patientenrelevant ist? Insbesondere bei Seltenen Erkrankungen sollten diese Daten, sofern bei der Nutzenbewertung schon vorhanden, die in den randomisierten klinischen Studien erhobenen Daten ergänzen, um ein valideres Bild vom wirklichen Zusatznutzen in einer breiteren Patientengemeinschaft zu bekommen.
- 5) Wenn RWD herangezogen werden - wie wird sichergestellt, dass Patientenvertreter:innen/ und Patientenorganisationen bei der Erstellung und der Lenkung der Register, in welchen

idealerweise die RWD gesammelt werden, beteiligt werden? Nur dann ist auch sichergestellt, dass diese Register auch die patientenrelevanten Daten einschließen.

6) Sind die validierten Fragebögen, die zur Ermittlung der behandlungs- und gesundheitsbezogenen Lebensqualität herangezogen werden, wirklich patientenzentriert erstellt worden (inwieweit haben Patientenorganisationen bzw. Patient:innen bei der Erstellung mitgearbeitet)?

7) Die Validierung eines Patient Reported Outcome Measures (PROM) ist aufwendig und sehr teuer. Bei seltenen Erkrankungen müssen häufig validierte PROMs herangezogen werden, die eigentlich für ein anderes Krankheitsbild entwickelt wurden und nicht 100 Prozent auf die seltene Erkrankung passen. Ist es nicht Aufgabe des Gesundheitssystems, in diesen Fällen Patientenorganisationen finanziell und logistisch zu unter-

stützen, damit diese (eventuell in Kooperation mit anderen Stakeholdern) auf ihren Krankheitsbereich besser passende „neue“ PROMs entwickeln und validieren können?

8) Bereits in der frühen Beratung des G-BA mit dem pharmazeutischen Unternehmen (pU) werden die für die Erstellung des Dossiers sinnvollerweise beizubringenden Daten diskutiert. Sowohl das IQWiG als auch der G-BA legen Wert auf die Patientenrelevanz der Endpunkte und der gelieferten Daten. Diese Patientenrelevanz sollte aber dann auch direkt bei der jeweiligen Patientengemeinschaft abgefragt werden. Tatsächlich erfolgt die Entscheidung über die Patientenrelevanz in Gremien und Ausschüssen, die nur zu einem kleinen Teil mit Patientenvertretern besetzt sind. Bei der frühen Beratung des G-BA mit dem pU sind keine krankheitsbezogenen Patientenvertreter:innen dabei.



Dr. Martin Danner ist Jurist und Bundesgeschäftsführer der Bundesarbeitsgemeinschaft SELBSTHILFE von Menschen mit Behinderung und chronischer Erkrankung und ihren Angehörigen e. V. (BAG SELBSTHILFE). Nach einem Studium in Heidelberg hat er einige Jahre als Rechtsanwalt mit der Spezialisierung im Gesundheitsrecht gearbeitet, bevor er die Leitung des Referats Gesundheitspolitik und Selbsthilfeförderung der BAG SELBSTHILFE übernommen hat. Er ist Sprecher der Patientenvertretung beim G-BA und unter anderem im Wissenschaftlichen Beirat des Ärztlichen Zentrums für Qualität in der Medizin und im IQWiG-Kuratorium aktiv.



Eva Stumpe ist Patientenvertreterin auf dem Gebiet der spinalen Muskelatrophie. Sie ist Kontaktperson bei der DGM e.V., Mitglied im Leitungsteam der Initiative SMA und Schatzmeisterin und Vorstandsmitglied von SMA Europe e.V. Frau Stumpe ist Mutter von zwei erwachsenen Kindern. Ihre Tochter Sarah lebt mit Spinaler Muskelatrophie Typ II. Zwei Jahre hat sie als Patientenvertreterin die Interessen von SMA-Patienten in Zulassungsverfahren bei der EMA vertreten. Nunmehr ist sie als Patientenvertreterin in mehreren Advisory-Boards der Industrie tätig. Beruflich ist sie Anwältin und engagiert sich im Familienunternehmen.

Primäre Endpunkte im AMNOG: Standortbestimmung und Ausblick am Beispiel Onkologie

Prof. Dr. Bernhard Wörmann | Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie (DGHO)

Die Festlegung von relevanten Endpunkten ist ein inhaltlich und methodisch kritischer Bestandteil der Konzeption klinischer Studien zu neuen Arzneimitteln. Die Selektion und Hierarchisierung der Endpunkte bestimmt das Design der Studie, die zu erhebenden Messparameter, die erforderliche Zahl an Studienteilnehmern, Abbruchkriterien und letztendlich den Zugang von Betroffenen zu einem neuen Arzneimittel. Die Studienergebnisse werden sowohl in den Verfahren zur Zulassung und zur Nutzenbewertung als auch zu Empfehlungen in Leitlinien genutzt – aber nicht mit derselben Methodik bewertet. Über die letzten 13 Jahren ist in Deutschland aus den AMNOG-Verfahren ein großer Erfahrungsschatz entstanden, allein in der Onkologie mit >300 abgeschlossenen Verfahren zu neuen Arzneimitteln bzw. neuen Indikationen. In diesem Artikel werden Stärken und Schwächen der häufigsten primären Endpunkte dargestellt und diskutiert.

Einleitung

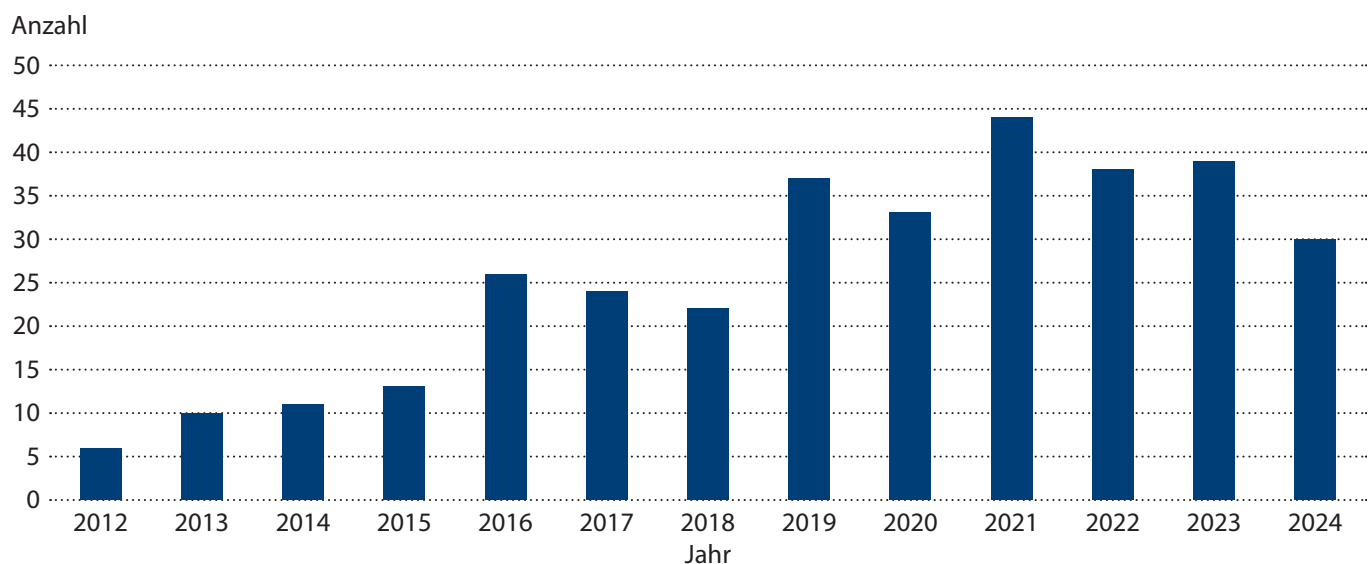
Vor allem die medikamentöse Therapie hat in den letzten beiden Jahrzehnten die Prognose von Patientinnen und Patienten (Pat.) mit einer Krebserkrankung deutlich verbessert. Höchstes Ziel bei der Behandlung dieser Pat. ist eine Heilung ohne negative Folgen. Voraussetzungen sind die Elimination der Krankheitsursache und die Verhinderung eines Rezidivs unter Einsatz nebenwirkungsarmer- oder freier Therapiestrategien. Bei vielen Krebserkrankungen ist dieses Ziel heute erreichbar. Die Chancen auf langfristige Heilung wurden insbesondere bei Pat. mit lokal begrenzter Erkrankung und hohem Rezidivrisiko erhöht, indem die wirksamen lokalen Maßnahmen wie Operation und/oder Bestrahlung mit dem Einsatz systemischer Therapie kombiniert wurden. Paradebeispiele sind Mamma-, Lungen- und kolorektales Karzinom.¹⁻⁶

Der relativ größte Fortschritt wurde bei Pat. mit fortgeschrittenen, metastasierten oder primär systemischen Krebserkrankungen erzielt. Eine neue Erfahrung beim Einsatz der hochwirksamen Arzneimittel in diesem Patientenkollektiv ist die zunehmende Transformation von sehr fortgeschrittenen Krebserkrankungen mit einer begrenzten Lebenserwartung von Monaten bis zu wenigen Jahren in chronische Krankheiten mit einer annähernd normalen Lebenserwartung. Paradebeispiele sind das metastasierte Melanom⁷, biologisch definierte Subgruppen des metastasierten nichtkleinzelligen Lungenkarzinoms^{3,8,9} und hämatologische Neoplasien wie die Chronische Myeloische Leukämie¹⁰, das Multiple Myelom¹¹ und die Chronische Lymphatische Leukämie.¹²

Erfahrungen aus dem AMNOG-Verfahren

Zwischen 2012 und 2024 wurden mehr als 300 Verfahren zu neuen Onkologika bzw. zu neuen Anwendungsgebiete-

Abgeschlossene Verfahren der frühen Nutzenbewertung in der Onkologie 2012–2024



Aufgenommen wurden abgeschlossene Verfahren zu neuen Arzneimitteln und zu neuen Indikationen. Jedes Verfahren wurde nur einmal gewertet. Neubewertungen nach Fristablauf, aufgrund neuer Daten oder regulatorischer Bestimmungen regulatorischer Bestimmungen wurden nicht aufgenommen.

Quelle: Eigene Darstellung

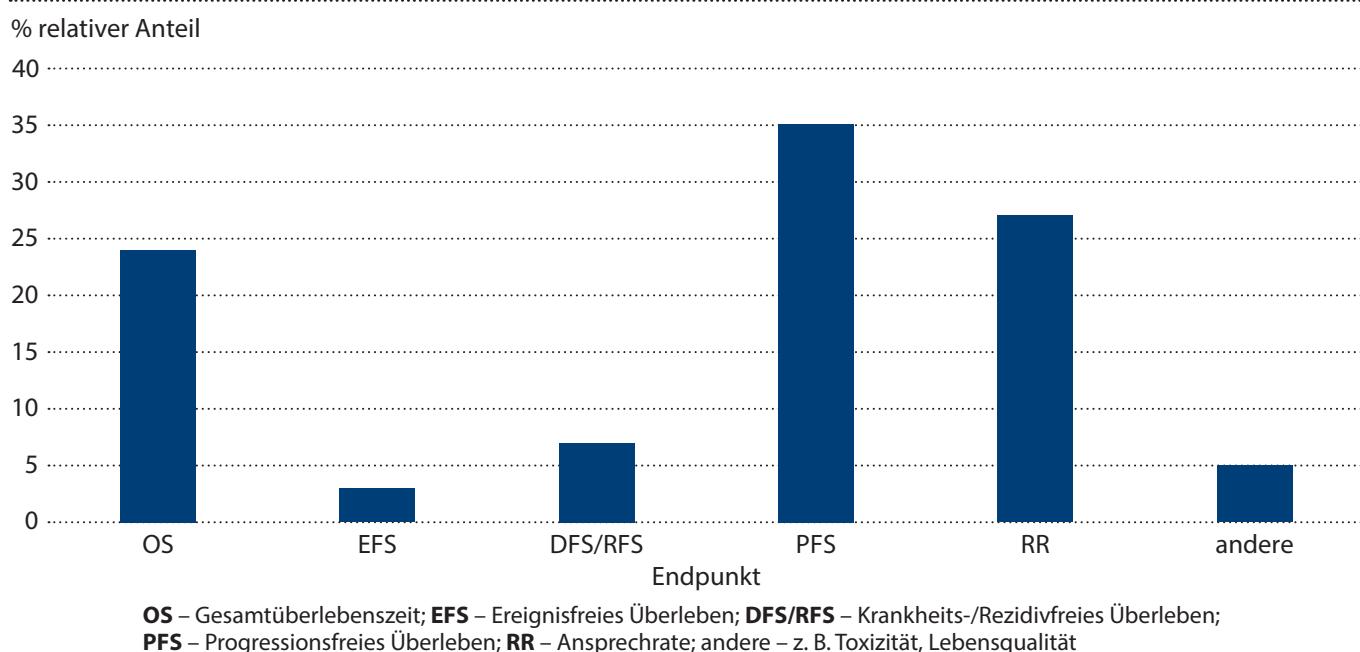
Abbildung 1: Überblick über die Zahl abgeschlossener Verfahren im Rahmen der frühen Nutzenbewertung 2012 bis 2024, begrenzt auf neue Arzneimittel bzw. neue Indikationen.



Prof. Dr. Bernhard Wörmann ist Arzt für Innere Medizin, Hämatologie und Internistische Onkologie, und hat die Zusatzbezeichnung Palliativmedizin. Seit 2010 ist er Medizinischer Leiter der DGHO Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie, seit 2015

Vorsitzender der Kommission „Nutzenbewertung von Arzneimitteln“ der AWMF. Ärztlich ist er im Ambulanten Gesundheitszentrum der Charité und in der Medizinischen Klinik mit Schwerpunkt Hämatologie, Onkologie und Tumorimmunologie am Campus Virchow in Berlin tätig.

Verteilung der primären Studienendpunkte in den abgeschlossenen Verfahren der frühen Nutzenbewertung in der Onkologie 2012–2024



Quelle: Eigene Darstellung

Abbildung 2: In der Mehrzahl der 333 ausgewerteten AMNOG-Verfahren war ein einziger primärer Endpunkt für die jeweilige Studie festgelegt. Zwei koprimäre Endpunkte wurden in 51 Verfahren verwendet.

ten im Rahmen des AMNOG-Prozesses durch den Gemeinsamen Bundesausschuss (G-BA) abgeschlossen. Die Anzahl aller abgeschlossenen Verfahren in der Onkologie liegt über 400, diese Zahl beinhaltet auch Neubewertungen, zum Beispiel nach Fristablauf.¹³ Abbildung 1 gibt einen Überblick über die Anzahl abgeschlossener Verfahren im Rahmen der frühen Nutzenbewertung 2012 bis 2024, begrenzt auf neue Arzneimittel bzw. neue Indikationen.

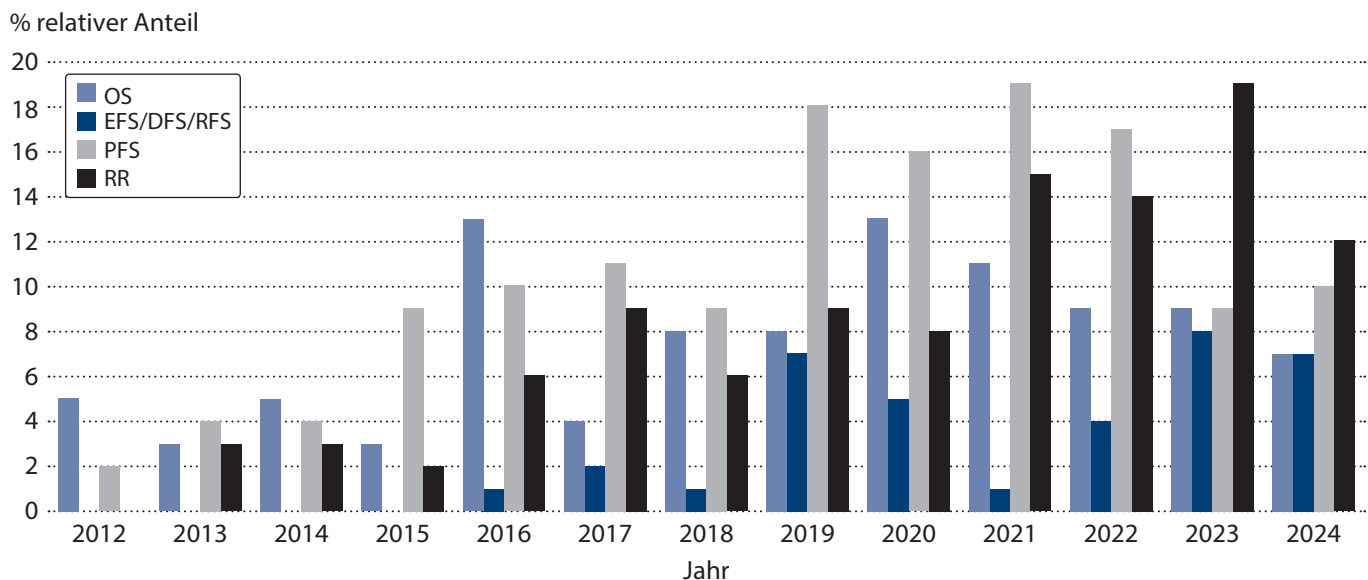
Primäre Endpunkte

In der Onkologie werden in den Zulassungsstudien zahlreiche, unterschiedliche Endpunkte verwendet. Von beson-

derer Bedeutung ist die Festlegung des primären Endpunktes. In der großen Mehrzahl der 333 ausgewerteten Verfahren war ein einziger primärer Endpunkt für die jeweilige Studie festgelegt. Bei 51 Verfahren wurden zwei koprimäre Endpunkte verwendet, zum Beispiel Gesamtüberlebenszeit und progressionsfreies Überleben. Abbildung 2 gibt einen Überblick über die Anzahl der verwendeten Endpunkte in den abgeschlossenen Verfahren der Jahre 2012 bis 2024.

Im untersuchten Zeitraum zeigt sich eine deutliche Verschiebung in der Häufigkeit der Verwendung primärer Endpunkte, siehe Abbildung 3.

Verteilung der primären Endpunkte in den abgeschlossenen Verfahren der frühen Nutzenbewertung in der Onkologie 2012–2024



Quelle: Eigene Darstellung

Abbildung 3: Die Verlängerung der Gesamtüberlebenszeit ist zwar für onkologische Patienten in der großen Mehrzahl der Indikationen der entscheidende Endpunkt. Sie war aber nur in 24 Prozent der abgeschlossenen Verfahren der primäre Endpunkt.

Rein deskriptiv ist die Anzahl von Verfahren mit dem primären Endpunkt ‚Gesamtüberleben (OS)‘ relativ konstant, während die Anzahl von Verfahren mit dem Endpunkt ‚Progressionsfreies Überleben (PFS)‘ vor allem in den Jahren 2019 bis 2022 deutlich zugenommen hat. In den letzten Jahren ist besonders die Anzahl von Verfahren mit dem primären Endpunkt ‚Ansprechrate (RR)‘ gestiegen.

Mortalität – Gesamtüberlebenszeit

Die Verlängerung der Gesamtüberlebenszeit ist für onkologische Pat. in der großen Mehrzahl der Indikationen der entscheidende Endpunkt. Sie war allerdings nur in 24 Pro-

zent der abgeschlossenen Verfahren der primäre Endpunkt. Die Verteilung dieses Endpunktes hat sich im Beobachtungszeitraum nicht wesentlich verändert, siehe Abbildung 3. Die Gesamtüberlebenszeit ist für die Erfassung der Wirksamkeit eines neuen Arzneimittels patientenrelevant, allerdings in der derzeit üblichen Form für eine Nutzenbewertung mit Schwächen behaftet.

Stärken

- Objektive Erfassung
- Gute Vergleichbarkeit mit Daten aus anderen Studien oder Registern
- Patientenrelevanz

Schwächen

- **Überlebensrate:** Die Überlebensrate wird selten als primärer Endpunkt oder überhaupt bewertet. Insbesondere bei der Chance einer nachhaltigen Verlängerung der Gesamtüberlebenszeit ist die Überlebensrate relevant für die Bewertung einer onkologischen Studie. Das traf in den letzten Jahren vor allem auf die Studien mit Immuncheckpoint-Inhibitoren zu.
- **Durchführbarkeit von Studien:** Für die Erfassung eines signifikanten Unterschieds ist eine sehr hohe Anzahl von Studienteilnehmern erforderlich, vor allem bei Erkrankungen mit guter Prognose.
- **Verzerrungspotenzial:** Die Erfassung der Gesamtüberlebenszeit erfordert die Erfassung aller nachfolgenden Therapien in dieser Indikation. Angesichts des weltweit sehr unterschiedlichen Zugangs zu weiteren neuen Arzneimitteln besteht ein hohes Verzerrungspotenzial. Auch ist bei einigen Studien in End-of-Life-Situationen ein Crossover-Design ethisch geboten. Es gibt derzeit keinen allgemein akzeptierten Konsens über die beste Methodik zur Berechnung von Crossover-Effekten. Bei hohen Crossover-Zahlen ändert sich auch der Inhalt der Fragestellung: Es wird nicht oder nicht nur der Wert eines neuen Arzneimittels oder Therapiekonzeptes, sondern sein Einsatz in einer frühen versus einer späten Krankheitsphase getestet.
- **Fehlende Validierung:** In vielen HTA-Methodologien ist das Ausmaß des relativen Unterschieds in der mittleren Überlebenszeit der relevante Bewertungsparameter.¹⁴ Damit führte auch 2024 ein absoluter Unterschied von <3 Monaten zum Vorschlag eines erheblichen Zusatznutzens, ohne nachhaltige Verlängerung der Überlebenszeit und vor allem auch ohne Erhöhung der Langzeitüberlebensrate.¹⁵

Morbidität – Krankheits-/Rezidivfreies Überleben (DFS/RFS) / Ereignisfreies Überleben (EFS)

Krankheits-/Rezidivfreies Überleben (DFS/RFS) und ereignisfreies Überleben (EFÜ, EFS) sind sinnvolle und gebräuchliche Endpunkte bei allen Therapieinterventionen mit kurativer Zielsetzung. Diese Endpunkte werden sowohl bei systemischen Erkrankungen wie akuten Leukämien als auch bei neoadjuvanter und adjuvanter Therapie eingesetzt.

Stärken

- Direkte Abbildung der Wirksamkeit der jeweiligen Intervention
- Kürzeres Zeitintervall bis zum Studienendpunkt im Vergleich zur Gesamtüberlebenszeit
- geringere Beeinflussung durch spätere Störfaktoren wie die Therapie im Rezidiv oder relevante, die Gesamtüberlebenszeit beeinflussende Komorbidität.

Schwächen

- **Composite Endpoint:** Alle zusammengesetzten Endpunkte haben ein höheres Verzerrungspotenzial durch unterschiedliche Gewichtung der eingeschlossenen Parameter. In einem System der Nutzenbewertung wie in Deutschland, bei dem HTA-Bewertung und Preisbildung im Wesentlichen auf dem Vergleich mit dem bisherigen Therapiestandard beruhen, ist die Einheitlichkeit der Definition von Composite Endpoints essenziell, aber auch schwerer zu erreichen.
- **Überlagerung der Erfassung der Wirksamkeit eines neuen Arzneimittels durch hohe Therapie-assoziierte Toxizität und Letalität.**

Morbidität – Progressionsfreies Überleben (PFÜ)

Progressionsfreies Überleben oder Zeit bis zum Progress waren die am häufigsten verwendeten, primären Endpunkte bei Studien zur Wirksamkeit neuer Arzneimittel in der Onkologie in den letzten Jahren, siehe Abbildung 1.

Progressionsfreies Überleben (PFÜ, PFS) ist ein sinnvoller und gebräuchlicher Endpunkt bei allen Therapieinterventionen mit nicht-kurativer Zielsetzung. Diese Endpunkte wurden in den letzten Jahren von der European Medicines Agency (EMA) und der US-amerikanischen Food and Drug Administration (FDA) bei sehr unterschiedlichen onkologischen Arzneimitteln und Therapieindikationen für die Zulassung neuer Arzneimittel akzeptiert. Die Zulassungsbehörden und HTA-Organisationen haben sich intensiv mit den methodischen Grundlagen beschäftigt.^{14, 16}

Stärken

- Direkte Abbildung der Wirksamkeit der jeweiligen Intervention bei fortgeschrittenen Erkrankungen und bei Therapie in nicht-kurativer Intention
- Rasche Erfassung der Wirksamkeit
- Reduktion der Anzahl erforderlicher Studienteilnehmer gegenüber dem Endpunkt Gesamtüberlebenszeit.

Schwächen

- Composite Endpoint: Alle zusammengesetzten Endpunkte haben ein höheres Verzerrungspotenzial durch unterschiedliche Gewichtung der eingeschlossenen Parameter.
- Notwendigkeit der transparenten Darstellung von Zensierung; der geschätzte Behandlungseffekt steigt mit dem Grad von Zensierungen und kann einen größeren Behandlungseffekt suggerieren.¹⁷
- Basierung auf technischen Endpunkten wie Bildgebung mit dem Risiko der Beeinflussung durch andere Therapie-induzierte Effekt wie Pseudo-Regression und Pseudo-Progression²⁰ [de Groot] und/oder durch Laborergebnisse mit dem Risiko der Überlagerung von Therapie-Nebenwirkungen
- Sehr heterogenes Bild bei der Korrelation von PFS mit dem Gesamtüberleben bei unterschiedlichen Tumorentitäten.¹⁸

Der Einwand einer fehlenden Patientenrelevanz des primären Endpunktes PFS kann durch Kombination mit der Auswertung sekundärer Endpunkte wie Symptomatik, Lebensqualität (Patient-Reported-Outcome) und Toxizität kompensiert werden.

Morbidität – Remissionsrate

Bei aggressiv verlaufenden Malignomen ist eine Verlängerung der Überlebenszeit nur bei Erreichen einer Remission, in der Regel einer kompletten Remission, möglich. Das kann bei metastasierten oder primär systemischen Erkrankungen durch medikamentöse Therapie erzielt werden. Bei indolent verlaufenden Malignomen ist diese Korrelation nicht so eindeutig. Die Bewertung der Remissionsrate als primärer Endpunkt per se ist schwierig, weil seine Verwendung im Rahmen von Zulassungsstudien vor allem mit der Qualität der Studien zusammenhängt, siehe Abbildung 4.

Die Grafik zeigt, dass der primäre Endpunkt der Remissionsrate vor allem bei nicht-randomisierten Studien verwendet wurde. Deren Zahl war in den letzten Jahren gestiegen.

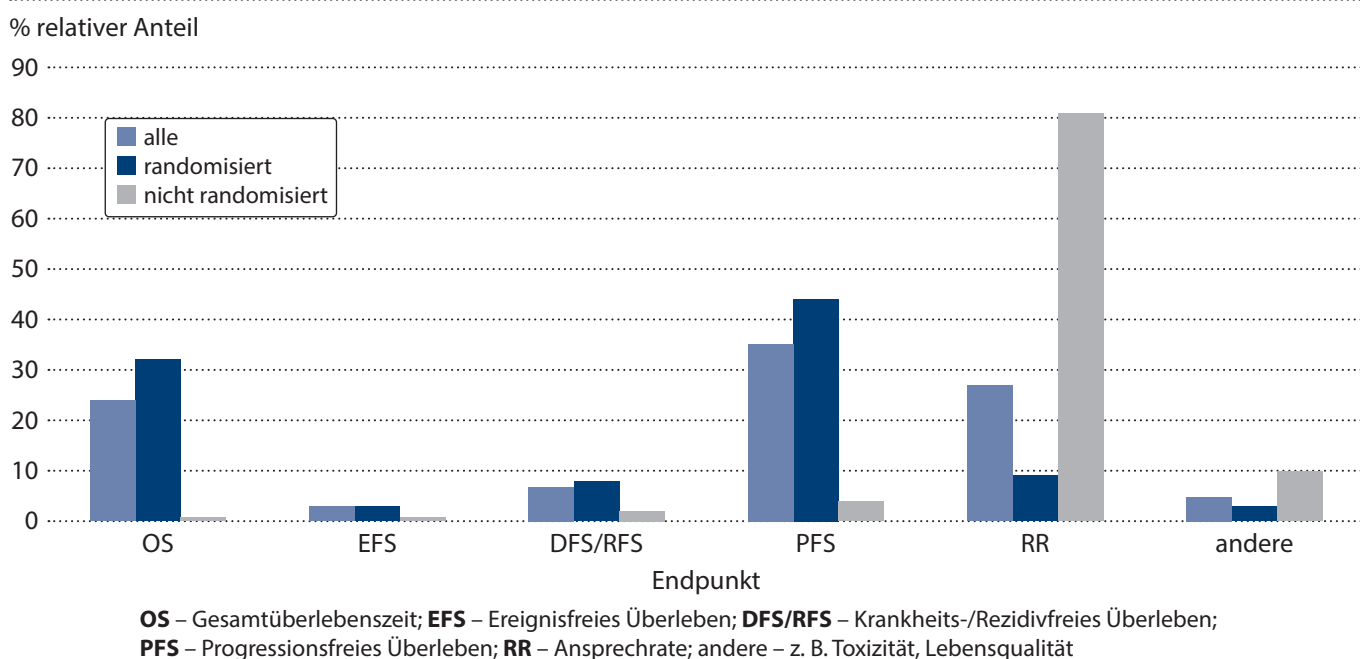
Stärken

- Direkte Abbildung der Wirksamkeit der jeweiligen Intervention
- Rasche Erfassung der Wirksamkeit
- Reduktion der Anzahl erforderlicher Studienteilnehmer gegenüber den Endpunkt Gesamtüberlebenszeit, krankheitsfreie und progressionsfreie Überlebenszeit.

Schwächen

- Basierung auf technischen Endpunkten wie Bildgebung mit dem Risiko der Beeinflussung durch andere Therapie-induzierte Effekt wie Pseudo-Regression und Pseudo-Progression [de Groot] und/oder durch Laborergebnisse mit dem Risiko der Überlagerung von Therapie-Nebenwirkungen

Korrelation zwischen Studiendesign und Auswahl primärer Endpunkte



Quelle: Eigene Darstellung

Abbildung 4: Der primäre Endpunkt der Remissionsrate ist vor allem bei nicht-randomisierten Studien verwendet worden. Deren Zahl ist in den vergangenen Jahren gestiegen.

- Keine direkte Korrelation mit Patient-Reported-Outcome.
- Schwache Korrelation mit dem Gesamtüberleben.

Morbidität – Symptomatik

Gerade im Bereich der Onkologie gibt es eine Vielzahl von sehr belastenden und die Lebensqualität einschränkenden Symptomen. Arzneimittel zur effektiven Symptomlinderung oder zur signifikanten Verlängerung der Zeit bis zum Auftreten belastender Symptome haben einen hohen Wert. Dennoch wird die Symptomatik nur selten als primärer Endpunkt definiert, konkret nur in 28 der 333 Verfahren. In Abbildung 1 sind diese Verfahren unter ‚andere‘

zusammengefasst. Eines dieser Ausnahmebeispiele ist die Stuhlfrequenz bei Pat. mit fortgeschrittenem Karzinoid bei der Bewertung von Telotristatethyl, weitere Beispiele sind die Symptomatik bei Pat. mit primärer Myelofibrose durch Fedratinib, Momelotinib oder Ruxolitinib.

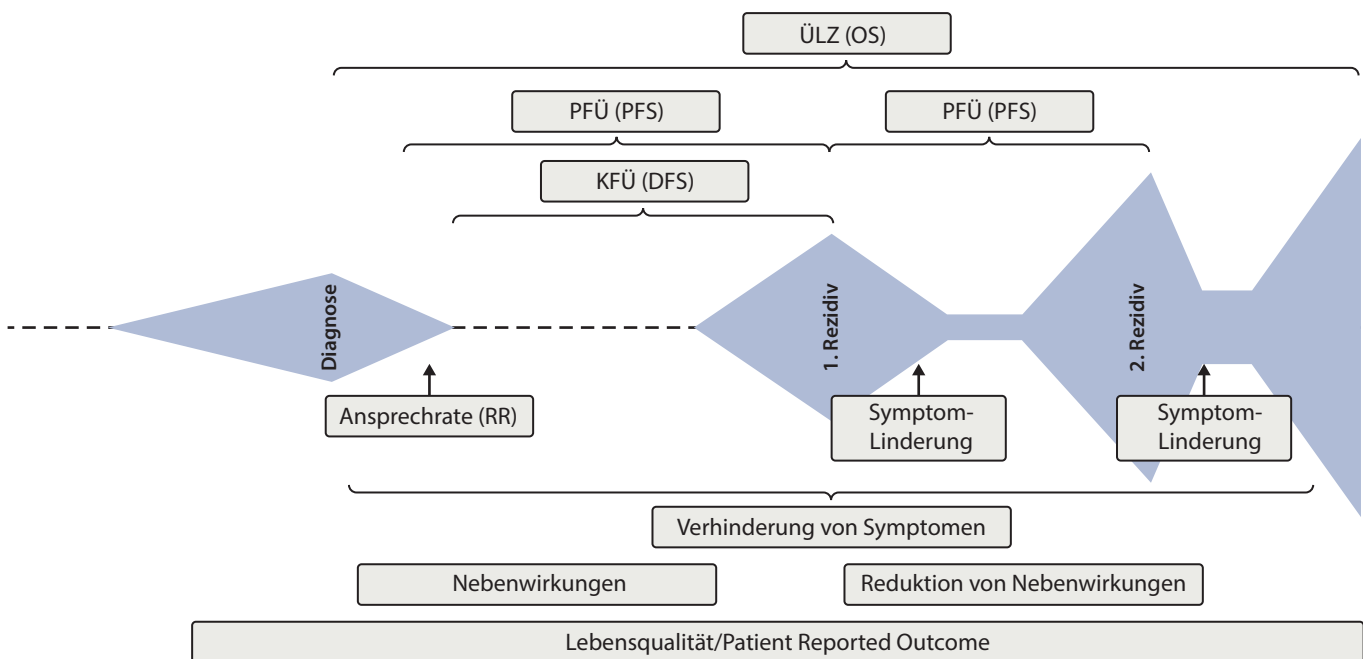
Lebensqualität / Patient – Reported Outcome (PRO)

Parameter der Lebensqualität waren in keiner der onkologischen Studien primärer oder koprimärer Endpunkt.

Zusammenfassung und Ausblick

Die Onkologie ist ein gutes Beispiel für die erforderliche

Patientenrelevante Endpunkte in der Onkologie



Quelle: Eigene Darstellung

Abbildung 5: Die Onkologie, hier orientiert am Krankheitsverlauf, ist ein Beispiel für die erforderliche Vielfalt von Endpunkten, um die Vorteile eines neuen Medikaments zu erfassen und zu bewerten.

Vielfalt von Endpunkten in der Erfassung und Bewertung der Vorteile eines neuen Medikamentes. Eine Übersicht über patientenrelevante Endpunkte gibt Abbildung 5, orientiert am Krankheitsverlauf.

Neue Arzneimittel werden als Grundlage der Zulassung, der Nutzenbewertung als Basis der Preisbildung und in Leitlinien bewertet. Jedes Verfahren hat seine eigene Methodologie. Aus Sicht der Versorgung ist es wichtig, –Parallelwelten‘ mit Absolutheitsanspruch in der Bewertung von Arzneimitteln zu vermeiden. Sich widersprechende und/oder nicht nachvollziehbare Bewertungen verunsichern Patienten und Verordner.

chern Patienten und Verordner.

Darüberhinaus ist es unerlässlich, Endpunkte auf ihre Wertigkeit zu überprüfen. Aus der Erfahrung von über 300 Verfahren zu neuen Arzneimitteln und Indikationen ergeben sich im Hinblick auf die HTA-Bewertung im Rahmen des AMNOG-Verfahrens für die Onkologie unter anderem folgende Herausforderungen:

- Anpassung der HTA-Kriterien an die verlängerten Überlebenszeiten, zum Beispiel durch Landmark-Analysen und/oder stärkere Integration weiterer Patienten-relevanter Endpunkte

- Entwicklung und Validierung einer Methodik zur gewichteten Bewertung unterschiedlicher Endpunkte unter Einbeziehung von Pat.
- Überprüfung der Kriterien für Nicht-Unterlegenheitsstudien.

Literatur

¹ Arbeitsgemeinschaft Gynäkologische Onkologie: Diagnostik und Therapie primärer und metastasierter Mammakarzinome: Endokrine und zielgerichtete Therapie metastasiertes Mammakarzinom. Status März 2024. AGO_2024D_26_Therapiealgorithmen_MASTER_schwarz (ago-online.de)

² AWMF Leitlinie „Früherkennung, Diagnostik, Therapie und Nachsorge des Mammakarzinoms“. Registernummer 032-045OL, Juli 2021. <https://go.sn.pub/flkeby>

³ Griesinger F et al.: Nicht-kleinzelliges Lungenkarzinom (NSCLC). Leitlinien von DGHO, OeGHO, SGMO und SGH+SSH, Status Januar 2025. <https://www.dgho-onkopedia.de/de/onkopedia/leitlinien/lungenkarzinom-nicht-kleinzellig-nscl>

⁴ Interdisziplinäre S3-Leitlinie: Prävention, Diagnostik, Therapie und Nachsorge des Lungenkarzinoms, 020-007, November 2022, <https://go.sn.pub/nzyfwn>

⁵ Kolonkarzinom. Leitlinien von DGHO, OeGHO, SGMO und SGH+SSH, Status Januar 2024. <https://go.sn.pub/mdg4um>

⁶ <https://go.sn.pub/i1v82z>

⁷ Wolchok JD, Chiarion-Sileni V, Rutkowski P et al.: Final, 10-Year Outcomes with Nivolumab plus Ipilimumab in Advanced Melanoma. N Engl J Med 392:11-22, 2025. DOI: 10.1056/NEJMoa2407417

⁸ Wu YL, Dziadziuszko R, Ahn JS et al.: Alectinib in Resected ALK-Positive Non-Small-Cell Lung Cancer: N Engl J Med 390_1265-1276, 2024. DOI: 10.1056/NEJMoa2310532

⁹ Planchard D, Jänne PA, Cheng Y et al.: Osimertinib with or without Chemotherapy in EGFR-Mutated Advanced NSCLC. N Engl J Med 389:1935-1948, 2023. DOI: 10.1056/NEJMoa2306434

¹⁰ Hochhaus A, Larson RA, Guilhot F, et al. Long-term outcomes of imatinib treatment for chronic myeloid leukemia. N Engl J Med 376:917-927, 2017. DOI: 10.1056/NEJMoa1609324

¹¹ Sonneveld P, Dimopoulos MA, Boccadoro M et al.: Daratumumab, Bortezomib, Lenalidomide, and Dexamethasone for Multiple Myeloma. N Engl J Med 390:301-313, 2024. DOI: 10.1056/NEJMoa2312054

¹² Wendtner C et al.: Chronische lymphatische Leukämie, September 2024. Chronische Lymphatische Leukämie (CLL) – Onkopedia

¹³ Nutzenbewertung nach § 35a SGB V (gefiltert) - Gemeinsamer Bundesausschuss

¹⁴ Allgemeine Methoden - Version 7.0

¹⁵ Nutzenbewertungsverfahren zum Wirkstoff Fruquintinib (Kolorektalkarzinom, vorbehandelte Patienten) - Gemeinsamer Bundesausschuss

¹⁶ European Medicines Agency: Guideline on the clinical evaluation of anticancer medicinal products, 18. November 2023. Anticancer GL Rev 6 adopted CHMP November 2023

¹⁷ Lengliné E, Baba J, de Boissieu P et al.: Composite event-free-survival as an endpoint in oncology drug evaluation: Review and guidance perspectives from the Haute Autorité de Santé (HAS). Eur J Cancer 204:114047, 2024. DOI: 10.1016/j.ejca.2024.114047

¹⁸ Belin L, Tan A, de Rycke Y, Dechartres A: Progression-free survival as a surrogate for overall survival in oncology trials: a methodological systematic review. Br J Cancer 122:1707-1714, 2020. DOI: 10.1038/s41416-020-0805-y

Patientenrelevanz: Ein Kommentar aus juristischer Sicht

Prof. Dr. Sebastian Kluckert | Bergische Universität Wuppertal – Lehrstuhl für Öffentliches Recht, insbesondere Öffentliches Wirtschaftsrecht und Sozialrecht | Unparteiischer Vorsitzender der Schiedsstelle nach § 130b Abs. 5 SGB V

Der Begriff der „Patientenrelevanz“ spielt eine zentrale Rolle im Kontext der Nutzenbewertung von Arzneimitteln und unterliegt einem Beurteilungsspielraum des Gemeinsamen Bundesausschusses (G-BA). Gerichte kontrollieren die Beurteilung des G-BA nur auf ihre Vertretbarkeit. Grenzen der Vertretbarkeit sind überschritten, wenn Entscheidungen gegen anerkannte Standards der evidenzbasierten Medizin und Gesundheitsökonomie verstoßen. Divergierende Einschätzungen innerhalb der Fachdisziplinen vergrößern den Beurteilungsspielraum des G-BA. Die Debatte um die Anerkennung bestimmter Endpunkte erweist sich daher nicht in erster Linie als eine juristische Problematik.

Patientenrelevanz“ ist ein Rechtsbegriff aus dem Recht der Nutzenbewertung von Arzneimitteln. Das Recht der Nutzenbewertung von Arzneimitteln findet seine parlamentsgesetzliche Grundlage in § 35a SGB V und es wird durch untergesetzliche Normen und Verfahrensvorschriften des Gemeinsamen Bundesausschusses (G-BA) konkretisiert. Während es bei der arzneimittelrechtlichen Zulassung vor allem um Patientensicherheit und Wirksamkeit von Arzneimitteln geht, bildet die Arzneimittelnutzenbewertung die Grundlage für die Bepreisung von Arzneimitteln mit neuen Wirkstoffen (vgl. § 130b SGB V, § 78 Abs. 3a AMG). Dabei soll der Arzneimittelpreis bzw. Erstattungsbeitrag im Verhältnis zu einer zweckmäßigen Vergleichstherapie nutzenadäquat sein.

Der Rechtsbegriff der Patientenrelevanz findet sich u. a. in folgenden Vorschriften: Nach § 2 Abs. 3 der Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (AM-NutzenV) ist „[d]er Nutzen eines Arzneimittels [...] der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustands, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“. Gemäß § 5 Abs. 5 Satz 1 AM-NutzenV wird der „Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie festgestellt als Verbesserung der Beeinflussung patientenrelevanter Endpunkte zum Nutzen gemäß § 2 Absatz 3“ AM-NutzenV. Der G-BA hat nach § 35a Abs. 3b Satz 4 SGB V für die anwendungsbegleitende Datenerhebung und Auswertung „Vorgaben [...] zu patientenrelevanten Endpunkten und deren Erfassung zu bestimmen“.

Beurteilungsspielraum und Vertretbarkeitskontrolle

Bei der Rechtsanwendung geht es im Wesentlichen um die Frage, ob ein konkret-individueller Lebenssachverhalt (bspw. eine durch Studien nachgewiesene Verbesserung beim progressionsfreien Überleben) von einem abstrakt-generell formulierten Gesetzesbegriff (z.B. „patientenrelevanter Endpunkt“) erfasst wird. Juristen sprechen davon, ob der Lebenssachverhalt unter das Gesetz subsumiert werden kann. Damit ist nichts anderes als die Frage gemeint, ob der Normgeber mit seiner Rechtsnorm den dem Rechtsanwender vor Augen schwebenden Lebenssachverhalt tatsächlich mit der Norm regeln wollte. Letztverbindlich entscheiden eine solche Frage die Gerichte.

Die Kompetenz der Gerichte ist jedoch dann beschränkt,



Prof. Dr. Sebastian Kluckert studierte BWL und Rechtswissenschaft in Berlin. Nach Promotion und Rechtsreferendariat arbeitete er als Wissenschaftlicher Assistent am Fachbereich Rechtswissenschaft der FU Berlin. Im Dezember 2016 erfolgte die Habilitation unter Zuerkennung der Lehrbefugnis für die Fächer Staats- und Verwaltungsrecht, Europarecht, Öffentliches Wirtschaftsrecht und Sozialrecht. Seit April 2018 ist er Professor an der Bergischen Universität Wuppertal. Zuvor war er auch als Rechtsanwalt in einer auf das Gesundheitswesen spezialisierten Kanzlei tätig. Seit dem 1. Juli 2023 ist er unparteiischer Vorsitzender der AMNOG-Schiedsstelle.

wenn das Gesetz – was selten ist – einer Behörde einen Handlungsspielraum in der Weise zugesteht, dass im Verhältnis zum Gericht vorrangig die Behörde selbst beurteilen darf, ob Lebenssachverhalt und Gesetz zusammenpassen. Ein solcher Handlungsspielraum wird Beurteilungsspielraum genannt. Dort wo ein Beurteilungsspielraum bejaht wird, prüft das Gericht zwar, ob die Behörde die Zuständigkeits-, Verfahrens- und Formvorschriften eingehalten hat, vom zutreffenden Sachverhalt ausgegangen ist und keine sachfremden Erwägungen angestellt hat. Ansonsten kommt es aber nur zu einer Vertretbarkeitskontrolle.

„Patientenrelevanz“ ist ein sog. unbestimmter Rechtsbegriff. Als unbestimmte Rechtsbegriffe werden solche Gesetzesbegriffe bezeichnet, die hochgradig auslegungsbedürftig sind, sich aufgrund einer geringen inhaltlichen Bestimmtheit schwer fassen lassen und hinter denen komplexe Wertungen, Bewertungen und Abwägungen ablaufen. Beispiele dafür sind Gesetzesbegriffe wie „Eignung“, „besonderer Härtefall“ oder „Verunstaltung des Landschaftsbildes“. Anders als bei rein deskriptiven Gesetzesbegriffen (wie „Hund“, „Katze“, „Maus“) ist bei unbestimmten Rechtsbegriffen die Frage, ob ein Lebenssachverhalt unter den Gesetzesbegriff zu subsumieren ist, nicht auf den ersten Blick zu beantworten. Unbestimmte Rechtsbegriffe sind ein möglicher Einfallstor für die Bejahung eines der Behörde vom Gesetzgeber eingeräumten Beurteilungsspielraums.

Im Rahmen der Nutzenbewertung von Arzneimitteln werden Beurteilungsspielräume durch das Bundessozialgericht (BSG) anerkannt. Nutzenbewertungsbeschlüsse werden vom BSG als Akte der untergesetzlichen Normsetzung qualifiziert. Das folgt aus der Rechtsnatur der Arzneimittelrichtlinie des G-BA (§ 92 Abs. 1 Satz 2 Nr. 6 SGB V), dessen Bestandteil die Nutzenbewertungsbeschlüsse werden (§ 35a Abs. 3 Satz 7 SGB V). Der Beurteilungsspielraum wird hier aus der einem Normgeber zustehenden Gestalt-

tungsfreiheit abgeleitet. Dies hat zur Folge, dass die sozialgerichtliche Kontrolle nicht ihre eigenen Wertungen an die Stelle der des G-BA setzen darf (vgl. BSG, Urt. v. 12.8.2021 – B 3 KR 3/20 R = BSGE 133, 1 Rn. 33). Allerdings gilt diese Einschränkung nur soweit die Wertung des G-BA vertretbar ist (Vertretbarkeitskontrolle).

Grenzen der Vertretbarkeit

Wo liegen nun die Grenzen der Vertretbarkeit einer Antwort auf die Frage, ob ein bestimmter Endpunkt unter den Gesetzesbegriff „patientenrelevant“ zu subsumieren ist? Unvertretbar wäre eine Beurteilung, die einen im höherrangigen Recht (insb. SGB V, AM-NutzenV) anerkannten Endpunkt als nicht patientenrelevant qualifiziert. So enthalten § 2 Abs. 3 und § 5 Abs. 2 Satz 3 AM-NutzenV sogenannte Regelbeispiele, d.h. ausdrücklich, aber nicht abschließend genannte Sachverhaltskonstellationen, die nach Ansicht des Gesetzgebers in der Regel unter den Gesetzesbegriff zu subsumieren sind. Das sind hier die Endpunkte Gesundheitszustand/Morbidität, Krankheitsdauer/Morbidität, Überleben/Mortalität, Nebenwirkungen und Lebensqualität. Daraus folgt, dass die Frage, ob ein umstrittener Endpunkt im Rahmen der Nutzenbewertung relevant ist, auch politisch durch den Gesetz- oder Verordnungsgeber entschieden werden kann.

Reicht es – sofern kein Widerspruch zum höherrangigen Recht festzustellen ist – für die Vertretbarkeit aus, dass es nach den einschlägigen Wissenschaften (insbesondere Medizin) vertretbar ist, einen bestimmten Endpunkt als nicht patientenrelevant zu qualifizieren? Bejahte man diese Frage, reichte es aus, dass der G-BA für seine Beurteilung erstzunehmende Stellungnahmen findet, auch wenn diese der überwiegend vertretenen Ansicht in der Wissenschaft („herrschende Meinung“) widersprechen. Nach zutreffender Ansicht gibt es im Rahmen der Nutzenbewer-

tung keine Vertretbarkeit einer Beurteilung des G-BA, die dem allgemein anerkannten Stand der HTA-Wissenschaft widerspricht. Diese Ansicht ist u. a. aus § 35a Abs. 1 Satz 8 Nr. 2 SGB V und § 7 Abs. 2 AM-NutzenV abzuleiten.

Nach diesen Bestimmungen bilden nämlich die internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie die Grundlage der Nutzenbewertung. Maßstab für die Beurteilung im Rahmen der Nutzenbewertung ist der allgemein anerkannte Stand der medizinischen Erkenntnisse. Daher müssen neben den gesetzlichen Regelbeispielen stehende Endpunkte, die nach den internationalen Standards als patientenrelevant anerkannt werden, auch vom G-BA als patientenrelevant anerkannt werden. Dazu gehören meines Erachtens auch Endpunkte, die nach den internationalen Standards als aussagekräftiges Surrogat eines (im Gesetz erwähnten oder weiteren international anerkannten) patientenrelevanten Endpunkts anerkannt sind.

Hinsichtlich der Surrogate findet sich allerdings in früheren Entscheidungen des BSG und des für die Nutzenbewertungen des G-BA erstinstanzlich zuständigen Landesozialgerichts Berlin-Brandenburg (zwischen 2011 und 2013) oftmals die Aussage, dass Studien, die als Primärziel bloße Surrogatparameter formuliert haben, zum Nachweis einer therapeutischen Verbesserung von vornherein nicht in Betracht kommen (wohl zuletzt BSG, Urt. v. 17.9.2013 – B 1 KR 54/12 R = BSGE 114, 217 Rn. 48; wohl zuletzt LSG Berlin-Brandenburg, Urt. v. 7.6.2013 – L 7 KA 164/09 KL, juris Rn. 126). Dieser kategorische Ausschluss dürfte mittlerweile stillschweigend aufgegeben worden sein.

Mögliche Impulse aus dem Unionsrecht

Impulse für die Vertretbarkeit der Beurteilung der Patientenrelevanz eines Endpunkts können sich auch aus der Europäischen Nutzenbewertung (EU-HTA) ergeben. Methodisch könnte sich ein verbindlicher Impuls als Aus-

fluss einer unionsrechtskonformen Auslegung und Anwendung des nationalen Rechts (AM-NutzenV) zeigen. Alternativ könnte sich ein solcher Impuls auch unterhalb der aus dem Anwendungsvorrang des Unionsrechts folgenden strengen Verbindlichkeit in der Weise entfalten, dass die Anerkennung eines Endpunkts als patientenrelevant im Rahmen des EU-HTA von der nationalen Rechtsprechung als Nachweis oder als starker Anhaltspunkt dafür betrachtet wird, dass der Endpunkt – nach den internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie bzw. nach dem anerkannten Stand der medizinischen Erkenntnisse – patientenrelevant ist.

Jedoch würden beide Impulse wohl eine Fortentwicklung des gegenwärtigen EU-HTA-Prozesses erfordern. Allerdings ist es bereits heute nach der Verordnung (EU) 2021/2282 über die Bewertung von Gesundheitstechnologien nicht vollkommen ausgeschlossen, dass der Festlegung der relevanten Parameter für den Bewertungsumfang (u.a. gesundheitsbezogene Endpunkte) im Rahmen der gemeinsamen klinischen Bewertungen eine irgendwie geartete unionsrechtliche Ausstrahlungswirkung auf nationale Bewertungsverfahren vom EuGH zugebilligt werden könnte, wenngleich eine solche Ausstrahlungswirkung nach den offiziellen Einordnungen zur Reichweite der Verordnung (EU) 2021/2282 gegenwärtig politisch nicht gewollt ist.

Klärung der Patientenrelevanz im Rechtsweg

Grundsätzlich ist die Klärung der Patientenrelevanz eines Endpunkts im Wege eines Gerichtsverfahrens möglich. Nutzenbewertungen des G-BA, die zu Unrecht (allerdings: Beurteilungsspielraum!) einen patientenrelevanten Endpunkt nicht anerkennen, für den nach der Studienlage eine Verbesserung eintritt, sind rechtswidrig. Rechtswidrige nachteilige Nutzenbewertungen können von einem pharmazeutischen Unternehmer mittels Klage (zuständig: LSG Berlin-

Brandenburg und BSG) angegriffen werden. Dieser Angriff kann im Rahmen der Anfechtung eines den Erstattungsbeitrag betreffenden Schiedsspruches nach § 130b Abs. 4 SGB V erfolgen, wobei dann die Anfechtung des Schiedsspruches mit dem zusätzlichen Antrag auf Feststellung der Unwirksamkeit des dem Schiedsspruch zugrundeliegenden Nutzenbewertungsbeschlusses des G-BA verbunden wird.

Alternativ ist auch eine allein gegen den Nutzenbewertungsbeschluss gerichtete Feststellungsklage mit dem G-BA als Beklagtem möglich, wenn es trotz der Differenzen zur Nutzenbewertung zu einer einvernehmlichen Vereinbarung eines Erstattungs Betrags zwischen pharmazeutischem Unternehmer und dem GKV-Spitzenverband gekommen ist. Soweit ersichtlich, gab es bisher keine Gerichtsverfahren, die explizit das Ziel verfolgten, einen Endpunkt als patientenrelevant zu erschließen.

Fazit

Patientenrelevanz stellt einen unbestimmten Rechtsbegriff mit Beurteilungsspielraum zu Gunsten des G-BA dar. Mit der Annahme eines Beurteilungsspielraums und der mit ihm einhergehenden verminderten gerichtlichen Kontrolle in Gestalt einer bloßen Vertretbarkeitskontrolle werfen Rechtsprechung und Rechtswissenschaft den Ball zurück in das Spielfeld von Medizin, Gesundheitsökonomie und weiteren HTA-Wissenschaften. Der Streit darüber, ob ein bestimmter Studienendpunkt patientenrelevant ist oder nicht, ist folglich weniger auf eine unzureichende rechtliche Verankerung zurückzuführen. Vielmehr begründet sich die Problematik in dem Umstand, dass unter den Vertretern der einschlägigen HTA-Wissenschaften die Qualifizierung eines bestimmten Studienendpunktes als patientenrelevant umstritten ist. Solange dies der Fall ist, bleibt die Ablehnung der Anerkennung der Patientenrelevanz eines Endpunkts durch den G-BA vertretbar und rechtmäßig.

Studienendpunkte in der frühen Nutzenbewertung aus Sicht des IQWiG

Dr. Daniela Preukschat, Dr. Sebastian Meller | Ressort Arzneimittelbewertung im IQWiG

Bei der frühen Nutzenbewertung ist die zentrale Frage, ob ein neues Arzneimittel im Vergleich zum bisherigen Therapiestandard einen patientenrelevanten Vorteil bietet. Ein solcher Vorteil kann auf Basis von patientenrelevanten Endpunkten oder validen Surrogaten abgeleitet werden. Das Problem ist: Daten zur Beantwortung dieser Frage fehlen weiterhin zu oft. Viel zu häufig liegen keine geeigneten Studien für die Bewertung vor. Wenn geeignete Studien vorliegen, sind dies in der Regel Zulassungsstudien, deren primäre Endpunkte oft nicht patientenrelevant sind bzw. nicht validierte Surrogatparameter enthalten.

Die Patientenrelevanz ist fest in der frühen Nutzenbewertung verankert. Der Nutzen eines Arzneimittels wird gemäß der Arzneimittel-Nutzenverordnung¹ definiert als „der patientenrelevante, therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustandes, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität. Der Zusatznutzen im Sinne dieser Verordnung ist ein Nutzen, der quantitativ oder qualitativ höher ist als der Nutzen, den die zweckmäßige Vergleichstherapie aufweist“ (Paragraf 2 Absatz 4 AM-NutzenV)¹. Patientenrelevant ist entsprechend dem Methodenpapier des IQWiG² „wie eine Patientin oder ein Patient fühlt, ihre oder seine Funktionen und Aktivitäten wahrnehmen kann oder ob sie oder er überlebt“.

Im Zentrum der Nutzenbewertung stehen demnach Endpunkte, welche von den Patientinnen und Patienten ganz unmittelbar wahrgenommen werden können. Patientenrelevante Endpunkte werden in der frühen Nutzenbewertung den vier Kategorien Mortalität, Morbidität, gesundheitsbezogene Lebensqualität und Nebenwirkungen zugeordnet. Davon abzugrenzen sind Endpunkte, die beispielsweise auf Bildgebung oder Laborbefunden beruhen und für Patientinnen und Patienten nicht spürbar sind. Immer wieder werden auch Änderungen der Darreichungsform als patientenrelevanter Vorteil per se vorgeschlagen. Ohne einen Nachweis, dass eine neue Darreichungsform patientenrelevante Endpunkte wie z.B. die gesundheitsbezogene Lebensqualität beeinflusst, kann ein solcher Vorteil weder gemäß SGB V noch aus der Definition der Patientenrelevanz des IQWiG abgeleitet werden. Nur in Ausnahmefällen, z.B. wenn an Stelle einer intrathekalen Gabe eine orale Gabe möglich ist,³ und damit eindeutig

von einer Verringerung von Komplikationen der Medikamentengabe ausgegangen werden kann, kann ein patientenrelevanter Vorteil abgeleitet werden.

Hoher Stellenwert Patienten-berichteter Endpunkte

Bei Endpunkten zur Morbidität und Lebensqualität sind Patienten-berichtete Endpunkte (engl. Patient-Reported Outcomes, PROs) von großer Bedeutung, ihre Rolle in klinischen Studien wurde durch ihre Bedeutung im AMNOG-Verfahren gestärkt. Bei PROs beurteilen die Patientinnen und Patienten selbst, wie sich der Einsatz des neuen Medikamentes beispielsweise auf ihre Symptome oder ihre gesundheitsbezogene Lebensqualität auswirkt. Ohne die Erhebung von PROs ergibt sich nur ein unvollständiges Bild zum Nutzen und Schaden von neuen Arzneimitteln. Schon bei der Entwicklung von PRO-Instrumenten ist die Patienteneinbindung unerlässlich. Nur dies stellt sicher,

dass alle aus Sicht der Patientin / des Patienten relevanten Aspekte verständlich und vollständig abgebildet sind.

PROs sind grundsätzlich gegenüber anthropometrischen Parametern, wie Körpergröße, als patientenrelevanter Endpunkt zu bevorzugen. Die Körpergröße (z-Score) wurde im Anwendungsgebiet Achondroplasie⁴ zwar als patientenrelevant eingestuft, es ist jedoch schwer abschätzbar, wie sich eine bestimmte Veränderung im Endpunkt Körpergröße (z-Score) letztendlich auf die Patientin bzw. den Patienten auswirkt, z. B. auf seine/ihre funktionellen Einschränkungen und Schmerzen. Für die Nutzenbewertung ließ sich deshalb der Zusatznutzen im Endpunkt Körpergröße (z-Score) nicht abschließend quantifizieren.

Auch bei Endpunkten zu Nebenwirkungen ist der Einsatz von PROs möglich, jedoch noch kein Standard. Zudem eignen sich nicht alle unerwünschten Ereignisse zur Erhe-



Dr. Sebastian Meller, seit 2021 wissenschaftlicher Mitarbeiter im Ressort Arzneimittelbewertung im IQWiG. Nach dem Studium der Biologie an der Universität zu Köln und anschließender Promotion im Fachbereich Biomedizin an der Universitätsklinik Bonn arbeitete er mehrere Jahre als Projektmanager am Zentrum für Klinische Studien Köln.



Dr. Daniela Preukschat, seit 2021 Bereichsleiterin im Ressort Arzneimittelbewertung (Bereich chronische Erkrankungen) im IQWiG. Nach dem Studium der Biologie sowie der Sportwissenschaften promovierte sie 2013 am Institut für Genetik der Universität zu Köln. Anschließend arbeitete sie zunächst im Bereich Evidenzbasierte Medizin des MDS (Medizinischer Dienst des Spitzenverbandes Bund der Krankenkassen e.V., Essen) und kam 2016 zum IQWiG.

bung mittels PROs, beispielsweise sind direkt beobachtbare/messbare Ereignisse (z.B. Netzhautriss) ungeeignet. Subjektiv empfundene Nebenwirkungen wie Übelkeit können hingegen sehr wohl mit PROs erhoben werden. Hierfür hat das National Cancer Institute (NCI) das PRO-Common Terminology Criteria for Adverse Events (PRO-CTCAE) System zur Bewertung von symptomatischer Toxizität bei Patientinnen und Patienten in onkologischen Studien entwickelt.⁵ Das System PRO-CTCAE wurde bereits in mehre-

ren Nutzenbewertungen dem IQWiG vorgelegt, erstmals in A20-87 und A23-86. Jedoch waren die entsprechenden Daten nicht verwertbar, da zum einen keine detaillierte Begründung für die Auswahl der aus dem PRO-CTCAE System verwendeten symptomatischen UEs vorlag und zudem die Ergebnisse lediglich deskriptiv dargestellt wurden ohne Berücksichtigung unterschiedlicher Beobachtungsdauern.^{6,7}

Surrogate, die vom IQWiG akzeptiert werden und Surrogate, für die Validierungen im Rahmen der frühen Nutzenbewertung vorgelegt wurden

Indikation	Endpunkte		Vom IQWiG akzeptiertes Surrogat?	Dossierbewertung (Beispiele)
	Surrogat	Patientenrelevanter Endpunkt		
HIV-Infektion	Virologisches Ansprechen	AIDS/Gesamtüberleben	JA	A12-04 Rilviripin
Diabetes Typ 1	HbA _{1c}	Mikrovaskuläre Folgekomplikationen	JA	A15-10 Insulin Degludec
Hepatitis C	SVR	Hepatozelluläres Karzinom	JA	A11-17 Boceprevir
Melanom	PFS	Gesamtüberleben	NEIN	A13-35 Debrafenib, A15-33 Pembrolizumab
Mammakarzinom	PFS	Gesamtüberleben	NEIN	A16-74 Palbociclib, in A22-66 kein Effekt im Gesamtüberleben (finale Analyse PALOMA-2)
Mammakarzinom - adjuvant	DFS	Gesamtüberleben	NEIN	A18-41, A21-11 Pertuzumab, in A22-103 Effekt im Gesamtüberleben im 3. Datenschnitt
Prostatakarzinom	PFS, MFS	Gesamtüberleben	NEIN	A18-80 Enzalutamid, A20-36 Apalutamid
interstitielle Lungenerkrankung	jährliche forcierte Vitalkapazität	Gesamtüberleben	NEIN	A20-71 Nintedanib

DFS: krankheitsfreies Überleben; **HbA_{1c}:** Hämoglobin A_{1c}; **MFS:** metastasenfreies Überleben; **PFS:** progressionsfreies Überleben; **SVR:** dauerhaftes virologisches Ansprechen

Quelle: IQWiG

Tabelle 1: Bisher waren die vorgelegten Validierungen oder Validierungsstudien überwiegend nicht geeignet, die Eignung der jeweiligen Surrogatendpunkte zu zeigen.

Nur wenige Surrogatvalidierungen vorgelegt

In frühen Nutzenbewertungen hat das IQWiG bisher drei Surrogatendpunkte als ausreichend valide Surrogate für jeweils unterschiedliche patientenrelevante Endpunkte akzeptiert: Virologisches Ansprechen als Surrogat für AIDS/Gesamtüberleben bei der HIV-Infektion, HbA_{1c}-Wert als Surrogat für mikrovaskuläre Folgekomplikationen bei Diabetes Typ 1 und das anhaltende virologische Ansprechen (englisch Sustained Virological Response [SVR]) als Surrogat für die Vermeidung des Auftretens eines hepatozellulären Karzinoms bei Hepatitis C-Infektion.⁸⁻¹⁰ Insgesamt wurden dem IQWiG jedoch nur wenige Validierungen von Surrogaten für die frühe Nutzenbewertung vorgelegt, diese waren vornehmlich im onkologischen Bereich (siehe Tabelle 1).

Hier ist insbesondere das progressionsfreie Überleben (englisch progression free survival [PFS]) zu nennen, für das am häufigsten Surrogatvalidierungen vorgelegt wurden (für die Indikationen Melanom, Mammakarzinom und Prostatakarzinom¹¹⁻¹⁵). Darüber hinaus wurden für das krankheitsfreie Überleben (englisch disease-free survival [DFS]) bei Mammakarzinom-adjuvant, das Metastasenfreie Überleben bei Prostatakarzinom und die jährliche forcierte Vitalkapazität bei interstitieller Lungenerkrankung jeweils Surrogatvalidierungen vorgelegt.¹⁵⁻¹⁸

Jedoch waren die vorgelegten Validierungen / Validierungsstudien nicht geeignet, die Eignung der jeweiligen Surrogatendpunkte für das Gesamtüberleben zu zeigen. Unabhängig von der Frage der Eignung als Surrogat für das Gesamtüberleben wurden einige Endpunkte bzw. Operationalisierungen, welche das Fortschreiten einer Krebserkrankung abbilden, als patientenrelevant eingestuft: so wurde die symptomatische Progression beim Prostatakarzinom, die über für den Patienten spürbare Symptome operationalisiert wurde, in zwei Nutzenbewer-

tungen akzeptiert.^{15,19} Auch Rezidive wurden grundsätzlich als patientenrelevanter Endpunkt in verschiedenen Nutzenbewertungen akzeptiert.^{16,17}

Die Erhebung des Endpunkts Rezidive beruht zwar auf einer Bildgebung, jedoch wird es als Scheitern des kurativen Ansatzes gewertet, wenn der Tumor während oder nach einer (adjuvanten) Therapie mit kurativer Absicht wieder feststellbar ist. Dies stellt für die Patientinnen und Patienten ggf. den Übergang in ein nicht mehr heilbares Krankheitsstadium dar und ist damit unmittelbar patientenrelevant. Zudem ist das Scheitern des kurativen Therapieansatzes auch bei Patientinnen und Patienten relevant, die zu Studienbeginn nicht tumorfrei sind und wurde vom IQWiG als eigenständiger Endpunkt bei hämato-onkologischen Patientinnen und Patienten mit diffusem großzelligem B-Zell-Lymphom (DLBCL)²⁰⁻²², aber auch bei Patientinnen und Patienten mit soliden Tumoren^{23,24} herangezogen.

Methodische Anforderungen für valide Surrogate

Surrogatvalidierungen sind grundsätzlich auf Basis von randomisiert kontrollierten Studien (RCTs) sowie in weiteren speziellen Situationen (wie sie bei der SVR gegeben waren) möglich. Erstere bedürfen in der Regel einer Metaanalyse von mehreren RCTs, in denen sowohl Effekte auf den Surrogatendpunkt als auch Effekte auf den interessierenden patientenrelevanten Endpunkt untersucht wurden. Anerkannte Validierungsverfahren sind korrelationsbasierte Verfahren wie die Betrachtung der Korrelation zwischen den Effekten auf Studienebene sowie der Surrogate Threshold Effect (STE).

Neben der verwendeten Methodik zur Surrogatvalidierung ist es dabei von elementarer Bedeutung, dass der zugrundeliegende Studienpool vollständig und geeignet ist. Die in den Nutzenbewertungen zu Pertuzumab (Mamma-

karzinom)^{16, 17} und Nintedanib (interstitielle Lungenerkrankung)¹⁸ vorgelegten Validierungsstudien zeigen, dass es prinzipiell möglich ist, eine Surrogatvalidierung gemäß der Anforderung des IQWiG durchzuführen. In dem Projekt zu Pertuzumab sollte das DFS und zu Nintedanib die jährliche forcierte Vitalkapazität je als geeignetes Surrogat für das Gesamtüberleben gezeigt werden. Bei Pertuzumab war der zugrunde gelegte Studienpool der Validierungsstudie nicht geeignet, da hier Studien ausgeschlossen wurden, die in dem in dieser Nutzenbewertung vorliegenden Anwendungsgebiet relevant gewesen wären. Bei Nintedanib war die methodische Durchführung der Validierung fehlerhaft, was zu einer Unterschätzung des STE führte. Nach selbst berechneter Korrektur zeigte sich, dass der Effekt auf das Surrogat nicht groß genug war, um einen Effekt auf das Gesamtüberleben abzuleiten.

Darüber hinaus gibt es spezielle Situationen, in denen ebenfalls eine Validität anerkannt werden kann.²⁵ Dafür ist es erforderlich, dass der Zusammenhang zwischen dem patientenrelevanten Endpunkt und dem Surrogatendpunkt biologisch/medizinisch eindeutig plausibel ist und weitere Kriterien erfüllt sind. Ein Beispiel ist die SVR bei Patientinnen und Patienten mit chronischer Hepatitis-C-Infektion, bei dem der Eintritt des Surrogatendpunkts zu einem deutlich reduzierten Risiko für das Auftreten von hepatozellulären Karzinomen führte. Zudem erreichte das Risiko bezüglich des eigentlichen Endpunkts ein minimales Niveau, nämlich das einer nicht erkrankten Population.¹⁰ Als Datengrundlage müssen in den speziellen Situationen Kohortenstudien vorliegen, die sich auf Personen unter Therapie beziehen und deren Nachbeobachtungszeit ausreichend lang ist, um das Risiko für das Eintreten des eigentlichen Endpunkts adäquat zu erfassen.

Nutzenbewertung und Therapieentscheidungen erfordern hinreichend sichere Daten

Problematisch ist aus Sicht des IQWiG, dass für die Zulassung neuer Arzneimittel von den Zulassungsbehörden häufig Surrogatendpunkte akzeptiert werden, die nicht ausreichend validiert sind und dass die Verwendung solcher Surrogatendpunkte über die letzten Jahrzehnte sowohl bei der Food and Drug Administration (FDA) als auch der Europäische Arzneimittel-Agentur (EMA) zugekommen hat.²⁶⁻²⁸ Für die (ggf. beschleunigte/bedingte) Zulassung wird oft die Position eingenommen, dass dies vertretbar sei und man die größere Aussageunsicherheit akzeptieren möchte, um ein neues Medikament schnell verfügbar zu machen.

Die Akzeptanz einer großen Aussageunsicherheit zum Zeitpunkt der Zulassung darf jedoch nicht dazu führen, dass niemals wirklich zuverlässige Daten generiert werden. Die Fragestellung der Nutzenbewertung und evidenzbasierte Therapieentscheidungen erfordern eine hinreichende Aussagesicherheit. Ziel der frühen Nutzenbewertung ist, diejenigen neuen Arzneimittel mit dem Urteil „Zusatznutzen belegt“ herauszufiltern, die hinreichend sicher einen Mehrwert für die Patientinnen und Patienten aufweisen und nicht nur „vielleicht“.

Ebenso erfordert eine evidenzbasierte medizinische Versorgung aussagekräftige Daten für gut begründete Therapieentscheidungen. Dafür braucht es aussagekräftige Endpunkte und in gewissen Fällen mehr bzw. bessere Surrogatvalidierungen. Des Weiteren trifft der oft geäußerte Vorwurf, dass in Leitlinien und Disease-Management-Programmen (DMP) beschriebene Therapieparameter von der Nutzenbewertung ignoriert werden, ebenfalls nicht den Kern des Problems: So legitimiert der in Leitlinie oder DMP empfohlene Einsatz eines Labor-Parameters zur Therapiesteuerung beim individuellen Patienten keines-

falls zwangsläufig seine Eignung als patientenrelevanter Endpunkt in einer klinischen Studie. Entscheidend für die Nutzenbewertung kann nur sein, ob man hinreichend sicher ist, dass sich ein Effekt im Surrogatendpunkt auch in einem Effekt im patientenrelevanten Endpunkt widerspiegeln wird.

Datenlage bei chronischen Erkrankungen ist unverändert schlecht

Insbesondere im Bereich der chronischen Erkrankungen zeigt sich unverändert oft ein erheblicher Mangel an geeigneten Daten für die frühe Nutzenbewertung. Von 2021 bis Anfang August 2024 wurden vom IQWiG insgesamt 169 Nutzenbewertungen im Bereich chronischer Erkrankungen durchgeführt. Bei 68 Prozent (115) der Projekte lagen keine geeigneten Daten vor (siehe Abbildung 1).

Problem sind hier zunächst nicht die Endpunkte, sondern vielmehr das Studiendesign. Auch wenn es sich in der Mehrzahl der Studien um RCTs handelte, sind die Daten für die Nutzenbewertung häufig nicht geeignet, da diese RCTs auf einem Vergleich gegenüber einem Placebo oder einer nicht adäquaten Therapie im Vergleichsarm basieren, anstatt auf einem direkten Vergleich mit der aktuellen Standardtherapie. In anderen Fällen findet sich gar kein Vergleich (1-armige Studien). Daten solcher Studien sind für die Nutzenbewertung in der Regel nicht geeignet. Bereits 2019 wurde bemängelt, dass für die Nutzenbewertung häufig keine geeigneten Daten verfügbar sind.²⁹ Leider ist – hier mit Blick auf die chronischen Erkrankungen – diesbezüglich keine positive Entwicklung zu sehen. Die Ursachen für diese Stagnation und mögliche Ansatzpunkte zur Veränderung (gezielter Einsatz positiver oder negativer [An-]Reize) sollten diskutiert werden.

In den 54 (32 Prozent) Nutzenbewertungen von 2021 bis

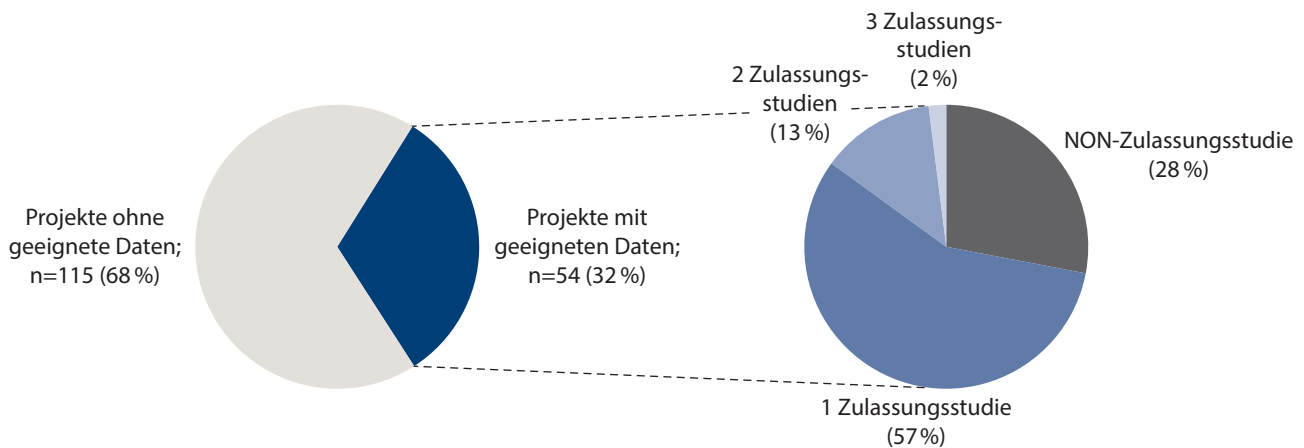
Anfang August 2024 in denen geeignete Daten für die Nutzenbewertung vorlagen, wurde in über 70 Prozent der Bewertungen auf Zulassungsstudien zurückgegriffen. In der Mehrzahl dieser Fälle lag jeweils nur eine Zulassungsstudie für die Nutzenbewertung vor. Für 15 (28 Prozent) der Nutzenbewertungen wurden Nicht-Zulassungsstudien (Abbildung 1) herangezogen. Insgesamt war in nur etwa einem Viertel der eingeschlossenen Studien der primäre Studienendpunkt uneingeschränkt patientenrelevant. Bei über 40 Prozent der eingeschlossenen Studien war der primäre Endpunkt nicht patientenrelevant, weil dieser, wie oben bereits für Zulassungsstudien beschrieben, beispielsweise auf einem (nicht ausreichend validierten) Surrogatendpunkt basierte (z.B. Änderung des HbA_{1c}-Werts bei Diabetes Typ 2-, oder der eGFR Slope bei Morbus Fabry Patientinnen und Patienten, siehe Abbildung 2).

Zur eGFR ist anzumerken, dass Änderungen der eGFR nur dann patientenrelevant sind, wenn hinreichend sicher ist, dass die eGFR in einen für die Patienten / den Patienten spürbaren Bereich abfällt. Dies war in einem der beiden Addenda zu Finerenon³⁰ erstmals der Fall. Der kombinierte Endpunkt der vorgelegten Studien zur renalen Morbidität umfasste die Einzelkomponenten Nierenversagen (definiert als bestätigt anhaltende Abnahme der eGFR auf < 15 ml/min/1,73 m² oder Nierenerkrankung im Endstadium), eGFR-Abnahme ≥ 57 Prozent und renal bedingter Tod. Aufgrund der mittleren eGFR-Ausgangswerte (ca. 43 ml/min/1,73 m²) der Patientinnen und Patienten konnte in der vorliegenden Datensituation von einer hinreichenden Patientenrelevanz der Komponente eGFR-Abnahme ≥ 57 Prozent ausgegangen und der kombinierte Endpunkt herangezogen werden.

In 30 Prozent der eingeschlossenen Studien konnte der primäre Endpunkt nur mit Einschränkungen herangezogen werden (z.B.: Veränderung der durchschnittlichen

Chronische Erkrankungen – oft liegen keine geeigneten Daten für die Nutzenbewertung vor; wenn Daten vorliegen, sind dies oft Zulassungsstudien

Projekte bei chronischen Erkrankungen ab 2021 bis August 2024 (n = 169)



Quelle: IQWiG

Abbildung 1: Von den 169 Nutzenbewertungen mit Bezug auf chronische Erkrankungen lagen im Zeitraum von Anfang 2021 bis Anfang August 2024 bei 68 Prozent (115) der Projekte keine geeigneten Daten vor. Problematisch waren dabei nicht primär die Endpunkte, sondern das Studiendesign.

monatlichen Migränetage gegenüber Baseline vs. Migränetage/Monat Reduktion um 50 Prozent oder nur einzelne Komponenten eines kombinierten Endpunkts wurden herangezogen, siehe Abbildung 2). Die Operationalisierung eines Endpunkts kann somit auch entscheidend dafür sein, ob ein Endpunkt als patientenrelevant eingeschätzt wird oder nicht.

Es zeigt sich also, dass die Tatsache, dass ein Endpunkt in (Zulassungs)-Studien untersucht wurde, nicht zwangsläufig bedeutet, dass es sich um einen patientenrelevanten Endpunkt handelt. Statt „Studienendpunkt→relevant“ sollte, wie bereits 2016 von Thomas Kaiser vorgeschlagen,³¹ besser gelten „relevanter Endpunkt→Studienendpunkt“. Grundsätzlich sollten HTA-Entscheidungen die Verwendung von Studienendpunkten, die eine hohe

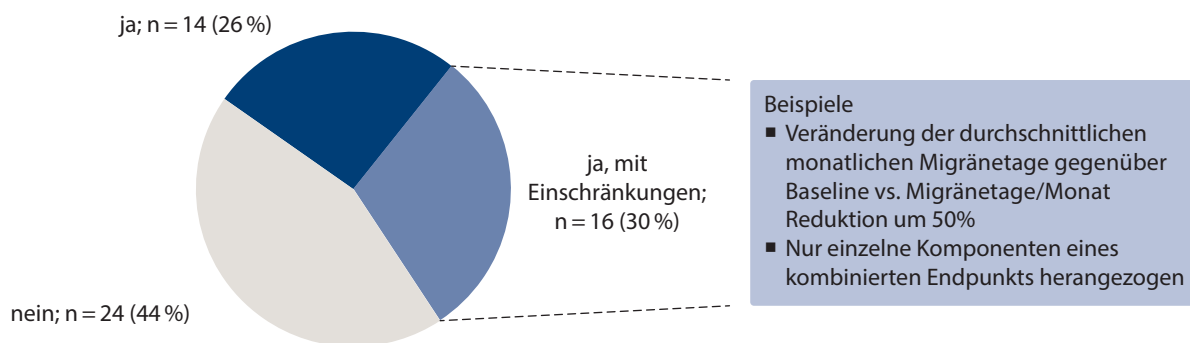
Unsicherheit mit sich bringen, nicht incentivieren und stattdessen vielmehr die Untersuchung von aussagekräftigen, patientenrelevanten Endpunkten belohnen.²⁹

Die Frage der Patientenrelevanz des primären Endpunkts bzw. dessen Operationalisierung ist jedoch nur selten unmittelbar fazitrelevant hinsichtlich der Frage „Besteht ein Zusatznutzen: ja oder nein?“. Lediglich in sechs Projekten (11 Prozent) von 2021 bis Anfang August 2024 wurde kein Zusatznutzen abgeleitet, obwohl sich auf Basis der vorgelegten Studien ein Vorteil für die Intervention beim primären (aber nicht patientenrelevanten) Endpunkt zeigte (z.B. Änderung des HbA1c-Werts bei Typ 2-Diabetes). Meist gibt es aber weitere relevante Endpunkte die einen Zusatznutzen zeigten (dies war in 24 Projekten [44 Prozent] der Fall). Ein Beispiel ist die Bewertung von

Nur etwa ein Viertel der primären Studienendpunkte sind uneingeschränkt patientenrelevant

Projekte bei chronischen Erkrankungen ab 2021 bis August 2024 (n = 169), davon n = 54 mit Daten

Wurde der primäre Endpunkt der in die Dossierbewertung eingeschlossenen Studien als patientenrelevant eingeschätzt?



Quelle: IQWiG

Abbildung 2: Die Operationalisierung eines Endpunkts kann auch entscheidend dafür sein, ob ein Endpunkt als patientenrelevant eingeschätzt wird oder nicht. Dass ein Endpunkt in (Zulassungs-)Studien untersucht wurde, bedeutet nicht zwangsläufig, dass es sich um einen patientenrelevanten Endpunkt handelt.

Risankizumab, in der zwar der primäre Endpunkt (endoskopische Remission) nicht akzeptiert wurde, sich aber positive Effekte im Inflammatory Bowel Disease Questionnaire (IBDQ; Gesamtscore und Subscore-Darmsymptome) und Short Form-36 (SF-36) zeigten.³²

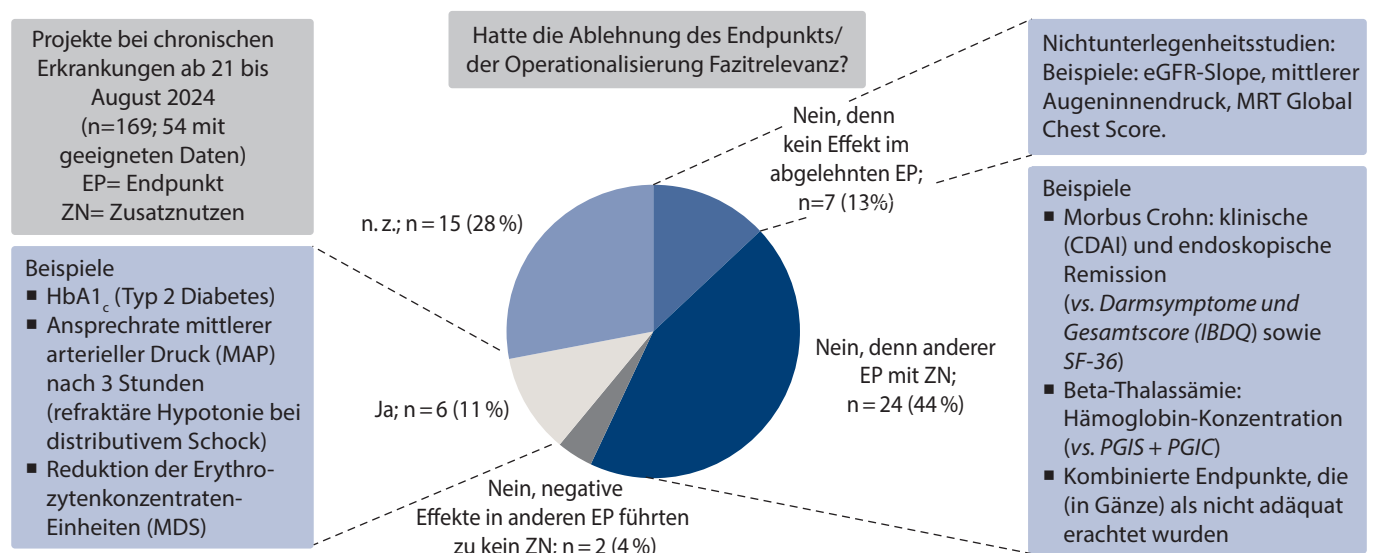
In anderen Bewertungen zeigten sich neben positiven Effekten auch negative Effekte, die in der Gesamtschau zum Urteil „kein Zusatznutzen“ führten (zwei Projekte [4 Prozent]). Bei sieben Projekten (13 Prozent) zeigte der primäre (nicht patientenrelevante) Endpunkt keinen Effekt und die Ablehnung hatte daher keinen Einfluss auf das Fazit der Nutzenbewertung (siehe auch Abbildung 3). Bei diesen Studien handelte es sich um Nichtunterlegenheitsstudien. Problematisch ist jedoch, wenn in diesen Fällen de facto keine aussagekräftigen Endpunkte zur Mortalität,

Morbidität oder Lebensqualität, d. h. zu Nutzenendpunkten, vorliegen. Dies war beispielsweise bei Vadadustat der Fall³³, aber auch in den neuesten Diabetesbewertungen (Typ2) zeigt sich, dass die Studien weiterhin nicht auf die Therapieziele ausgerichtet sind.³⁴

Verbesserungen bei PRO-Analysen notwendig

Die Datenqualität von PROs für die Nutzenbewertung ist oft nicht ausreichend, so dass nur ein Teil der vorgelegten PRO-Analysen letztlich für die Nutzenbewertung geeignet ist.³⁵ Ein weiteres häufiges Problem ist, dass die PROs zu kurz erhoben werden. Der Grund ist vermutlich, dass die für die Nutzenbewertung vorgelegten Studien stark auf die Zulassung zugeschnitten sind. So werden die PROs z. B. nur bis zur Progression der Erkrankung erfasst. Aber für die

In nur 11 Prozent der herangezogenen Studien war die Ablehnung des primären Endpunkts fazitrelevant (bezüglich der Frage Zusatznutzen ja / nein?)



CDAI: Crohn's Disease Activity Index; eGFR: geschätzten glomerulären Filtrationsrate; IBDQ: Inflammatory Bowel Disease Questionnaire; MDS: myelodysplastische Syndrome; PGIC/PGIS: Patient Global Impression of Change bzw. Severity; SF-36: Short Form-36

Quelle: IQWiG

Abbildung 3: In der überwiegenden Zahl der Bewertungen ist die Ablehnung des primären Studienendpunktes aufgrund seiner nicht-gegebenen Patientenrelevanz nicht fazitrelevant für die Frage nach dem Zusatznutzen (ja oder nein?)

Betroffenen sind die Symptome und Lebensqualität selbstverständlich auch nach einer Verschlechterung der Erkrankung relevant. Ob es der Patientin oder dem Patienten mit der zu bewertenden Intervention langfristig besser oder schlechter geht als mit der Vergleichsbehandlung lässt sich nur beurteilen, wenn die entsprechenden PROs über den Progress hinaus erhoben werden.³⁶

Von besonderer Bedeutung ist zudem, dass im Erhebungszeitraum die PROs vollständig erhoben werden. So führen zu geringe Rücklaufquoten der entsprechenden Fragebögen oft dazu, dass die Daten für die Nutzenbewer-

tung nicht herangezogen werden können. Auch in einer Diskussionsrunde verschiedener Interessenvertreter wurde seitens der Patientenvertretungen geäußert, dass Betroffene durchaus willens sind, Zeit in die PRO-Erfassung zu investieren, wenn man ihnen konkret erläutert, wozu ihre Angaben benötigt werden. Das zeigt: Eine längere Erhebung solcher Zielgrößen ist grundsätzlich machbar – und läuft den Interessen der Betroffenen keineswegs zuwider.³⁷

Die Analyse bzw. Operationalisierung von PRO-Endpunkte muss im Einzelfall sorgfältig diskutiert werden, da

zahlreiche Aspekte die Aussagekraft der Ergebnisse beeinflussen können. So ist beispielsweise je nach Indikation und Therapieziel einmal die (dauerhafte) Verbesserung, ein anderes Mal die Verschlechterung primär relevant. Betrachtet man stetige Auswertungen oder Responderanalysen zu einem bestimmten Auswertungszeitpunkt? Welcher Zeitpunkt wäre am besten geeignet für eine Responderanalyse?³⁸ Welches Responsekriterium ist geeignet, eine spürbare Veränderung anzuzeigen? Für die Frage nach dem Responsekriterium hat sich die vom IQWiG vorgeschlagene patientenindividuelle Änderung von 15 Prozent der Spannweite des Erhebungsinstruments als praxistauglich erwiesen.³⁹ Ziel dieses vom IQWiG entwickelten Ansatzes war es auch, Klarheit für die Hersteller zu schaffen und ergebnisgesteuerte Berichterstattung zu verhindern. Seit ihrer Einführung wird diese Responseschwelle regelmäßig von Herstellern in Studien bzw. Dossiers angewendet.

Fazit

- IQWiG-Grundsätze sind unverändert: Patientenrelevanz ist das Paradigma der Nutzenbewertung.
- Stand zu Surrogaten: wenige vorgelegte Validierungen, hauptsächlich im onkologischen Bereich. Zahlreiche Veröffentlichung der letzten Jahre zeigen: Zulassungsbehörden akzeptieren Surrogate, deren Validität nicht hinreichend gezeigt ist – mehr Validierungen (da, wo nötig) wünschenswert.
- Statt „Studienendpunkt→relevant“ sollte gelten „relevanter Endpunkt→ Studienendpunkt“
- Verbesserungen bei PRO-Analysen notwendig:
 - grundsätzliche methodische Qualität (Vollständigkeit und Qualität der Daten) & Dauer der PRO-Erhebung oft verbesserungswürdig (z. B. über Progress hinaus).
 - Sinnhaftigkeit von Analysen muss im Einzelfall diskutiert werden (z. B. Symptomlast, [dauerhafte] Verbesserung/Verschlechterung; Umgang mit „intercurrent events“; betrachteter Zeitpunkt).

- Insgesamt sind (noch) mehr hochwertige Daten zu PROs notwendig.

Literatur

- ¹ Bundesministerium für Gesundheit (2023) Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsvereinbarungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung - AM-NutzenV). <https://go.sn.pub/f5cp5w>. Accessed 09.10.2023.
- ² Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2023) Allgemeine Methoden; Version 7.0. <https://go.sn.pub/mkwigq>. Accessed 06.10.2023.
- ³ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2021) Risdiplam (spinale Muskelatrophie) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/doprfg>. Accessed 11.07.2023.
- ⁴ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2024) Vosoritid (Achondroplasie); Addendum zum Projekt A23-92 (Dossierbewertung). <https://doi.org/10.60584/A24-08>. Accessed 23.02.2024.
- ⁵ Institute NC Patient-Reported Outcomes version of the Common Terminology Criteria for Adverse Events (PRO-CTCAE). <https://go.sn.pub/n7lwyc>. Accessed 07.10.2024.
- ⁶ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2020) Durvalumab (kleinzelliges Lungenkarzinom) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/9oyg37>. Accessed 11.07.2023.
- ⁷ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2023) Sacituzumab govitecan (Mammakarzinom); Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://doi.org/10.60584/A23-86>. Accessed 17.11.2023.
- ⁸ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2012) Rilpivirin – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/zphzud>. Accessed 11.07.2023.
- ⁹ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2015) Insulin degludec (neues Anwendungsgebiet) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/3lai7e>. Accessed 11.07.2023.
- ¹⁰ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2011) Boceprevir – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/0605wp>. Accessed 11.07.2023.
- ¹¹ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2013) Dabrafenib – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/qnv9l4>. Accessed 11.07.2023.
- ¹² Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2015) Pembrolizumab – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/0lfwif>. Accessed 11.07.2023.

- ¹³ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2017) Palbociclib (Mammakarzinom) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/0r0kcc..> Accessed 11.07.2023.
- ¹⁴ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2017) Ribociclib (Mammakarzinom) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/kqadrk>. Accessed 11.07.2023.
- ¹⁵ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2020) Apalutamid (Prostatakarzinom) – Nutzenbewertung gemäß § 35a SGB V (Ablauf Befristung); Dossierbewertung. <https://go.sn.pub/7dbyze>. Accessed 11.07.2023.
- ¹⁶ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2018) Pertuzumab (Mammakarzinom) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/nnvllq>. Accessed 11.07.2023.
- ¹⁷ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2021) Pertuzumab/Trastuzumab (Mammakarzinom, adjuvant) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/x6kivq>. Accessed 11.07.2023.
- ¹⁸ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2020) Nintedanib (andere chronische progredient fibrosierende interstitielle Lungenerkrankungen) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/4bj058>. Accessed 11.07.2023.
- ¹⁹ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2024) Niraparib/Abirateronacetat (Prostatakarzinom); Addendum zum Projekt A23-107 (Dossierbewertung). <https://go.sn.pub/g13v0o>. Accessed 06.05.2024.
- ²⁰ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2023) Lisocabtagen maraleucel (DLBCL, HGBL, PMBCL und FL3B, Zweitlinie); Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/uz4974>. Accessed 16.09.2023.
- ²¹ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2023) Axicabtagen-Ciloleucel (DLBCL und HGBL, Zweitlinie); Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/w6w8p2>. Accessed 10.10.2023.
- ²² Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2024) Polatuzumab Vedotin (Kombination mit Rituximab, Cyclophosphamid, Doxorubicin und Prednison; bisher unbehandeltes DLBCL); Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/n4pkoh>. Accessed 02.04.2024.
- ²³ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2024) Nivolumab (NSCLC, neoadjuvant); Addendum zum Projekt A23-74 (Dossierbewertung). <https://go.sn.pub/81yl50>. Accessed 01.02.2024.
- ²⁴ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2022) Pembrolizumab (Mammakarzinom) – Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung. <https://go.sn.pub/6f2qkt>. Accessed 11.07.2023.
- ²⁵ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2025) [A24-61] Wissenschaftliche Ausarbeitung zu klinischen Studien im Therapiegebiet Wundbehandlung. <https://go.sn.pub/t18565>. Accessed 07.01.2025.
- ²⁶ Schuster Bruce C, Brhlikova P, Heath J, McGettigan P (2019) The use of validated and nonvalidated surrogate endpoints in two European Medicines Agency expedited approval pathways: A cross-sectional study of products authorised 2011–2018. *PLoS Med* 16(9):e1002873. doi:10.1371/journal.pmed.1002873.
- ²⁷ Gyawali B, Hey SP, Kesselheim AS (2020) Evaluating the evidence behind the surrogate measures included in the FDA's table of surrogate endpoints as supporting approval of cancer drugs. *EClinicalMedicine* 21:100332. doi:10.1016/j.eclinm.2020.100332.
- ²⁸ Wallach JD, Yoon S, Doernberg H (2024) Associations Between Surrogate Markers and Clinical Outcomes for Nononcologic Chronic Disease Treatments. *JAMA* 331(19):1646-1654. doi:10.1001/jama.2024.4175.
- ²⁹ Wieseler B, McGauran N, Kaiser T (2019) New drugs: where did we go wrong and what can we do better? *BMJ* 366:l4340. doi:10.1136/bmj.l4340.
- ³⁰ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2023) Finerenon (Niereninsuffizienz, Stadium 3 und 4); Addendum zum Projekt A23-15 (Dossierbewertung). <https://go.sn.pub/qmqozn>. Accessed 23.08.2023.
- ³¹ Thomas Kaiser; Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2016) Klinisch relevante Studienendpunkte und Surrogate bei der frühen Nutzenbewertung. Interdisziplinäre Plattform zur Nutzenbewertung Heft 2 Januar 2016
- ³² Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2023) Risankizumab (Morbus Crohn); Addendum zum Projekt A22-133 (Dossierbewertung). <https://go.sn.pub/42elj9>. Accessed 11.07.2023.
- ³³ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2024) Vadadustat bei symptomatischer Anämie infolge chronischer Nierenerkrankung: Chance vertan. <https://go.sn.pub/a2uk9c>.
- ³⁴ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2024) Insulin icodec (Diabetes mellitus Typ 2) – Nutzenbewertung gemäß § 35a SGB V. <https://go.sn.pub/byfekc>.
- ³⁵ Kramer L, Moos M, Thaa B (2024) Health-related quality of life (HRQoL) in German early benefit assessment: The importance of disease-specific instruments. *Z Evid Fortbild Qual Gesundhwes* 186:1-9. doi:10.1016/j.zefq.2024.02.003.
- ³⁶ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2022) Kritik an Studienplanung: Patientenberichtete Endpunkte oft viel zu kurz erhoben. <https://go.sn.pub/em0o9l>.
- ³⁷ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2023) Langfristige Erhebung patientenberichteter Endpunkte in onkologischen Studien: wichtig und machbar. <https://go.sn.pub/bdb35h>.
- ³⁸ Pe M, Alanya A, Falk RS (2023) Setting International Standards in Analyzing Patient-Reported Outcomes and Quality of Life Endpoints in Cancer Clinical Trials-Innovative Medicines Initiative (SISAQOL-IMI): stakeholder views, objectives, and procedures. *Lancet Oncol* 24(6):e270-e283. doi:10.1016/S1470-2045(23)00157-2.
- ³⁹ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2021) Klinische Relevanz patientenberichteter Endpunkte: Neuer Schwellenwert erweist sich als praxistauglich. <https://go.sn.pub/jmstw4>.

Innovation in Studienendpunkten – Studienendpunkte für Innovationen

Dr. Jutta Wendel-Schrief | MSD Sharp & Dohme GmbH

Klinische Studien müssen weltweit sowohl den Anforderungen der Zulassungsbehörden, als auch den Anforderungen im Rahmen des Marktzugangs und der Erstattung gerecht werden. Betrachtet man die frühe Nutzenbewertung in AMNOG, weichen die Anforderungen an pivotale Studien oft von den Anforderungen der EMA ab. Ein entscheidender Aspekt bei der Planung klinischer Studien stellen die Endpunkte dar. EU-HTA verfolgt einen patientenzentrierten HTA-Ansatz, während Deutschland einen Sonderweg mit der Festlegung von patientenrelevanten Endpunkten beschreitet. Die Operationalisierung von Morbiditätsendpunkten erfordert eine klare Definition der klinischen Relevanz und die Berücksichtigung der Heterogenität der Patientenpopulation. Ein Konsens zwischen medizinischen Fachgesellschaften, klinischer Forschung und Patientenvertretern ist essenziell, um geeignete Endpunkte zu formulieren und relevante Daten zur Lebensqualität zu erfassen. MSD beteiligt sich aktiv an der Weiterentwicklung sowohl klinisch als auch HTA-relevanter Endpunkte. Mit EUHTA ergibt sich die Chance einer Harmonisierung und Weiterentwicklung von AMNOG hin zur Patientenzentrierung.

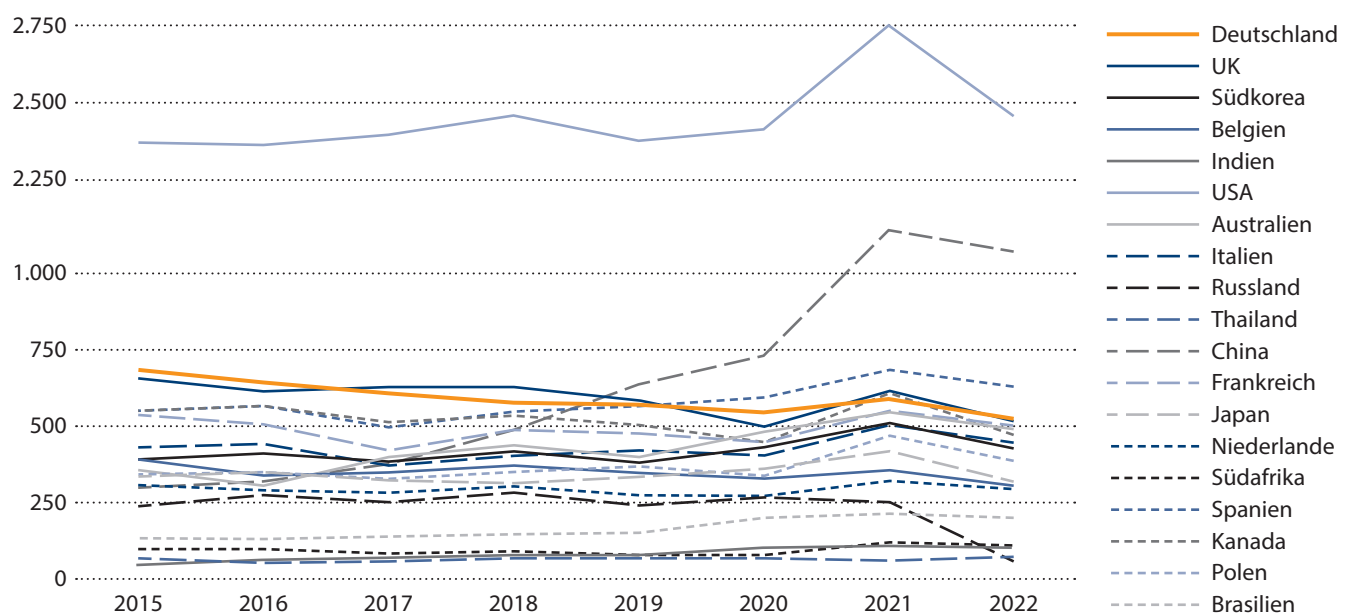
Pharmazeutische Unternehmen als Quelle der Innovation in Deutschland und der Welt
Pharmazeutische Unternehmen spielen seit Jahrzehnten eine zentrale Rolle bei der Entwicklung innovativer Arzneimittel, die die Gesundheitsversorgung weltweit voranbringen. Gemäß einer Studie der European Federation of Pharmaceutical Industries and Associations (EFPIA) investierte die pharmazeutische Industrie im Jahr 2023 fast 50 Milliarden Euro in Forschung und Entwicklung in Europa, wobei etwa die Hälfte (48,4 Prozent) dieser Summen in klinische Studien floss.

In Deutschland allein investierte die Pharmaindustrie im Jahr 2020 rund 7 Milliarden Euro in Forschung und Entwicklung und rangiert mit 542 klinischen Studien weltweit auf Rang 6 bei der Durchführung von Industrie-initiierten klinischen Studien. Laut einer Analyse von Charles Rivers Associates ist der Privatsektor für knapp zwei Drittel der Investitionen in Forschung und Entwicklung verantwortlich.

Eine große Herausforderung für pharmazeutische Unternehmen (pU): Verschiedene Anforderungen ausbalancieren

Das Design klinischer Studien stellt Sponsoren immer wieder vor erheblichen Herausforderungen. Klinische Studien müssen nicht nur den Anforderungen der medizinischen Wissenschaft, sondern auch den Anforderungen der Zulassungs- und Erstattungsbehörden gerecht werden. Für ein global tätiges Unternehmen wie MSD Sharp & Dohme (MSD), das Medikamente in über 140 Länder vertreibt, bedeutet dies, dass diese Anforderungen berücksichtigt werden müssen, um Zulassung und Marktzugang zu sichern. Hier weichen trotz International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use (ICH) und europäische HTA-Verordnung/EUnetHTA wichtige Standards immer noch voneinander ab.

Zahl klinischer Studien von Pharma-Unternehmen im Ländervergleich



Quelle: vfa auf Basis von clinicaltrials.org

siehe <https://www.vfa.de/de/arzneimittel-forschung/so-funktioniert-pharmaforschung/amf-standortfaktoren.html>

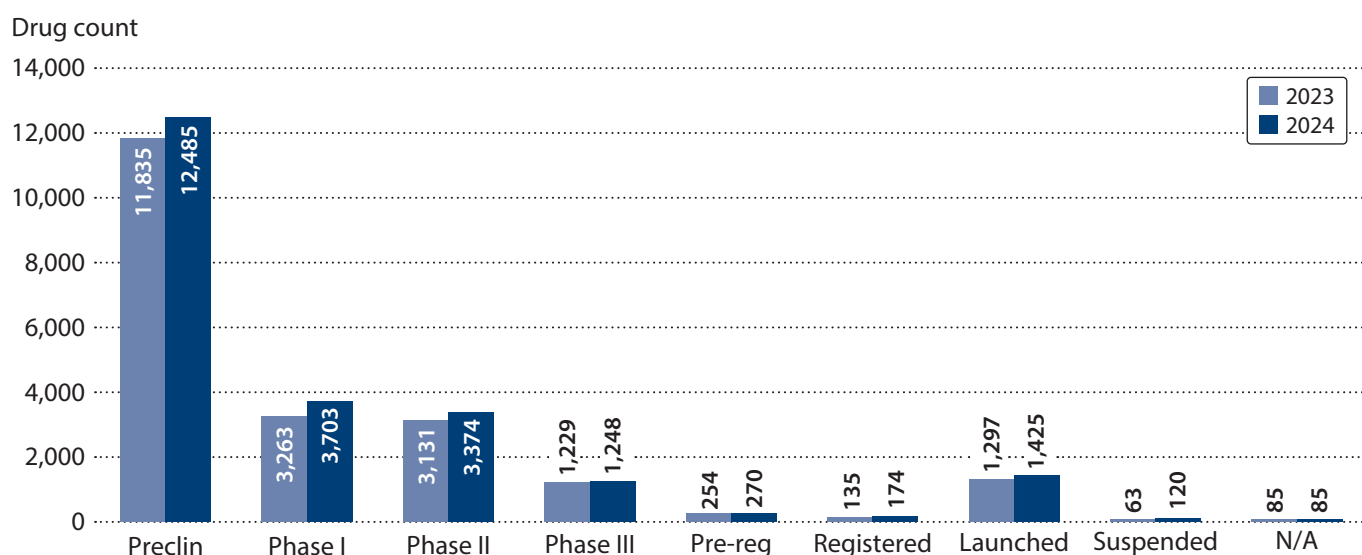
Abbildung 1: Im Jahr 2020 lag Deutschland bei der Zahl Industrie-initiierten Studien mit 542 klinischen Studien im Ländervergleich weltweit auf Rang 6.



Dr. Jutta Wendel-Schrief ist seit Januar 2015 Leiterin des Geschäftsbereichs Market Access bei MSD Sharp & Dohme GmbH. Jutta Wendel-Schrief studierte Biologie und promovierte zum Dr. phil. nat. im Bereich der Biochemie an der Johann Wolfgang von Goethe-Universität in Frankfurt am Main. Nach mehrjähriger Forschungstätigkeit am Max-

Planck-Institut für Biophysik zum Thema Transportproteine arbeitete sie seit 1990 in verschiedenen Firmen im pharmazeutischen Umfeld, in unterschiedlichen Bereichen und Indikationen, hauptsächlich in Marketing und Vertrieb. Von 2010 bis 2015 leitete sie bei der MSD Sharp & Dohme GmbH die Business Unit Klinik- und Spezialprodukte.

Pipeline der in der Entwicklung befindlichen Medikamente, Vergleich 2024 zu 2023



N/A = not applicable and is applied to companion diagnostics prelaunch

Quelle: Pipeline by development phase, Pharmaprojects January 2024

Abbildung 2: Im Durchschnitt schaffen es nur ein bis zwei von je 10.000 aller in der Entwicklung befindlichen Medikamente bis zur Marktreife.

Bei MSD sind weltweit ungefähr ein Drittel der Mitarbeiter:innen im Bereich Forschung und Entwicklung tätig und arbeiten zur Zeit unter anderem an der Planung und Durchführung von über 2.000 klinischen Studien in 47 Ländern und vier Regionen, darunter 130 klinische Studien in Deutschland, mit Rekrutierung für 20 weitere Studien. Eine Vielzahl interner und externer Experten, darunter auch Kliniker, Operations-Spezialisten, Forschungspartner und Patient:innen, sind an der Erstellung der Studienprotokolle beteiligt. Ziel dabei ist es, durchführbare Studienprotokolle zu erstellen, die den weltweiten Anforderungen entsprechen. Auch deutsche Anforderungen haben einen hohen Stellenwert.

Trotz aller Anstrengungen schaffen es im Durchschnitt nur ein bis zwei von je 10.000 aller in der Entwicklung befindlichen Medikamente bis zur Marktreife,

Operationalisierung von Morbiditätsendpunkten in der Nutzenbewertung

Die Messung der Wirksamkeit, Verträglichkeit, Nützlichkeit und Schädlichkeit von Therapien erfolgt im AMNOG anhand von patientenrelevanten Endpunkten, die durch die Kategorien Mortalität, Morbidität, Nebenwirkungen und gesundheitsbezogene Lebensqualität bewertet werden. Die Frage ist, welche Endpunkte als patientenrelevant erachtet werden und wie sie operationalisiert werden können.

nen. Vor allem bei Morbiditätsendpunkten gab es in der Vergangenheit viel Interpretationsspielraum.

1. Klinische Relevanz: Definitionen müssen präzise und aussagekräftig sein, um die beabsichtigten klinischen Effekte angemessen zu messen.

2. Heterogenität des Patientenkollektivs: Es ist wichtig, dass die Endpunkte die Heterogenität der Patient:innen und ihre verschiedenen Krankheitsverläufe angemessen erfassen (Berücksichtigung von Subgruppen).

3. Konsens über Endpunkte: Die Definition von Morbiditätsendpunkten erfordert einen Konsens zwischen verschiedenen Interessengruppen, einschließlich medizinischer Fachgesellschaften, Forschern, Zulassungsbehörden, HTA-Behörden und Patientenvertretern.

4. Datenerfassung und -verfügbarkeit: Die Operationalisierung der Morbiditätsendpunkte muss im medizinischen Alltag umsetzbar sein damit die erforderlichen Daten zur Messung und Bewertung dieser Endpunkte erhoben werden können. Dies gilt vor allem auch für die Erfassung der Lebensqualität.

5. Patientenperspektive: Es ist wichtig, die Perspektive, Bedürfnisse und Präferenzen der Patient:innen bei der Festlegung und Bewertung patientenrelevanter Endpunkte angemessen zu berücksichtigen.

Die Einigung auf geeignete Endpunkte gelingt nicht immer und kann zeitaufwendig und ressourcenintensiv sein.

Im Rahmen des AMNOG kommt erschwerend hinzu, dass die Anforderungen des G-BA bezüglich der Operationalisierung und Messung von patientenrelevanten Endpunkten teilweise die klinische Fragestellung nicht vollständig abbilden. Dies kann dazu führen, dass Medikamente, die für Patient:innen eine wichtige Versorgungslücke schließen, im AMNOG keine Chance auf einen Zusatznutzen haben und somit möglicherweise nicht in Deutschland verfügbar sind.

Beispiel Ereignisfreies Überleben (EFS)

Ein Beispiel hierfür ist der Endpunkt „Ereignisfreies Überleben“ (EFS), der unter anderen den Krankheitsprogress bzw. Rezidive erfasst. In der kurativen Therapiesituation bildet das EFS aus Sicht des IQWiG und G-BA das Scheitern der potenziellen Heilung ab, welches grundsätzlich als patientenrelevant erachtet wird. Als Voraussetzung für die Akzeptanz des Endpunkts fordert der G-BA den Nachweis der Tumorkontrolle bzw. Tumorfreiheit. Insbesondere bei inoperablen Patienten ist die Krankheitsbeurteilung (Tumorkontrolle) im Laufe der Behandlungszeit aufgrund der oft durchgeführten Radiotherapie jedoch erschwert, da es im Zuge der Bestrahlung zu Gewebeveränderungen im Tumorgewebe kommt.

Die Fragestellung der Nutzenbewertung spiegelt somit nicht die der Kliniker wider, was im vorliegenden Beispiel dazu führt, dass sich die Forderungen des G-BA für die pUs nicht umsetzen lassen. Eine engere Abstimmung zwischen G-BA, Fachgesellschaften, Patientenvertretern und der klinischen Forschung ist zwingend erforderlich, um diesem Problem entgegenzuwirken. Ziel muss es sein, klinisch relevante und praktikable Morbiditätsendpunkte zu generieren. Der Bedarf hierfür ist hoch – insbesondere für inoperable Patient:innen mit kurativer Therapieintention.

Das Phänomen der Diskrepanz zwischen den Kriterien des G-BA und der Wirklichkeit im Klinikalltag besteht auch bei operablen Patient:innen. Der G-BA fordert als Voraussetzung für die Akzeptanz von EFS den Nachweis der Tumorfreiheit (R0-Resektion), was dazu führt, dass alle nicht vollständig resezierten Patient:innen in den EFS-Analysen als Ereignis gezählt werden. Nach dem Verständnis des IQWiG und G-BA ist für diese Patient:innen der kurative Therapieansatz gescheitert – und dass, obwohl die Patient:innen durchaus noch eine Chance auf Heilung besäßen.

Bei der Entwicklung des Studienprotokolls stellt sich für den pU die herausfordernde Frage, ob die Operationalisierung des Endpunkts nah am Versorgungsalltag sein soll und somit angemessen die variierenden Krankheitsverläufe der Patient:innen widerspiegelt oder ob der Erfolg der Nutzenbewertung bei der Erfassung des Endpunkts im Mittelpunkt stehen soll. Auch hier steht zur Diskussion, ob die Anforderungen des G-BA an derartige Endpunkte überdacht werden sollten, um den Praxisalltag besser abzubilden und den Patient:innen in den Mittelpunkt zu rücken.

Die Adressierung dieser Herausforderungen erfordert eine sorgfältige Abwägung und Zusammenarbeit zwischen allen beteiligten Parteien, einschließlich der Zulassungsbehörden, der pharmazeutischen Industrie, medizinischer Fachgesellschaften und Patientenvertretungen, um sicherzustellen, dass die Nutzenbewertung im Rahmen des AMNOG angemessen und umfassend erfolgt. Nur so kann sichergestellt werden, dass die Daten der Patient:innen, die Zeit, Engagement und Hoffnung in Studien investiert haben, auch sinnvoll benutzt werden können. Für pUs sowie für Patient:innen sind nicht bewertbare Datensätze eine Verschwendung wertvoller Ressourcen.

Patientenrelevante Endpunkte und Innovationen von morgen

Innovationen von morgen bringen das System vor neue Herausforderungen. Beispielsweise erforscht MSD gemeinsam mit Moderna individualisierte Neoantigen-therapien (INT) – auch bekannt als „therapeutische Impfstoffe gegen Krebs“. Das Ziel der INT ist es, das körpereigene Immunsystem bei bereits an Krebs erkrankten Patient:innen so zu aktivieren, dass es bei der Krebsbekämpfung unterstützen kann, um das Risiko eines Rückfalls zu minimieren. Die INT soll dem Immunsystem helfen, individuelle Neoantigene auf den Krebszellen der Patient:innen zu erkennen und zu

zerstören. Die aktuellen Forschungsergebnisse deuten darauf hin, dass die INT-Immunzellen so aktivieren könnten, dass sie die veränderten Neoantigene auf den Krebszellen erkennen und angreifen können. Aufgrund des Gedächtnisses des Immunsystems könnte die INT zu einer langfristigen, spezifischen Erkennung der Krebszellen führen und das Risiko eines Rückfalls verringern. In der Studie V940-001 wird die INT in Bezug auf den primären Endpunkt das Rezidivfreie Überleben (RFS) untersucht. Sekundäre Endpunkte umfassen das fernmetastasefreie Überleben (DMFS), das Gesamtüberleben (OS) und die Sicherheit der Therapie.

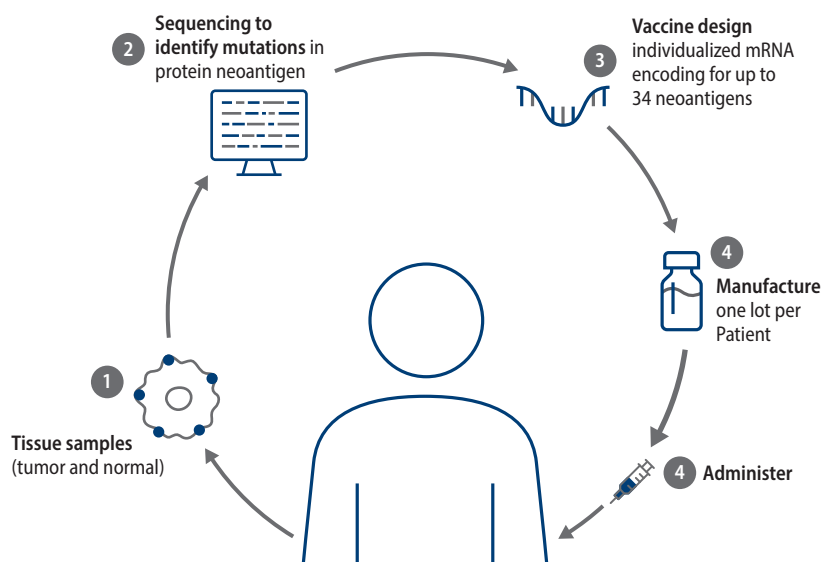
Operationalisierung von symptombezogenem progressionsfreiem Überleben (Progression-free survival PFS) (PFS+) mit Hilfe künstlicher Intelligenz

Ein wichtiger Endpunkt in der Onkologie stellt in einigen Indikationsgebieten der Endpunkt PFS dar, der per se in AMNOG nicht als patientenrelevant anerkannt wird. Eine Weiterentwicklung des Endpunktes (PFS+), der PFS mit Krankheitssymptomatik bzw. Lebensqualität verknüpft, könnte ein Weg zur Akzeptanz des Endpunktes in der frühen Nutzenbewertung darstellen, da er den direkten Einfluss des Progresses auf das Befinden des Patienten/der Patientin nachweisen würde. Damit wäre PFS+ im Sinne des AMNOG patientenrelevant. Der Einsatz von künstlicher Intelligenz könnte dabei methodisch unterstützen, um zum Beispiel anhand bestehender Datensätze Muster zu identifizieren, die diesen Zusammenhang nachweisen.

Wird die Zukunft der AMNOG-Endpunkte durch EU-HTA patientenzentriert?

Der medizinische Fortschritt schreitet stetig voran und in bestimmten Fällen sind randomisierte kontrollierte Studien (RCT) daher aus praktischen und ethischen Gründen

Individualisierte Neoantigentherapie – Studie mit Fokus auf den primären Endpunkt Rezidivfreies Überleben (RFS)



Advantages of ,RNA-based Cancer Therapies

Speed and efficiency

Needle-to-needle generation of personalized treatments can be produced in just a few weeks.

Integrated Manufacturing

Opportunities to scale and optimize within Moderna's existing Massachusetts manufacturing facility

Customizable

Identify and encode multiple patient-specific neoantigens to design an individualized treatment

Quelle: Moderna

Abbildung 3: Das Ziel der INT ist es, das körpereigene Immunsystem bei bereits an Krebs erkrankten Patient:innen so zu aktivieren, dass es bei der Krebsbekämpfung unterstützen kann, um das Risiko eines Rückfalls zu minimieren.

nicht durchführbar. Diese Dynamik muss auch bei der Bewertung des Zusatznutzens von Arzneimitteln abbildbar sein, um die Abkopplung vom wissenschaftlichen Fortschritt in der Arzneimittelentwicklung zu vermeiden. Für besondere Therapiesituationen, wie z.B. Gen- und Zelltherapien, wie auch INTs, für die Studien höchster Evidenzstufe unmöglich oder unangemessen sind, bedarf es daher einer Weiterentwicklung des AMNOG. Mit dem Start der europäischen HTA-Bewertung ab Januar 2025 für onkologische Therapeutika und Arzneimittel für neuartige Therapien (ATMP) bietet sich die Chance für Synergieeffekte. Methodische Bewertungsgrundsätze müssen besser an den europäischen Rahmen angeglichen werden. Es braucht

mehr Flexibilität bezüglich der Berücksichtigung von patientenzentrierten Endpunkten und spezifischen Studiendesigns (zum Beispiel indirekte Vergleiche) in Orientierung an europäischen Leitlinien, damit medizinischer Fortschritt auch in Zukunft in Deutschland eine Chance hat.

Literatur

¹ The pharmaceutical industry in figures, EFPIA, edition 2024, siehe the-pharmaceutical-industry-in-figures-2024.pdf Seite 5.

² vfa auf Basis von clinicaltrials.org siehe <https://go.sn.pub/7yka34>.

³ Fowler A, Grieve K, Maos A, Wilsdon T. Quantifying public and private investment in European biopharmaceutical research and development. Health Aff Sch. 2024 May 31 siehe Quantifying public and private investment in European biopharmaceutical research and development - PubMed Seite 2.

⁴ <https://go.sn.pub/xdyrkx>.

Patientenzentrierte Outcomes in der EU-HTA versus AM-Nutzenverordnung und IQWiG-Methodenpapier

	EU: patient-centered outcomes	AM-NutzenV: (& Verfahrensordnung)	IQWiG Methodenpapier Version 7.0
Mortalität	✓	✓	✓
Morbidität	✓	✓	✓
QoL	✓	✓	✓
Nebenwirkungen/Sicherheit	✓	✓	✓
Empfindungen (<i>feelings</i>)	✓	✗	✓
Überzeugungen (<i>beliefs</i>)	✓	✗	✗
Präferenzen	✓	✗	✗
Bedürfnisse (<i>needs</i>)	✓	✗	✗
Funktionen und Aktivitäten	✓	✗	✓

Quelle: vfa

Abbildung 4: Mit dem Start der europäischen HTA-Bewertung ab Januar 2025 bietet sich die Chance, methodische Bewertungsgrundsätze besser an den europäischen Rahmen anzugleichen.

⁵ Klinische Studien von MSD - msd.de

⁶ Citeline Clinical, Pharma R&D Annual Review 2024, siehe <https://go.sn.pub/1pfji1>. Figure 2 Seite 9.

Patientenrelevante Endpunkte: Fragen aus Sicht der Medizinischen Dienste

Dr. Andreas Rhode | Medizinischer Dienst Westfalen-Lippe

st das Herausstellen der Mortalität bei der Nutzenbewertung sachgerecht?

Die patientenrelevanten Endpunkte untersuchen Mortalität, Morbidität und die Lebensqualität. Bei vielen vormals rasch tödlich verlaufenden Erkrankungen hat der medizinische Fortschritt dazu geführt, dass sich das Überleben der Patienten verbessert hat. Außerdem verfälscht der Einsatz von weiteren Arzneimitteln, die in späteren Therapielinien gegeben werden, dann die Aussagen zum Gesamtüberleben. Lebensverlängerung stellt in diesen Situationen weder das primäre Therapieziel dar, noch ist es ein sinnvoller Studienendpunkt.

Das ereignisfreie, aber auch das progressionsfreie Überleben als Surrogatparameter werden mittlerweile als klinisch sinnvolle Endpunkte angesehen und Analysen zeigen auch eine Korrelation mit dem Gesamtüberleben (Michael Untch et al., Disease-free survival (DFS) as a surrogate for overall survival (OS) in patients (pts) with HR+/HER2-early breast cancer (EBC): A correlation analysis. JCO 41, 535-535(2023). DOI:10.1200/JCO.2023.41.16_suppl.535).

Es gibt jedoch auch genügend Beispiele, bei denen sich ein Vorteil im Progression-free survival (PFS) schließlich nicht signifikant im Gesamtüberleben abgebildet hat.

Muss das Kriterium der Mortalität bei der Nutzenbewertung neu gefasst werden?

Surrogatparameter als sinnvolles Messinstrument in der Nutzenbewertung

Bei vielen Erkrankungen kann sowohl die Mortalität, aber auch die Morbidität sinnvoll nicht im zeitlichen Rahmen einer Studie kontrolliert werden.

Kann in diesem Zusammenhang dann auf Surrogatparameter (wie PFS, DFS, aber auch laborchemische Parameter wie HbA1c etc.) zurückgegriffen werden?

Welche Voraussetzungen müssen diese Surrogatparameter erfüllen, damit tatsächlich Rückschlüsse auf patientenrelevante Parameter möglich sind?

Patientenrelevante Endpunkte als sekundärer Endpunkt

In der Entwicklung von Arzneimitteln werden die Studien in der Phase III auf die Zulassung und nicht auf die Nutzenbewertung ausgelegt. Daher sind die im Nutzenbewertungsverfahren dann möglicherweise zu bewertenden Parameter lediglich als sekundärer Endpunkt gesetzt gewesen (Beispiel primärer Endpunkt PFS, sekundärer Endpunkt Mortalität, Lebensqualität etc.).

Wie verzerrt sind die Ergebnisse bezogen auf patientenrelevante Endpunkte, wenn diese lediglich den sekundären Endpunkt dargestellt haben?

Patientenrelevante Endpunkte in indirekten Vergleichen

Da für die Zulassung häufig Studien durch die pU gewählt werden, die keinen direkten Vergleich zu einer zweckmäßigen Vergleichstherapie darstellen, da dieses durch die Zulassungsbehörden nicht gefordert wird und die zVT sich in den unterschiedlichen Ländern auch unterscheiden kann, fehlen im Nutzenbewertungsverfahren dann direkte Vergleiche zur zVT. Die pU wählen dann häufig indirekte Vergleiche. Hier werden zwei Studien über einen Brückenkompator miteinander untersucht. Dabei können sich in den Studien aber die Parameter für die gesetzten Endpunkte unterscheiden.

Wie kann man bezogen auf die patientenrelevanten Endpunkte mit indirekten Vergleichen umgehen?

Was sind die Mindestvoraussetzungen für eine Vergleichbarkeit?

Starre patientenrelevante Patientenendpunkte versus höchst individuelles Therapieansprechen

Bamberger 2020 berichtet beispielsweise für Epilepsien oder Depressionen von einem höchst individuellen Therapieansprechen. Es würden bei neurologischen und psychischen Erkrankungen Therapieprädiktoren fehlen, nach denen sich Patientenkollektive definieren ließen. Dadurch sei ein Zusatznutzen nicht nachweisbar (Bamberger M. Welchen patientenrelevanten Nutzen haben neue Arzneimittel in der Neurologie und Psychiatrie? Psychopharmakotherapie 27:289-294, 2020.).

Werden die patientenrelevanten Patientenendpunkte tatsächlich in der Nutzenbewertung zu starr gesetzt?

Unterläuft dieses gegebenenfalls den klinischen Mehrwert der Arzneimittel?



Dr. Andreas Rhode ist Facharzt für Psychiatrie und Psychotherapie und zertifizierter Gesundheitsökonom. Nach einer mehrjährigen oberärztlichen Tätigkeit am St. Marien-Hospital Hamm (Lehrstuhl für Psychotherapie und Psychosomatik der Universität Witten / Herdecke) übernahm er 2011 die Leitung der Fachklinik Release. Seit 2021 Leitung der Sozialmedizinischen Expertengruppe 6 Arzneimittelversorgung der Gemeinschaft der Medizinischen Dienste sowie in dieser Funktion für den GKV-SV beratende Tätigkeit im Unterausschuss Arzneimittel des G-BA. Zudem Mitglied der Off-Label-Expertengruppe und der Off-Label-Expertengruppe Long Covid des BfArM.

Erhebung von patientenberichteten Endpunkten: Einblicke aus der Psychiatrie

Professor Dr. Peter Falkai, Klinik für Psychiatrie und Psychotherapie, Klinikum der Universität München |
Professor Dr. Jörg Ruof, European Access Academy, Basel

Eine am Patientenwohl ausgerichtete moderne Psychiatrie stützt sich wesentlich auf differenzierte patientenberichtete Erhebungsinstrumentarien (PROMs = Patient Reported Outcome Measures). Eine Dominanz eines singulären Endpunktes wie zum Beispiel das Gesamtüberleben in der Onkologie ist aufgrund des komplexen Verlaufs psychischer Erkrankungen bei diesen jedoch nicht absehbar. Es sind vielmehr die verschiedenen Dimensionen der patientenberichteten Outcomes gleichermaßen zu berücksichtigen wie u.a.: generische Instrumente; krankheitsbezogene Instrumente und die Erhebung der PREMs (Patient Reported Experience Measures), d.h. der Erfahrungen, welche der Patient mit der Gesundheitsversorgung macht. Sprunginnovationen sind im Bereich psychischer Erkrankungen derzeit weniger absehbar. Die schrittweise Verbesserung der Versorgung psychiatrischer Patienten auf Basis der Erhebung von PROMs und PREMs und entsprechender Surrogatparameter wie „bezahlte Arbeit“ und „stabile Partnerschaft“ stehen im Zentrum der Forschungsbemühungen.

Bedeutung von Patient-Reported-Indicators in der Psychiatrie

Die europäische Regulation zum Health Technology Assessment (EU HTAR) hat eine Stärkung der European Health Union im Blick.¹ Es ist davon auszugehen, dass sich mittel- bis langfristig die Diskussion um die Zielgrößen für klinische Studien und die Patientenversorgung weiter vertieft und vereinheitlicht. Patientenberichteten Endpunkten kommt dabei eine zentrale Rolle zu. Dabei werden Patient-Reported-Outcomes (PROs) in der Methoden-Guidance der EU HTA-Koordinierungsgruppe als „any report of the status of the patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else“² definiert.

Bei psychischen Erkrankungen bilden diese Erhebungsinstrumente sowohl in der klinischen Routine als auch in der Forschung eine wichtige Rolle. So befasst sich seit 2017 eine Arbeitsgruppe innerhalb der OECD mit der Entwicklung internationaler Standards für Patient-reported-Indicators bei psychischen Erkrankungen.³ Die Harmonisierung der PRO-Erhebungsinstrumente ist dabei eine Voraussetzung, um Erfahrungen über die nationalen Grenzen hinweg zu teilen und bei Erkrankungen wie Depression, Bipolaren Störungen oder Schizophrenie international Zielgrößen zu definieren.

Die systematische Integration von PRO-Instrumentarien in der psychiatrischen Versorgung ermöglicht zum Beispiel bei suizidgefährdeten Patienten die Erfassung früher Warnsignale, oder im ambulanten oder teilstationären Bereich die Erhebung anderer wichtiger „Between-visit“-Informationen.⁴ Allerdings erfordert der zielführende Umgang mit diesen Instrumenten Sachkenntnis und Erfahrung. Die Kooperation und Motivation der Patienten spielt dabei eine wichtige Rolle, ebenso das „buy in“ seitens der

Mitarbeiter einer Klinik oder ambulanten Versorgungseinrichtung.

Differenzierung von PROM und PREM

Hinter der Begrifflichkeit PRO bzw. PROM (Patient Reported Outcome Measures) verbergen sich verschiedene Konzepte, die differenziert zu betrachten sind. In einer systematischen Übersichtsarbeit wurde der Outcome von Lebensstilinterventionen, wie gesunde Diät und körperliche Aktivität, bei schweren psychiatrischen Erkrankungen untersucht. Die Metaanalyse umfasst 21 Studien mit insgesamt 5907 Patienten. Dabei ergab sich kein Effekt der genannten Interventionen auf die Lebensqualität. Die krankheitsbezogenen Instrumente zur Erfassung des Schweregrades der Depression bzw. der Angstsymptomatik zeigten jedoch einen positiven Effekt.⁵ Konzeptionell sind entsprechend generische Erhebungsinstrumente mit einer breiten, umfassenden Perspektive von krankheitsbe-

Definition von PROMs und PREMs

For decades, clinician-rated outcome measures have been the central source of data informing clinical practice and policy

- Patient Reported Outcome Measures (PROMs) more directly assess the lived experiences of service users, capturing their perspectives on their health status and essential subjective constructs such as goal attainment, quality of life and social inclusion
- Patient Reported Experience Measures (PREMs) assess their experiences of using health services, including communication, responsiveness and recovery orientation

Quelle: Differenzierung von PROM und PREM nach Roe (2022)⁷

Abbildung 1: Von PROMs sind die PREMs zu unterscheiden, die vor allem auf die Erfahrungen des Patienten abheben.

zogenen, auf die spezifische Erkrankung ausgerichteten Erhebungsinstrumenten zu differenzieren. Weiterhin sind von den generischen oder krankheitsbezogenen PROMs



Prof. Dr. Peter Falkai ist seit 2012 Lehrstuhlinhaber der Klinik für Psychiatrie und Psychotherapie am Klinikum der LMU München. Präsident der Deutschen Gesellschaft für Psychiatrie, Psychotherapie und Nervenheilkunde von 2011 bis 2012, Mitglied der Leopoldina Nationale Akademie der Wissenschaften.



Prof. Dr. Jörg Ruof ist Mitglied im Beirat der Plattform Nutzenbewertung. Er ist Facharzt für Allgemeinmedizin, Master of Business Administration und Master of Public Health. Außerordentliche Professur am Institut für Epidemiologie, Sozialmedizin und Gesundheitssystemforschung der Medizinischen Hochschule Hannover. 2017 Gründung der r-connect GmbH. 2021 Gründung der „European Access Academy“.

die Patient-Reported Experience Measures (PREM) zu unterscheiden. Diese werden ebenfalls von den Patienten selbst berichtet, sind jedoch weniger auf den eigenen Gesundheitszustand als auf die Erfahrungen der Patienten mit der Gesundheitsversorgung ausgerichtet.^{6, 7}

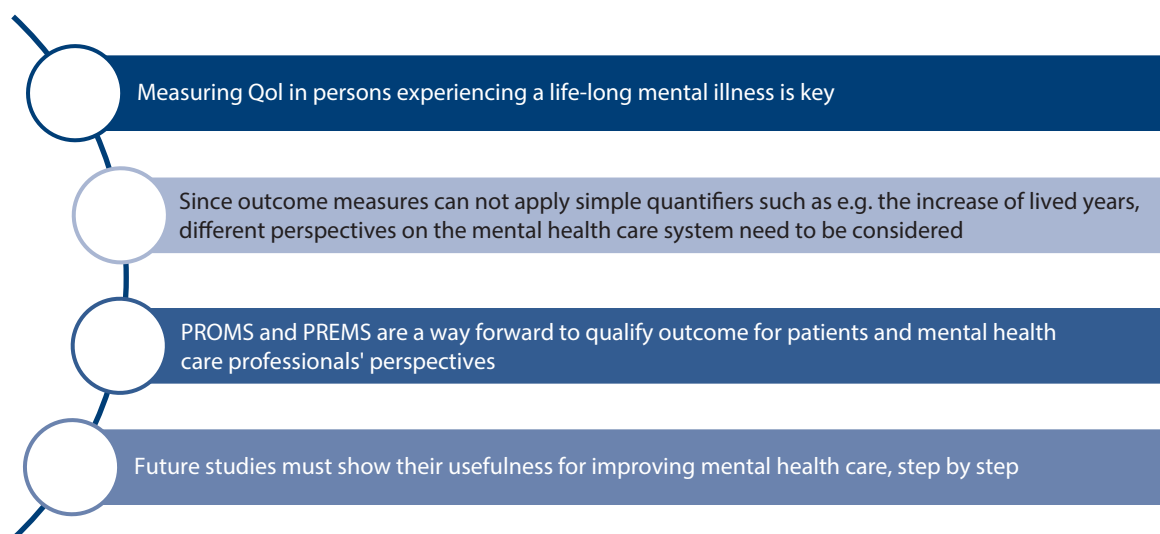
In einer 2023 publizierten, an der Université Paris Cité durchgeführten Studie wurden PROMs und PREMs von 248 Patienten erfasst, die an einer psychiatrischen Universitätsklinik aufgenommen wurden. Alle drei Dimensionen: i) generische PROMs; ii) krankheitsspezifischen PROMs; und iii) PREMs wurden erhoben. Dabei zeigte sich eine Verbesserung der PROMs während des Klinikaufenthaltes. Die PREMs lagen sehr nahe bei den parallel erhobenen Ratings der Ärzte, waren aber nur schwach mit den etablierten kli-

nischen PRO-Messinstrumenten korreliert. Die Autoren empfehlen entsprechend die Messung aller drei Dimensionen als Indikatoren für den Therapieerfolg und die Qualität der Patientenversorgung.⁸

Trends in der patientenzentrierten Gesundheitsversorgung

Aktuelle Trends hin zu „User-Centered Health Care-Models“ oder auch dem „Shared-Decision-Making“ zeigen die Relevanz der patientenzentrierten, partnerschaftlichen Gesundheitsversorgung. Ethische Erwägungen, die Berücksichtigung der Patientenrechte und vor allem die Ausrichtung auf bessere Therapieerfolge durch die aktive und gleichberechtigte Einbindung der Patienten, machen

Überblick zum Status der Outcomes-Messung in der Psychiatrie



Quelle: Eigene Darstellung

Abbildung 2: Eine Dominanz eines singulären Endpunktes ist in der Psychiatrie nicht absehbar. Stattdessen müssen verschiedene Dimensionen der patientenberichteten Outcomes gleichermaßen berücksichtigt werden.

diese Trends unumkehrbar. Die Integration unterschiedlicher Perspektiven (Patient/ Angehörige/Care Giver) ist dabei ebenso wichtig wie die Betrachtung der verschiedenen Dimensionen der patientenberichteten Endpunkte.

Eine Befragung von Patienten und des MDK⁹ ergab:

- Mortalität ist in der Regel kein geeigneter Outcome bei psychischen Erkrankungen (mit Ausnahme von Suizidstudien).
- Rückfallfreies und (voll) funktionsfähiges Langzeitüberleben (Recovery-Konzept) sind hingegen sinnvoll.
- Surrogatparameter hierfür sind bezahlte Arbeit und stabile Partnerschaft (sekundäre Endpunkte). Primäre Endpunkte sind weiterhin die Messung der Psychopathologie und der Funktion.

PROs sind eine weitere wichtige Möglichkeit, primär die Wirksamkeit von Arzneimitteln oder sonstigen Therapieinterventionen zu überprüfen

Zusammenfassend ist zu betonen, dass sich eine am Patientenwohl ausgerichtete moderne Psychiatrie wesentlich auf differenzierte patienten-zentrierte Erhebungsinstrumentarien stützt. Eine Dominanz eines singulären Endpunktes wie zum Beispiel das Gesamtüberleben in der Onkologie ist dabei nicht absehbar und wahrscheinlich bei der Komplexität psychischer Erkrankungen auch nicht sinnvoll.

Es sind vielmehr die verschiedenen Dimensionen der patientenberichteten Outcomes gleichermaßen zu berücksichtigen, wie unter anderem: generische breit angelegte Instrumente; krankheitsbezogene Instrumente und die Erhebung der PREMs, d.h. der Erfahrungen, die der Patient mit der Gesundheitsversorgung macht. Sprunginnovationen sind derzeit weniger absehbar. Die schrittweise Verbesserung der Versorgung psychiatrischer Patienten auf Basis der Erhebung von PROMs und PREMs und entspre-

chender Surrogatparameter wie „bezahlte Arbeit“ und „stabile Partnerschaft“ stehen im Zentrum der Forschungsbemühungen.

Literatur

¹ Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU. <https://go.sn.pub/kp5l9f> [Zugriff: 25.1.2025]

² HTA Coordination Group: Guidance on outcomes for joint clinical assessments. Adopted on 10 June 2024 by the HTA CG pursuant to Article 3(7), point (d), of Regulation (EU) 2021/2282 on Health Technology Assessment <https://go.sn.pub/qahh23> [Zugriff: 25.1.2025]

³ de Bienassis K, Kristensen S, Hewlett E, et al. Patient-reported indicators in mental health care: towards international standards among members of the OECD. *Int J Qual Health Care* 2021; 34 (Suppl 1): ii7-ii12.

⁴ Fowler JC, Madan A, Frueh BC et al. Lessons learned while integrating patient-reported outcomes in a Psychiatric Hospital. *Psychotherapy* 2019; 56: 91-99

⁵ Pape LM, Adriaanse MC, Kol J et al. Patient-reported outcomes of lifestyle interventions in patients with severe mental illness: a systematic review and meta-analysis. *BMC Psychiatry* 2022; 22: 261 <https://doi.org/10.1186/s12888-022-03854-x> [Zugriff: 25.1.2025]

⁶ de Bienassis K, Kristensen S, Hewlett E, et al. Measuring patient voice matters: setting the scene for patient-reported indicators. *Int J Qual Health Care* 2021; 34 (Suppl 1): ii3-ii6.

⁷ Roe D, Slade M, Jones N. The utility of patient-reported outcome measures in mental health. *World Psychiatry*. 2022 Feb;21(1):56-57

⁸ Scanferla E, de Bienassis K, Pachoud B, Gorwood P. How subjective well-being, patient-reported clinical improvement (PROMs) and experience of care (PREMs) relate in an acute psychiatric care setting? *European Psychiatry* 2023; 66: e26 1-9 <https://doi.org/10.1192/j.eurpsy.2023.12> [Zugriff: 25.1.2025]

⁹ Persönliche Kommunikation mit Dr. Andreas Rhode, Medizinischer Dienst Westfalen-Lippe

Endpunkte – Herausforderungen aus der Sicht des vfa

Dr. Andrej Rasch | Verband forschender Arzneimittelhersteller

Im Rahmen der AMNOG-Nutzenbewertung bestehen zahlreiche praktische und methodische Herausforderungen im Umgang mit Endpunkten wie Gesamtüberleben, PROs oder Surrogatendpunkten. Neben der Frage nach grundsätzlicher Relevanz besteht viel Diskussionsbedarf auch hinsichtlich der Gewichtung und Gesamtabwägung einzelner Zielgrößen und Effekte. Es bedarf hier einer stärkeren Fokussierung auf akzeptierte und etablierte Methoden, die den internationalen Standards der evidenzbasierten Medizin entsprechen, einer höheren Transparenz der Einordnung und Abwägung von Endpunkten sowie einer Berücksichtigung der Besonderheiten von Therapiesituationen.

Welche Zielgrößen sind relevant für Patienten?“ ist die Leitfrage der Herbsttagung 2024 der Plattform zur Nutzenbewertung. Die Frage nach den Endpunkten ist eine der zentralen Fragestellungen im AMNOG-Verfahren. Doch welchen Stellenwert haben die Endpunkte für die Feststellung eines Zusatznutzens im AMNOG-Verfahren? Hierarchisch muss nüchtern festgestellt werden, dass es zunächst auf das zum AMNOG-Verfahren kompatible Studiendesign einer Studie (also regelhaft randomisiert-kontrolliert) von ausreichender Dauer und die richtige Umsetzung der Vorgabe zur zweckmäßigen Vergleichstherapie ankommt. Sind diese zentralen Bedingungen jedoch erfüllt, kommt es tatsächlich auf die Endpunkte und die dafür gezeigten Studieneffekte an.

Aus rechtlicher Sicht gilt dabei nach Paragraph 2 Abs. 3 AM-NutzenV die Vorgabe, dass der „Nutzen eines Arzneimittels [...] der patientenrelevante therapeutische Effekt insbesondere hinsichtlich der Verbesserung des Gesundheitszustands, der Verkürzung der Krankheitsdauer, der Verlängerung des Überlebens, der Verringerung von Nebenwirkungen oder einer Verbesserung der Lebensqualität“ ist. Der Zusatznutzen ist demnach festzustellen über eine Beeinflussung „patientenrelevanter Endpunkte“. Der Status eines Endpunkts in den zulassungsrelevanten Studien (primär oder sekundär) hat zugleich keine Konsequenz für die Relevanz in der Nutzenbewertung.

Nach rund 13 Jahren der Bewertungspraxis konnten rund 57 Prozent der neuen Arzneimittel ihren Zusatznutzen belegen. Dieser Anteil variiert jedoch deutlich je nach Therapiegebiet. Während für onkologische Arzneimittel ein Zusatznutzen in rund 73 Prozent belegt werden konnte, lag der Anteil des belegten Zusatznutzens für Krankheiten des Nervensystems oder psychiatrische Erkrankungen

bei 46 Prozent bzw. nur 29 Prozent liegt. Offenkundig konnte hier die Kombination aus einem kompatiblen Studiendesign und Endpunkten für onkologische Arzneimittel im AMNOG bessere Ergebnisse erzielen als bei anderen Indikationen.

Gesamtüberleben

Dabei ist festzustellen, dass vor allem in der Onkologie der Endpunkt „Gesamtüberleben“ von herausragender Bedeutung ist. Er ist ohne alle Zweifel patientenrelevant und damit auch maßgeblich für den Nachweis und Einstufung des Zusatznutzens. Die Empirie für onkologische Erkrankungen zeigt dabei durchaus eine Abhängigkeit zwischen der Größe des Vorteils beim Gesamtüberleben (in Kombination mit Vorteilen bei anderen Endpunktkategorien) und dem Ausmaß des Zusatznutzens in den Entscheidungen des G-BA.



Dr. PH Andrej Rasch ist seit 2013 beim Verband Forschender Arzneimittelhersteller e.V. (vfa) als Senior Manager Nutzenbewertung/HTA-Koordination tätig. Davor war er als Forschungsbereichsleiter Arzneimittel beim Wissenschaftlichen Institut der AOK (WIdO), Methodiker beim Gemeinsamen Bundesausschuss (G-BA) und wissenschaftlicher Mitarbeiter an der Fakultät für Gesundheitswissenschaften der Universität Bielefeld tätig.

Auch für den Endpunkt „Gesamtüberleben“ besteht jedoch eine Reihe an praktischen und methodischen Herausforderungen. So ist nicht in allen Therapiesituationen eine angemessene Studien-Power im vertretbaren Zeitrahmen erreichbar. Dies wird im Rahmen der Nutzenbewertung häufig dadurch zusätzlich verstärkt, dass eine Studienpopulation in Teilfragestellungen seitens des G-BA aufgeteilt wird oder nur ein Teil der pivotalen Studienpopulation als verwertbar herangezogen wird. Dies, ohne eine erkennbare Berücksichtigung der Power-Problematik.

In einigen Situationen ist zudem ein erlaubter Therapiewechsel ethisch unvermeidbar, was zur erschwerten Interpretation der Überlebensdaten führt. Zwar bestehen zum Umgang mit dem erlaubten Therapiewechsel methodische Lösungen, keiner dieser Ansätze wird jedoch bislang in der AMNOG-Bewertung akzeptiert. Ebenso herausfordernd sind hinsichtlich Interpretation der Ergebnisse zum Gesamtüberleben die Anforderungen an die Erfassung der Folgetherapien.

Wenn eine bestimmte Therapiesituation (noch) keine Aussagen zum Gesamtüberleben in einem vertretbaren Zeitrahmen erlaubt, stellt sich zudem die Frage, wie ein patientenrelevanter Behandlungserfolg gemessen werden kann. In der Entscheidungspraxis des G-BA werden inzwischen auch Endpunkte wie EFS (event-free survival), DFS (disease-free survival) oder RFS (relapse-free survival) grundsätzlich akzeptiert, jedoch nur zur Abbildung des Scheiterns eines kurativen Therapieansatzes.

Nicht bewertungsrelevant ist hingegen weiterhin der Endpunkt PFS (progression-free survival), wobei zum PFS von Beginn an unterschiedliche Auffassungen innerhalb des G-BA hinsichtlich der Patientenrelevanz bestehen. Auch solche Endpunkte wie TTST (time to subsequent therapy), CR (complete response, bis auf den alten Sonderfall der Bewertung zum Basallzellkarzinom) oder MRD (mi-

nimal residual disease) bleiben ohne Berücksichtigung. Diese, für die Zulassungsentscheidungen gewichtigen oder gar primären Endpunkte, werden im AMNOG-Verfahren, unabhängig der Besonderheiten einzelner Therapiesituationen, durchgehend als nicht patientenrelevant und damit nicht bewertungsrelevant eingestuft.

Surrogatendpunkte

Zusätzlich zur Frage einer direkten Patientenrelevanz stellt sich bei einigen Endpunkten die Frage, ob sie für bestimmte Therapiesituationen als Ersatz (Surrogate) für andere Endpunkte herangezogen werden können. Bei AMNOG-Nutzenbewertungen gelten für eine Surrogatvalidierung die Vorgaben des IQWiG, die mit dem Rapid Report A10-05 „Aussagekraft von Surrogatendpunkten in der Onkologie“ im Jahr 2011 publiziert wurden (IQWiG 2011).

Für eine Surrogatvalidierung wird dabei idealerweise eine Meta-Analyse von mehreren RCTs mit hoher Ergebnis-sicherheit sowie ein hohes Korrelationsmaß auf Studien- und individueller Ebene oder anderenfalls die Anwendung des Konzepts eines Surrogate-Threshold-Effekts (STE) mit bestimmten Schwellenwerten vorausgesetzt. Obwohl die Anforderungen des G-BA keine expliziten Schwellenwerte vorgeben, beziehen sie sich auf dieselbe Methodik, wie sie das IQWiG vorschlägt. Soweit der Anspruch. Die Wirklichkeit ist zugleich ernüchternd, da diese Anforderungen seit nunmehr 13 Jahren nicht erfüllt werden konnten. Ob dies nun an den konservativen Anforderungen selbst liegt oder einer (ggf. daraus folgend) überschaubaren Anzahl an Validierungsversuchen, kann nicht abschließend beantwortet werden.

Dennoch ist festzuhalten, dass Surrogatendpunkte in einigen Fällen unverzichtbar sind, da sonst zentrale Fragestellungen schlicht nicht untersucht werden können, um den Zugang zu innovativen Behandlungen zu ermög-

lichen. Auch im Rahmen der AMNOG-Nutzenbewertung wurden daher in der Vergangenheit in absoluten Ausnahmefällen einige Surrogatendpunkte auch ohne nach o.g. Methoden formale Validierung akzeptiert und zur Ableitung des Zusatznutzens herangezogen.

So wurden z.B. der Endpunkt Sustained Virologic Response bei chronischer Hepatitis C, das virologische Ansprechen bei einer HIV-Infektion oder HbA_{1c} bei Diabetes Mellitus Typ 1 als ausreichend valide Surrogatendpunkte sowohl von IQWiG als auch vom G-BA eingestuft. Die Berücksichtigung dieser Surrogatendpunkte war zwar stets nachvollziehbar, dennoch blieb die jeweilige Entscheidung bei einer wenig transparenten Einzelabwägung.

Patientenberichtete Endpunkte

Patientenberichtete Endpunkte (PROs) nehmen insgesamt einen zunehmend hohen Stellenwert ein. Für viele Therapiegebiete gehört eine Erfassung der Morbidität und der gesundheitsbezogenen Lebensqualität mit solchen Endpunkten inzwischen zum Standard klinischer Prüfungen. So zeigt sich in einer Analyse der Zulassungsentscheidungen in der EU für onkologische Arzneimittel in den Jahren 2017 bis 2020, dass PROs in rund 78 Prozent der Fälle in den pivotalen Studien eingeschlossen wurden (Teixeira et al. 2022).

Auch in der AMNOG-Nutzenbewertung zeigt sich die zunehmende Bedeutung von PROs. So am Beispiel des nicht-kleinzelligen Lungenkarzinoms, wo in den vom G-BA berücksichtigten Studien in 95 Prozent verwertbare Daten zu mindestens einem PRO-Instrument vorlagen (Brand et al. 2022). Auch für die Untergruppe der patientenberichteten Endpunkte, die gesundheitsbezogene Lebensqualität, zeigt sich ein ermutigendes Bild. So war der Anteil der Verfahren mit Daten zur Lebensqualität in den letzten Jahren zunehmend und lag seit 2014 bei über 70 Prozent. Beson-

ders hoch war dieser Anteil bei Nutzenbewertungen zu onkologischen Arzneimitteln (Kramer et al. 2024).

Die Frage nach grundsätzlicher Relevanz hört jedoch nicht bei der Art eines Endpunktes auf, sondern kann sich auch auf seine Operationalisierung erstrecken. So kann eine, als eindeutig relevant erscheinende PRO-Zielgröße, in einer Bewertung doch unberücksichtigt bleiben. Dies lässt sich am Beispiel von Nutzenbewertungen im Therapiegebiet einer mittelschweren bis schweren Plaque-Psoriasis zeigen. Hier wird der Endpunkt PASI 90, der für eine 90-prozentige Verbesserung der Krankheitssymptome und nahezu symptomfreie Haut steht, vom IQWiG seit Jahren nicht berücksichtigt, da hier formal nicht auszuschließen sei, dass Psoriasis-Symptome weiterhin vorhanden sind und die Patient:innen beeinträchtigen. Aus diesem Grund werden vom IQWiG ausschließlich Auswertungen zu PASI 100 (komplette Remission) herangezogen. Diese Einschätzung stand von Beginn an im Widerspruch zu den Leitlinien und der Versorgungspraxis, wo auch PASI 75 und PASI 90-Antworten als Therapieziele dienen, da eine Abwesenheit von kutanen Symptomen nicht bei allen Patienten erreicht werden kann (Nast et al. 2021). Auch vom G-BA werden daher die entsprechenden Ergebnisse zu PASI 75 und PASI 90 berücksichtigt.

Auch für die Erfassung der PROs besteht eine Reihe an Herausforderungen. So liegen für manche besondere Therapiesituationen, wie seltene Erkrankungen, keine validierten und etablierten Instrumente vor. Eine Anwendung verfügbarer Fragebögen aus anderen Therapiegebieten wird dabei stets kritisch gesehen. Bei der Interpretation der Studienergebnisse bleibt die mögliche Power-Problematik unberücksichtigt. Herausfordernd ist zudem die Erfassung und Einhaltung hoher Rücklaufquoten insb. bei terminalen Lebensphasen und nach Progress einer lebensbedrohlichen Erkrankung (Böhme et al 2022).

Bei der Dauer der PRO-Erfassung bestanden bis zuletzt unterschiedliche Auffassungen. Einerseits zwischen dem IQWiG, das eine möglichst lange Dokumentation bis zum Studienende propagiert, und der klinischen Fachwelt, die eine Erfassung nach Progress zwar für wichtig hält, jedoch im sinnvollen Maße und nicht uneingeschränkt bis zum Lebensende.

Umgang mit verfügbarer Evidenz

Hinsichtlich der grundsätzlichen Akzeptanz der Daten ist auf die bestehende Regelung im § 5 Abs. 5 AM-NutzenV hinzuweisen, die besagt: „Können zum Zeitpunkt der Bewertung valide Daten zu patientenrelevanten Endpunkten noch nicht vorliegen, erfolgt die Bewertung auf der Grundlage der verfügbaren Evidenz unter Berücksichtigung der Studienqualität mit Angabe der Wahrscheinlichkeit für den Beleg eines Zusatznutzens und kann eine Frist bestimmt werden, bis wann valide Daten zu patientenrelevanten Endpunkten vorgelegt werden sollen.“

Die Regelung zielt zum einen auf die Möglichkeit einer Befristung ab, was an sich gelebte Praxis ist. Zum anderen erfolgt die Maßgabe, die Bewertungen auf der Grundlage der verfügbaren Evidenz vorzunehmen. In der Praxis zeigt sich jedoch, dass verfügbare Daten regelhaft nicht herangezogen werden, wenn sie nicht den patientenrelevanten Endpunkten eingeordnet werden. Hier stellt sich die Frage, ob vor allem bei besonderen Therapiesituationen eine Bewertung unter Berücksichtigung der verfügbaren Evidenz erfolgen sollte.

Gewichtung der Endpunkte und Effekte

Neben der grundsätzlichen Frage zur Relevanz eines Endpunktes stellt sich im Rahmen einer Nutzenbewertung zudem die Frage, wie relevant ein Endpunkt oder Effekt ist. Das IQWiG unterscheidet in den eigenen Methoden drei

hierarchische Kategorien der Zielgröße: 1. Gesamtmortalität, 2. schwerwiegende (bzw. schwere) Symptome und Nebenwirkungen sowie gesundheitsbezogene Lebensqualität sowie 3. nicht schwerwiegende (bzw. nicht schwere) Symptome und Nebenwirkungen.

Allein die Einordnung einer Zielgröße als schwerwiegend oder nicht schwerwiegend ist jedoch nicht immer hinreichend transparent und trivial. So ist am bereits skizzierten Beispiel eines Anwendungsgebiets wie mittelschwere bis schwere Plaque-Psoriasis zu beobachten, dass eine pauschale Einstufung des Endpunktes PASI 100 (Komplettremission) unter „nicht schwerwiegende / nicht schwere Symptome“ durchaus Fragen aufwerfen kann. In vielen Fällen basiert die hierarchische Zuordnung einer Zielgröße (so z.B. aus einem EORTC QLQ-C30 Fragebogen bei onkologischen Erkrankungen) nur durch die formale Einordnung in die Kategorien Morbidität oder Lebensqualität. Dies kann zur systematischen Verzerrung der Einstufungen bei Endpunkten der Morbiditätskategorie führen, da sie im Zweifel unter „nicht schwerwiegend“ verortet werden und damit eine höhere Hürde bei der Bewertung haben.

Diesbezüglich wirft sicherlich auch die Methodik zur Bestimmung des Ausmaßes eines Zusatznutzens eine Reihe an kritischen Fragen auf. So steht dieser Sonderweg zur Ausmaßbestimmung für relative Effektmaße von Beginn an in der Kritik, insb. aufgrund der (im institutseigenen Binnenkonsens) festgelegten pauschalen Schwellenwerte für obere Konfidenzintervallgrenzen, normative Setzungen oder die Annahme von zwei Studien über alle Therapie-situationen.

Zwar wird auf die IQWiG-Methodik zur Ausmaßbestimmung vom G-BA seit 2011 nicht abgestellt (dies wird in allen tragenden Gründen zu den Beschlüssen explizit erwähnt), dennoch ist davon auszugehen, dass sie die

Nutzenbewertungen nachhaltig prägt. Auch die festgesetzten Schwellenwerte für stetige Zielgrößen in Kombination mit dem konservativen Ansatz einer verschobenen Hypothesengrenze entsprechen nicht den international anerkannten Kriterien oder Standards der evidenzbasierten Medizin und stellen so eine zusätzliche Herausforderung dar (IQWiG 2022).

Problematisch ist zuletzt auch die Bewertung der grundsätzlichen Relevanz der PRO-Effekte. Die Forderung von etablierten und validierten MID-Schwellen (minimal important difference) wurde bei Responderanalysen mit einer starren 15 Prozent-Formel ersetzt. Sind demnach in einer Studie Responderanalysen präspezifiziert und entspricht das Responsekriterium mindestens 15 Prozent der Skalenspannweite des verwendeten Erhebungsinstruments, so werden diese Analysen für die Bewertung herangezogen.

Dieser „One-size-fits-all“-Ansatz ist jedoch aus vielerlei Hinsicht umstritten, insb. da auch er als Sonderweg den Entwicklungsansatz der internationalen Wissenschaft zur Verbesserung von Bewertungsstandards mittels sinnvoller Qualitätskriterien außer Acht lässt und bekannte Unterschiede der Patientensicht auf bedeutsame Ergebnisse nicht hinreichend berücksichtigt (Böhme et al 2022, Schlichting et al 2022).

Die IQWiG-Methodik führt zudem dazu, dass selbst bei gewährleistetester klinischer Relevanz durch das vordefinierte Responderkriterium ein statistisch signifikanter Effekt für manche PROs nicht zwangsläufig zum anerkannten Effekt führt. Denn zusätzlich zum o.g. Responsekriterium von 15 Prozent greift ein weiteres Relevanzkriterium, nämlich das des Schwellenwertes für das obere Konfidenzintervall (für nicht schwerwiegende Symptome). Dies führt zu einer Doppelung der Relevanzkriterien und einer überkonservativen Einordnung der PRO-Effekte.

Gesamtabwägung

Schließlich steht eine Gesamtabwägung der Endpunkte und Therapieeffekte im Fokus einer Entscheidung zum Zusatznutzen. Sie erfolgt vom G-BA stellvertretend für die Patient:innen und deren Präferenzen, jedoch ohne eines formalen und hinreichend transparenten Verfahrens bei der Gewichtung. Studien zur Messung von Patientenpräferenzen wurden im AMNOG-Verfahren bislang nicht berücksichtigt.

Einige Einordnungen werfen hier Fragen auf, so z.B. im Falle einer Einordnung der Therapievorteile als geringer Zusatznutzen. Ein solcher liegt im Sinne der AM-NutzenV dann vor, wenn eine „bisher nicht erreichte moderate und nicht nur geringfügige Verbesserung des therapierelevanten Nutzens [...] erreicht wird, insbesondere eine Verringerung von nicht schwerwiegenden Symptomen der Erkrankung oder eine relevante Vermeidung von Nebenwirkungen“.

In der Bewertungspraxis des G-BA fallen darunter jedoch auch Bewertungen mit einer Verlängerung des Gesamtüberlebens, einer Vermeidung von Rezidiven bei onkologischen Erkrankungen, häufigeren Komplettremissionen der schweren Plaque-Psoriasis bei Kindern und Jugendlichen oder auch multiplen Vorteilen bei Patient:innen mit mittelschwerem bis schwerem aktivem Morbus Crohn. Eine Abwägung der Effekte ist zuletzt mit der Einführung der sog. „Leitplanken“ im GKV-FinStG noch entscheidender geworden, da bereits eine kleine Veränderung in der Einordnung des Zusatznutzenausmaßes (auch über Abbildung von methodischen Restunsicherheiten) über den Geltungsrahmen dieser Leitplanken in den Verhandlungen entscheiden kann.

Europäische Perspektive

Eine perspektivische und doch unmittelbar bevorstehende

Herausforderung ist zudem der Prozess des europäischen HTA. Dieser startet im Januar 2025 mit der Bewertung von Arzneimitteln für neuartige Therapien (ATMP) sowie onkologischen Arzneimitteln. Ab dem Jahr 2028 sollen dann Bewertungen von Orphan Drugs und ab 2030 für andere Medikamente folgen. Eine wesentliche Unsicherheit ist dabei die unsichere Anzahl an nationalen PICO-Fragestellungen, die sich u.a. aus angeforderten und verfügbaren Endpunkten, aber auch möglichen Operationalisierungen und Auswertungen der Endpunkte zusammensetzt.

Dies mündet zugleich in der Herausforderung, wie das nationale „Delta-Dossier“ für den AMNOG-Prozess auszugestalten sein wird. Ebenso spannend bleibt es, ob zukünftig die angestrebte Harmonisierung der methodischen Anforderungen erreicht wird und welche Wechselwirkungen sich aus dem unterschiedlichen Umgang mit Endpunkten im europäischen HTA und der AMNOG-Nutzenbewertung ergeben. So zum Beispiel auch hinsichtlich des verankerten Ansatzes von „patientenzentrierten Endpunkten“ (inklusive z.B. Präferenzen oder Bedürfnisse), der im europäischen HTA verfolgt wird.

Fazit

Abschließend können zentrale Herausforderungen im Umgang mit Endpunkten im Rahmen der Nutzenbewertung wie folgt festgehalten werden. Es bedarf:

- einer stärkeren Fokussierung auf akzeptierte und etablierte Methoden, die den internationalen Standards der evidenzbasierten Medizin entsprechen,
- einer höheren Transparenz der Einordnung und Abwägung von Endpunkten,
- einer Berücksichtigung der Besonderheiten von Therapiesituationen im Rahmen der Nutzenbewertung.

Literatur

¹ Böhme et al. Es braucht neue Ansätze für Lebensqualitätsdaten, in Monitor Versorgungsforschung (01/22), S. 43-47

² Brand et al. Value in Health, Volume 25, Issue 12S (December 2022), <https://go.sn.pub/y32ial>.

³ IQWiG. Aussagekraft von Surrogatendpunkten in der Onkologie. Rapid Report. IQWiG-Berichte – Jahr: 2011 Nr. 80

⁴ IQWiG. Dokumentation und Würdigung der Anhörung zum Entwurf der Allgemeinen Methoden 6.1, Version 1.0 vom 22.03.2022

⁵ Kramer et al. Health-related quality of life (HRQoL) in German early benefit assessment: The importance of disease-specific instruments, in: ZEFQ (2024), <https://doi.org/10.1016/j.zefq.2024.02.003>

⁶ Nast et al. Deutsche S3-Leitlinie zur Therapie der Psoriasis vulgaris, adaptiert von EuroGuiDerm – Teil 1: Therapieziele und Therapieempfehlungen. J Dtsch Dermatol Ges 2021 Jun;19(6):934-951.

⁷ Schlichting et al. Is IQWiG's 15% Threshold Universally Applicable in Assessing the Clinical Relevance of Patient-Reported Outcomes Changes? An ISPOR Special Interest Group Report. Value Health 2022 Sep;25(9):1463-1468.

⁸ Teixeira et al. (2022) A review of patient-reported outcomes used for regulatory approval of oncology medicinal products in the European Union between 2017 and 2020. Front. Med. 9:968272.

Patientenrelevante Endpunkte im Kontext des europäischen HTA-Prozesses

Dr. Johanna Seeger | Gemeinsamer Bundesausschuss

Die ersten gemeinsamen europäischen Bewertungen (Joint Clinical Assessments, JCA) und Beratungen stehen ab Januar 2025 bevor und erfordern klar definierte methodische Anforderungen und ein gemeinsames Verständnis zentraler Konzepte. Die Erfahrungen aus Übungsverfahren zur Bestimmung des Bewertungsumfanges im sogenannten „Scoping“-Verfahren flossen in die Entwicklung verschiedener Guidance-Dokumente ein. Hinsichtlich der Endpunkte leistet die „Guidance on Outcomes for JCA“ einen wichtigen Beitrag zur Harmonisierung der Anforderungen zu Endpunkten im Rahmen des JCA. Die Deutungshoheit über die Relevanz von Ergebnissen zu bestimmten Endpunkten verbleibt jedoch bei den Mitgliedstaaten.

Einleitung
Mit dem Geltungsbeginn der Verordnung (EU) 2021/2282 (EU-HTA-Verordnung) am 12. Januar 2025 stehen die ersten gemeinsamen europäischen Bewertungen (Joint Clinical Assessments, JCA) und Beratungen (Joint Scientific Consultations, JSC) für Arzneimittel kurz bevor. Der Geltungsbereich der EU-HTA-Verordnung wird zunächst auf Arzneimittel zur Therapie onkologischer Erkrankungen und Arzneimittel für neuartige Therapien (Advanced Therapy Medicinal Products, ATMP) begrenzt sein, jedoch nach schrittweiser Ausweitung ab dem 13. Januar 2030 alle neuen Arzneimittel einschließen.¹

Eine Voraussetzung für die Umsetzung der EU-HTA-Verordnung ist die Entwicklung eines gemeinsamen Verständnisses der EU-Mitgliedstaaten und die Definition gemeinsamer methodischer Anforderungen als Grundlage für die Durchführung von JCA und JSC. Hierzu werden in den Subgruppen der HTA-Koordinierungsgruppe Guidance-Dokumente zu verschiedenen prozessualen und methodischen Aspekten entwickelt.

Das Rückgrat von JSC und JCA bildet das PICO-Schema, welches die Fragestellung der europäischen HTA-Bewertung anhand der Parameter „Population, Intervention, Comparator, Outcomes“ definiert. Eine geeignete Operationalisierung von Endpunkten ist für die Generierung aussagekräftiger Studiendaten für das JCA-Verfahren essenziell und stellt somit auch einen elementaren Beratungsaspekt in den JSC dar.

Endpunkte im Scoping-Prozess

Die Festlegung des Bewertungsumfanges für einen JCA erfolgt im sogenannten „Scoping“-Verfahren. Hierbei werden die nationalen Fragestellungen der Mitgliedstaaten anhand von PICO-Schemata abgefragt. Anschließend

erfolgt eine Konsolidierung mit dem Ziel, die Bedarfe der einzelnen Mitgliedstaaten durch eine möglichst geringe Anzahl an PICOs abzubilden.

Eine Erprobung dieses Prozesses erfolgte im Rahmen des Projektes „EUnetHTA 21“ beispielhaft für drei Arzneimittel.² Die Ergebnisse dieser „PICO exercises“ bildeten die Grundlage für die Weiterentwicklung der „Guidance on the Scoping Process“, welche Orientierung für Mitgliedstaaten sowie Assessoren und Co-Assessoren bei der Definition der nationalen PICOs und der anschließenden Konsolidierung geben soll.

In Bezug auf die Endpunkte zeigte sich in den „PICO exercises“ eine große Heterogenität der Endpunktformulierung durch die einzelnen Mitgliedstaaten (Tabelle 1).



Dr. Johanna Seeger arbeitet seit 2021 als Referentin in der Abteilung Arzneimittel des Gemeinsamen Bundesausschusses (G-BA). Seit 2022 ist sie Vertreterin des G-BA in den Subgruppen der HTA-Koordinierungsgruppe für gemeinsame klinische Bewertungen, für gemeinsame wissenschaftliche Beratungen sowie für die Entwicklung methodischer und verfahrenstechnischer Leitfäden. 2009 schloss sie ihr Studium an der Evangelischen Hochschule Berlin als „Bachelor of Nursing“ ab. In den Jahren 2012 bis 2016 studierte sie Pharmazie an der FU Berlin. 2017 erfolgte ihre Approbation zur Apothekerin, 2021 schloss sie ihre Promotion ab.

Zunächst fällt auf, dass die in Deutschland übliche Kategorisierung der Endpunkte in die Kategorien Mortalität, Morbidität, gesundheitsbezogene Lebensqualität und Nebenwirkungen nicht vorgenommen wurde. Zudem unterschieden sich die einzelnen Endpunkte stark hinsichtlich ihres Konkretisierungsgrades. Teilweise wurden konkrete Messinstrumente genannt, teilweise sehr allgemein formuliert: „Any other patient centred outcome measured by patient-reported outcomes measures“. Weiterhin erfolgte in bestimmten Fällen eine Vermischung von Erhebungsmethoden und Endpunkten („Radiological tumor assessment, including overall response rate and duration of response“). Hinsichtlich der Sicherheitsendpunkte zeichnete sich die Möglichkeit der Definition eines indikationsunabhängigen, standardisierten Sets ab.

Die Erfahrungen aus den „PICO exercises“ in Bezug auf die Formulierung von Endpunkten flossen in die „Guidance on Outcomes for Joint Clinical Assessments (JCA)“ ein.³

Guidance on Outcomes for JCA: Konkretisierung der Anforderungen an Endpunkte

Dieses von der Subgruppe für die Entwicklung methodischer und verfahrenstechnischer Leitfäden entwickelte Guidance-Dokument definiert zentrale Konzepte und Anforderungen an Endpunkte im JCA. Zudem werden Vorgaben an die Formulierung von Endpunkten durch die Mitgliedstaaten im Rahmen des Scoping-Verfahrens gemacht. Durch eine einheitliche Formulierung soll die Konsolidierung von Endpunkten unterstützt werden. Beispielsweise empfiehlt die Guidance, auf die Forderung von Effektmaßen zu verzichten. Wenn die Mitgliedstaaten dennoch konkrete Effektmaße oder Messinstrumente zur Beantwortung ihrer nationalen Fragestellung benötigen, wird die Angabe einer Präferenz mit der Formulierung „[Outcome of interest] measured preferably as [insert measure]“ empfohlen.

Ergebnisse der EUnetHTA 21 „PICO exercises“ zu den Arzneimitteln Pluvicto und Pombiliti zum PICO-Parameter „Outcomes“

Lutetium (¹⁷⁷ Lu) Vipivotidtetraacetat (Pluvicto)	Cipaglucosidase alfa (Pombiliti)
Pluvicto in combination with androgen deprivation therapy (ADT) with or without androgen receptor (AR) pathway inhibition is indicated for the treatment of adult patients with progressive prostate-specific membrane antigen (PSMA)-positive castration-resistant prostate cancer (mCRPC) who have been treated with AR pathway inhibition and taxane based chemotherapy.	Pombiliti (cipaglucosidase alfa) is a long-term enzyme replacement therapy used in combination with the enzyme stabiliser miglustat for the treatment of adults with late-onset Pompe disease (acid α -glucosidase [GAA] deficiency).
- Overall survival - Radiological tumor assessment, including overall response rate and duration of response - Progression free survival (radiological, clinical or PSA) by investigator and blinded independent committee review - Symptomatic skeletal event, including time to first skeletal event - Prostate specific antigen levels - Pain measured by a patient-reported outcome measure such as a numeric rating scale or a visual analogue scale - Fatigue - Health-related quality of life, measured preferably by generic and disease specific questionnaires, ie EORTC QLQ C30 plus, if possible, EORTC PR25 or FACT-P, FACT-G - Health status measured preferably by EQ-5D-5L - Any other patient centred outcome measured by patient-reported outcomes measures - Adverse events (total) - Serious adverse events - Severe adverse events (Grade $\geq$ 3) - Discontinuation and interruption due to adverse events - Adverse events of special interest (AESI) - Suspected unexpected serious adverse reaction (SUSAR)	- Overall survival - Ventilator-free survival - Changes in mobility (incl. measurement by 6MWT and documented use of wheelchair) - Changes in respiratory function (incl. measurement by FVC in sitting and upright positions) - Changes in muscle strength (by validated scales) - Changes in motor function (by validated scales, e.g. quick motor function test) - Respiratory symptomatology associated with Pompe disease - Gastrointestinal symptomatology associated with Pompe disease - Quality of life (as assessed using disease-specific (preferably) and/or generic questionnaires) - Health status (measured preferably by the EQ-5D) - Patient-reported outcomes to include R-PACT scale, and any other patient-centered outcome assessed by means of a patient-reported outcome measure - Adverse events (AEs) (incl. hypersensitivity, infusion reactions, immunogenicity) - Serious AEs (SAEs) - Severe AEs - Discontinuation and interruption of treatment due to AEs - Mortality due to AEs

6MWT: 6-minute walking test; AE: Adverse event; EQ-5D: EuroQoL five-dimension scale questionnaire; FVC: Forced vital capacity; R-PACT: Rasch-built Pompe-specific Activity.

Quelle: EUnetHTA 21: D5.4 JCA without HTD submission (PICO exercises); <https://www.eunetha.eu/d5-4/>

Tabelle 1: Die Erprobung des Prozesses, um zu einer möglichst geringen Zahl an PICO zu gelangen, erfolgte in den „PICO exercises“. Dabei wurde eine große Heterogenität der Endpunktformulierung durch die Mitgliedstaaten deutlich.

Weiterhin macht das Guidance-Dokument Vorgaben zum Umgang mit Surrogatendpunkten und kombinierten Endpunkten im Scoping-Verfahren sowie zu den diesbezüglichen Anforderungen an das europäische Dossier sowie den JCA-Bericht. Hinsichtlich der Validität, Reliabilität und Interpretierbarkeit von Messinstrumenten beschreibt die Guidance, welche Angaben im Dossier gemacht und im JCA-Bericht dargelegt werden müssen, um den Mitgliedstaaten die Beurteilung der Eignung eines Messinstrumentes zu ermöglichen. Schließlich wird ein einheitliches Set an Endpunkten der Kategorie Nebenwirkungen definiert, zu denen unabhängig von den eingereichten und konsolidierten PICO-Schemata immer Ergebnisse im Dossier und JCA-Report dargestellt werden müssen.

Wichtig ist, dass im Rahmen des JCA keine Wertungsentscheidungen, beispielsweise bezüglich der Patientenrelevanz bestimmter Endpunkte oder der Validität von Surrogatendpunkten oder Messinstrumenten getroffen werden. Die Beurteilung der (Patienten-)Relevanz von Endpunkten sowie die Schlussfolgerungen aus den im JCA-Bericht dargestellten Ergebnissen und damit verbundene Wertungsentscheidungen liegen in der Verantwortung der Mitgliedstaaten.

Fazit

Auch wenn wenige Wochen vor Beginn der ersten JCA-Verfahren noch einige offene Fragen bestehen, hat die intensive Auseinandersetzung mit den verschiedenen Perspektiven der Mitgliedstaaten im Zuge der vorbereitenden Arbeit der Subgruppen der HTA-Koordinierungsgruppe relevante Fortschritte auf dem Weg zu einem gemeinsamen Verständnis ermöglicht. Die „Guidance on Outcomes for JCA“ leistet einen wesentlichen Beitrag zur Harmonisierung der Anforderungen an Endpunkte. Dies soll die Mehrfachnennung identischer oder fast identischer Endpunkte

im Scoping-Verfahren und eine daraus folgende Duplikation der Darstellung von Ergebnissen im europäischen Dossier und im JCA-Report minimieren.

Welche konkreten Endpunkte in den Bewertungsumfang eingeschlossen werden sowie die Entscheidung darüber, welche Endpunkte für die Beantwortung der nationalen Fragestellung als relevant erachtet werden, liegt in der Verantwortung der Mitgliedstaaten. Der Prozess zur Festlegung des Bewertungsumfanges für einen JCA kann dazu führen, dass für ein zur Beantwortung der deutschen Fragestellung herangezogenes PICO auch Ergebnisse zu von anderen Mitgliedstaaten geforderten Endpunkten im JCA-Bericht dargestellt werden, die vom G-BA jedoch nicht als relevant erachtet werden. Aus der Darstellung von Ergebnissen zu einem Endpunkt im JCA-Report ergibt sich somit keine Verpflichtung, diese Ergebnisse für die nationale Bewertung heranzuziehen.

In Bezug auf das AMNOG-Verfahren ist zudem der Grundsatz der Gleichbehandlung von Produkten, welche in den Geltungsbereich der EU-HTA-Verordnung fallen und solchen, die ausschließlich auf nationaler Ebene bewertet werden, relevant. Das ist insbesondere bedeutsam, da es zunächst indikationsabhängig parallel Verfahren mit und ohne vorgeschalteten JCA geben wird, in denen keine unterschiedlichen Bewertungsmaßstäbe angelegt werden sollten.

Literatur

¹ Regulation (EU) 2021/2282 of the European Parliament and of the Council of 15 December 2021 on health technology assessment and amending Directive 2011/24/EU. <https://go.sn.pub/bomyql>

² EUnetHTA 21: D5.4 JCA without HTD submission (PICO exercises). <https://go.sn.pub/4jvktg>

³ European Commission: Guidance on Outcomes for Joint Clinical Assessments. <https://go.sn.pub/1pvyyz>

Ein überprüfender Zugriff auf die Internetquellen erfolgte am 15. November 2024.

Validierung von Surrogatendpunkten

Prof. Tomasz Burzykowski | Professor für Biostatistik und statistische Bioinformatik am Data Science Institute der Universität Hasselt

Ein Surrogatendpunkt (oder Surrogat-Outcome) soll ein klinisches Ergebnis ersetzen, um neue Behandlungen zu bewerten, wenn er günstiger, einfacher, häufiger oder früher messbar ist. Ein solcher Surrogatendpunkt soll den klinischen Nutzen, Schaden oder das Fehlen eines solchen vorhersagen. Für eine quantitative Bewertung der Evidenzstärke zur Surrogatfähigkeit müssen zwei Punkte nachgewiesen werden: die prognostische Relevanz des Surrogats für das klinische Ergebnis und dass die Behandlungseffekte auf das Surrogat zuverlässig die Effekte auf das klinische Ergebnis vorhersagen. Dieser Prozess der Erbringung der Evidenz wird als Validierung des Surrogats bezeichnet. Zu diesem Zweck wird meist der sogenannte meta-analytische Ansatz verwendet, bei dem Daten aus mehreren randomisierten klinischen Studien analysiert werden.

In den von der Koordinierungsgruppe der Mitgliedstaaten (MS) für die Bewertung von Gesundheitstechnologien (HTA) verabschiedeten „Leitlinien zu Ergebnissen für gemeinsame klinische Bewertungen“ (HTA CG, 2024) heißt es: „Surrogatendpunkte können einen direkten patientenbezogenen Nutzen widerspiegeln oder auch nicht, und ihre klinische Relevanz und Eignung für die gemeinsame klinische Bewertung (Joint Clinical Assessment, JCA) muss von den Mitgliedstaaten berücksichtigt werden.“ Was sind Surrogatendpunkte?

In der Leitlinie heißt es: „Ein Surrogatendpunkt ist ein Ergebnis, das ein relevantes Ergebnis ersetzen soll, das in einer spezifischen klinischen Studie nicht beobachtet werden kann. (...) Von einem Surrogatendpunkt wird erwartet, dass er lediglich den Behandlungseffekt eines Ergebnisses vorhersagt, das in einer klinischen Studie nicht beobachtet wurde.“ Bei der Erörterung der Verwendung von Surrogatendpunkten heißt es in der Leitlinie, dass diese „(...) gegebenenfalls zusätzlich zu patientenrelevanten Endpunkten angefordert werden können. Es sollten jedoch nach Möglichkeit nur Surrogatendpunkte herangezogen werden, wenn ihre Validität zuvor eindeutig nachgewiesen wurde.“ Damit wird das Konzept der „Validität eines Surrogatendpunkts“ vorangebracht.

Die Validität bezieht sich auf „die Stärke der Assoziation zwischen dem Surrogatendpunkt und dem relevanten Endpunkt sowie auf die Assoziation zwischen den Behandlungseffekten auf den Surrogatendpunkt und den relevanten Endpunkt“. Da die Leitlinie besagt, dass Entwickler von Gesundheitstechnologien „die Stärke des Zusammenhangs nachweisen“ sollten, wenn ein Surrogatendpunkt verwendet wird. Dieses Kriterium entspricht dem sogenannten meta-analytischen Rahmenwerk zur Validierung von Surrogatendpunkten (Buyse et al., 2000).

Tatsächlich wird dieser Ansatz ausdrücklich in der Leit-

linie erwähnt: „Es gibt mehrere andere nützliche Ansätze zur Validierung von Surrogatendpunkten. Im Allgemeinen basieren diese Methoden auf einem meta-analytischen Ansatz.“ Es ist erwähnenswert, dass die Leitlinie auch ein Validitätsmaß nennt, das spezifisch für den meta-analytischen Ansatz ist, nämlich den Surrogat-Schwelleneffekt (Surrogate Threshold Effect, STE): „Das Konzept der Surrogat-Schwelleneffekts ist hilfreich für die Entscheidungsfindung, da sie den Mindesteffekt in Bezug auf den Surrogatendpunkt darstellt, der erforderlich ist, um zu dem Schluss zu kommen, dass mit hoher Sicherheit auch ein Effekt auf den patientenrelevanten Endpunkt vorliegt.“

Im Folgenden werden die in der HTA-CG-Leitlinie erwähnten Konzepte und Methoden näher erläutert.



Tomasz Burzykowski ist Professor für Biostatistik und statistische Bioinformatik am Data Science Institute der Universität Hasselt (Belgien). Außerdem ist er Vizepräsident für Forschung beim International Drug Development Institute (IDDI), einem Auftragsforschungsinstitut, das Dienstleistungen für klinische Studien anbietet. Er hatte Gastprofessuren am Karolinska-Institut (Schweden), an der Technischen Universität Warschau (Polen) und an der Medizinischen Universität Białystok (Polen). In den Jahren 2023-2024 war er Präsident der Internationalen Gesellschaft für klinische Biostatistik.

Im meta-analytischen Ansatz (Buyse et al., 2000) werden Daten aus historischen klinischen Studien verwendet, um die Stärke der Assoziation zwischen den Behandlungseffekten auf den relevanten Endpunkt und auf den Surrogatendpunkt zu bewerten. Dies wird als Assoziation auf Studienebene (Trial-Level-Assoziation) bezeichnet. Die Stärke der Assoziation wird mit dem Korrelationskoeffizienten R oder dem Bestimmtheitsmaß R^2 gemessen, die sich aus einem linearen Regressionsmodell ergeben, das an die geschätzten Behandlungseffekte angepasst wurde.

Je stärker die Assoziation, desto präziser ist die Vorhersage des Behandlungseffekts auf den relevanten Endpunkt basierend auf dem Effekt auf den Surrogatendpunkt. Die Konzentration auf die Stärke der Assoziation ergibt sich aus der Tatsache, dass – wie bereits erwähnt – die Vorhersage das Hauptziel der beabsichtigten Verwendung eines Surrogats ist.

Darüber hinaus werden beim meta-analytischen Ansatz Daten aus historischen klinischen Studien verwendet, um die Stärke der Assoziation zwischen dem relevanten Ergebnis und dem Surrogatendpunkt zu bewerten. Dies wird als Assoziation auf individueller Ebene (Individual-Level-Assoziation) bezeichnet. Sie wird in der Regel anhand eines Assoziationsmaßes (z. B. Korrelationskoeffizient, Odds Ratio) erfasst, das auf einem bivariaten Modell basiert, das an die individuellen Patientendaten angepasst wird und mögliche Behandlungseffekte berücksichtigt. Je stärker die Assoziation ist, desto genauer ist die Vorhersage des relevanten Ergebnisses auf der Grundlage des für einen Patienten beobachteten Surrogatwertes.

Ein Surrogat, das auf individueller Ebene gültig ist, kann nützlich sein, um z.B. die Behandlungsbedürfnisse eines Patienten individuell anzupassen (Buyse et al., 2022). Für ein solches Surrogat folgt jedoch nicht unmittelbar, dass behandlungsbedingte Veränderungen des Surrogats eng

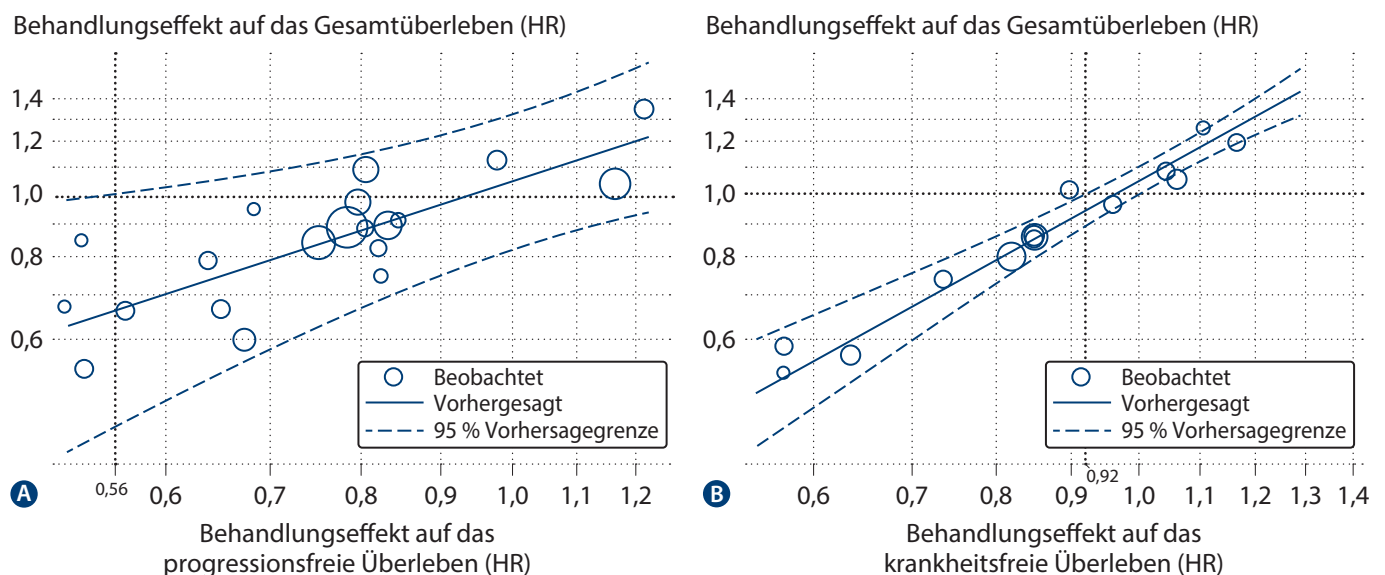
mit den entsprechenden Veränderungen am klinischen Endpunkt verbunden sind, d. h. dass das Surrogat auf Studienebene gültig ist. Im letzteren Fall ist eine direkte Bewertung der Stärke der Assoziation auf Studienebene erforderlich (Buyse, 2016; Buyse et al., 2022).

Um die Konzepte und Herausforderungen im Zusammenhang mit der Validierung von Surrogatendpunkten zu veranschaulichen, betrachten wir drei Beispiele für eine solche Validierung anhand eines meta-analytischen Ansatz-

zes. Insbesondere werden zwei Meta-Analysen von randomisierten klinischen Studien zum Magenkarzinom und eine Meta-Analyse von randomisierten Studien zum kolorektalen Karzinom diskutiert.

Die erste Meta-Analyse von Studien zum Magenkarzinom schloss Patienten mit fortgeschrittenem Karzinom ein (Paoletti et al., 2013). Ziel war die Bewertung des progressionsfreien Überlebens (PFS) als Surrogat für das Gesamtüberleben (OS). Es lagen Daten von 4.069 Patienten aus 20

Assoziation auf Studienebene bei fortgeschrittenem (Abb. 1a) und reseziertem (Abb. 1b) Magenkarzinom



Quelle: 1a: Paoletti et al. (2013); 1b: Oba et al. (2013)

Abbildung 1: Panel A: Assoziation auf Studienebene für das Beispiel des fortgeschrittenen Magenkarzinoms. Panel B: Assoziation auf Studienebene für das Beispiel des resezierten Magenkarzinoms. Die Kreise (deren Größe proportional zum Stichprobenumfang der Studie ist) stellen die Paare der geschätzten Behandlungseffekte auf das Surrogat und das Gesamtüberleben dar. Die durchgezogene gerade Linie ist die lineare Regression, die den vorhergesagten Behandlungseffekt auf das Gesamtüberleben angibt. Die gestrichelte vertikale Linie zeigt den Surrogat-Schwelleneffekt an. Es ist zu beachten, dass beide Achsen auf einer logarithmischen Skala liegen.

relevanten randomisierten Studien mit dokumentiertem OS und PFS vor. Die Studien untersuchten den Zusatz von experimentellen Chemotherapeutika zu bereits vorhandenen Kontroll- oder Standardtherapien. Der Spearman-Rangkorrelationskoeffizient, der die Assoziation zwischen OS und PFS auf individueller Ebene quantifiziert, betrug 0,853 (95 % KI [0,852, 0,854]).

Wie in der HTA-CG-Leitlinie [HTA CG, 2024] erwähnt, „gibt es keinen allgemein akzeptierten Schwellenwert für die Feststellung ausreichender Korrelationen zwischen dem Surrogat und dem patientenzentrierten Ergebnis auf Studienebene (d.h. Korrelation der Wirkungen) und auf Patientenebene (d.h. Korrelation der Ergebnisse). Eine Korrelation von mindestens 0,85 wird jedoch als „hoch“ beschrieben und kann als Kriterium für die Validierung von Surrogat-Endpunkten verwendet werden.“ Geht man von einem Schwellenwert von 0,85 aus, ist das PFS ein gültiges Surrogat auf individueller Ebene für das OS.

Die Assoziation auf Studienebene zwischen den Behandlungseffekten auf das PFS und OS wird in Abbildung 1A dargestellt. Das lineare Regressionsmodell war: $\ln(\text{HROS}) = 0,042 + 0,779 \times \ln(\text{HRPFS})$, wobei $\ln(\text{HROS})$ und $\ln(\text{HRPFS})$ die Logarithmen des Hazard Ratios (HR) für das OS bzw. PFS darstellen. Nach Bereinigung des Schätzfehlers wurde R^2 auf 0,61 (95 % KI [0,04, 1,00]) geschätzt. Das breite Konfidenzintervall spiegelt die Unsicherheit dieser Schätzung wider, was teilweise auf die kleinen Stichprobengrößen einiger der in die Meta-Analyse einbezogenen Studien zurückzuführen ist. Der Korrelationskoeffizient (R), der R^2 entspricht, beträgt 0,781. Bei Verwendung des Schwellenwerts von 0,85 ist PFS kein gültiges Surrogat auf Studienebene für das OS.

Die Ergebnisse wurden extern validiert, indem 12 Studien verwendet wurden, die nicht in die Meta-Analyse einbezogen waren und für die die Behandlungseffekte aus

Berichten extrahiert wurden, die nach Abschluss der Meta-Analyse veröffentlicht wurden. Abbildung 2A zeigt die gleiche Regressionslinie wie in Abbildung 1A. Die beobachteten Behandlungseffekte auf das Überleben (HROS) werden für diese 12 Studien dargestellt, ebenso wie die aus den Behandlungseffekten des Surrogats (HRPFS) in diesen Studien vorhergesagten Behandlungseffekte auf das OS, einschließlich ihrer 95 %-Vorhersageintervalle. Wie in Abbildung 2A zu sehen ist, waren die Vorhersageintervalle breit und schlossen für alle Studien den Wert von 1 (kein Behandlungseffekt auf das OS) ein, was bedeutet, dass die beobachteten Effekte auf das PFS in keiner der 12 Studien einen signifikanten Effekt auf das OS hätten vorhersagen können. Dennoch zeigten drei der 12 Studien einen statistisch signifikanten Effekt der Behandlung auf das Überleben (Paoletti et al., 2013). Insgesamt scheint das PFS kein gültiges Surrogat auf Studienebene für das OS bei fortgeschrittenem Magenkarzinom zu sein.

Die zweite Meta-Analyse von Studien zum Magenkarzinom schloss Patienten mit reseziertem Karzinom ein (Oba et al., 2013). Ziel war die Bewertung des krankheitsfreien Überlebens (DFS) als Surrogat für das OS. Es lagen Daten von 3.371 Patienten aus 14 Studien mit dokumentiertem OS und DFS vor. In den Studien wurde eine adjuvante Chemotherapie mit einer alleinigen Operation verglichen. Basierend auf individuellen Patientendaten wurde der Spearman-Rangkorrelationskoeffizient, der die Assoziation zwischen OS und DFS auf individueller Ebene quantifiziert, auf 0,974 (95 % KI [0,971, 0,976]) geschätzt. Bei Verwendung des Schwellenwerts von 0,85 ist DFS ein gültiges Surrogat auf individueller Ebene.

Auf Studienebene gab es ebenfalls eine enge Assoziation zwischen den Behandlungseffekten auf das DFS und das OS (siehe Abbildung 1B). Das lineare Regressionsmodell war: $\ln(\text{HROS}) = 0,047 + 1,239 \times \ln(\text{HRDFS})$.

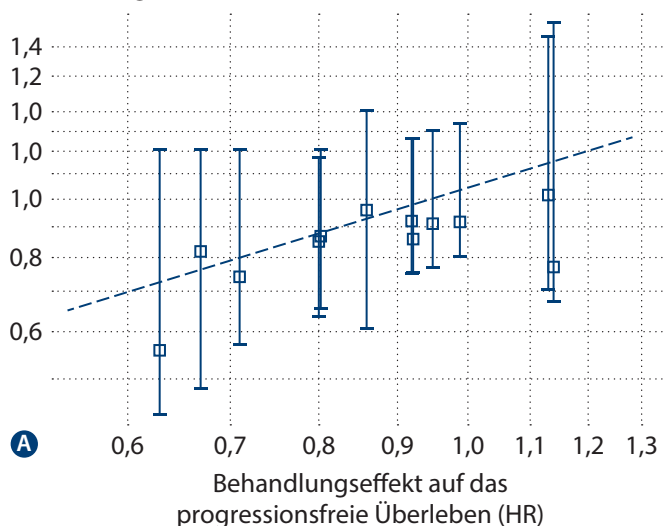
Nach Bereinigung des Schätzfehlers wurde R^2 auf nahezu 1 (95 % KI [0.999, 1.000]) geschätzt, wobei der entsprechende Wert des Korrelationskoeffizienten R ebenfalls nahe bei 1 lag. (Die erzielten numerischen Ergebnisse sind jedoch mit Vorsicht zu genießen, da der geschätzte Wert von R^2 sehr nahe an der Obergrenze von 1 liegt.) Man

könnte also schlussfolgern, dass DFS ein gültiges Surrogat auf Studienebene für das OS ist.

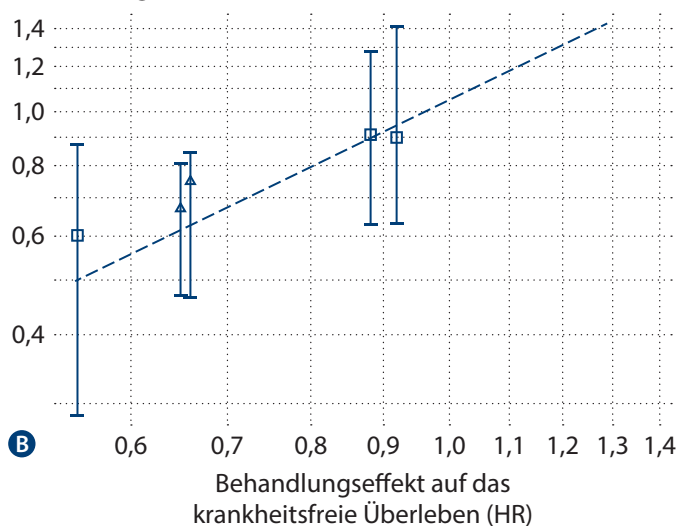
Die Ergebnisse wurden extern validiert, indem fünf Studien verwendet wurden, die nicht in die Meta-Analyse einbezogen waren. Drei davon stammten aus veröffentlichten Berichten, und zwei basierten auf individuellen Patienten-

Validierung der Surrogate progressionsfreies und krankheitsfreies Überleben beim fortgeschrittenen (Abb. 2a) bzw. resezierten (Abb. 2b) Magenkarzinom

Behandlungseffekt auf das Gesamtüberleben (HR)



Behandlungseffekt auf das Gesamtüberleben (HR)



Quelle: 2a: Paoletti et al. (2013); 2b: Oba et al. (2013)

Abbildung 2. Panel A: Validierungsdaten für das Beispiel des fortgeschrittenen Magenkarzinoms. Panel B: Validierungsdaten für das Beispiel des resezierten Magenkarzinoms. Die gestrichelte gerade Linie ist die lineare Regression, die den vorhergesagten Behandlungseffekt auf das Gesamtüberleben angibt. Die Kästen sind die geschätzten Behandlungseffekte auf das Gesamtüberleben für die Validierungsstudien. Die vertikalen Intervalle sind die 95 %-Vorhersageintervalle für den Behandlungseffekt auf das Gesamtüberleben. Es ist zu beachten, dass beide Achsen auf einer logarithmischen Skala liegen.

daten, die nach Abschluss der Meta-Analyse verfügbar wurden. Abbildung 2B zeigt die gleiche Regressionslinie wie in Abbildung 1B. Die beobachteten Behandlungseffekte auf das Überleben (HROS) werden für diese fünf Studien dargestellt, ebenso wie die aus den Behandlungseffekten des Surrogats (HRDFS) in diesen Studien vorhergesagten Behandlungseffekte auf das Überleben, einschließlich ihrer 95 %-Vorhersageintervalle. Es besteht eine sehr gute Übereinstimmung zwischen den beobachteten und vorhergesagten Behandlungseffekten, und in den drei Studien, in denen die Vorhersagegrenzen von HROS den Wert 1 ausschlossen, erreichten die beobachteten Effekte auf das Überleben tatsächlich statistische Signifikanz (Paoletti et al., 2013). Insgesamt ist das DFS ein gültiges Surrogat sowohl auf individueller als auch auf Studienebene für das OS bei resezierbarem Magenkarzinom.

Abbildung 1 veranschaulicht die zuvor erwähnte Verbindung zwischen der Stärke der Assoziation auf Studienebene und der Genauigkeit der Vorhersage des Behandlungseffekts auf das klinische Ergebnis. In Abbildung 1A sind die Vorhersagegrenzen für die Auswirkung auf das OS breit; in diesem Fall war der Wert von $R^2=0,61$ (mit $R=0,781$) deutlich kleiner als 1. Andererseits sind die Vorhersagegrenzen in Abbildung 1B eng; in diesem Fall waren R^2 (und R) sehr nahe bei 1. Je höher also R^2 oder R , desto präziser ist die Vorhersage. Das ist der Grund, warum im meta-analytischen Ansatz die Validität eines Surrogats auf Studienebene durch Quantifizierung der Stärke der Assoziation auf Studienebene beurteilt wird. Ein ähnliches Argument gilt für die Assoziation auf individueller Ebene.

Die in Abbildung 1 dargestellten Vorhersagegrenzen veranschaulichen auch das Konzept des Surrogat-Schwelleneffekts (STE). Der STE ist der minimale Behandlungseffekt auf das Surrogat, der eine Vorhersage eines (von Null verschiedenen) Effekts auf das klinische Ergebnis ermög-

licht. Praktisch ausgedrückt ist der STE für die betrachteten Beispiele der Wert der HR für das Surrogat, bei dem die obere Vorhersagegrenze, die aus der Regressionslinie abgeleitet wurde, gleich 1 wird. In Abbildung 1 ist der Wert durch die gestrichelte vertikale Linie dargestellt. Für das Beispiel des fortgeschrittenen Magenkarzinoms war der STE gleich 0,56. Das bedeutet, dass ein HR-Wert für das PFS, der kleiner ist als 0,56, mit 95-prozentiger Sicherheit einen HR-Wert für das OS vorhersagen würde, der kleiner als 1 ist. Der STE von 0,56 impliziert eine deutliche Verringerung des Risikos um 44 %. Betrachtet man die Bandbreite der HR-Werte für das PFS, die in den Studien in Abbildung 1A beobachtet wurden, so stellt $HR=0,56$ einen ziemlich extremen Behandlungseffekt dar. Das ist eine Folge der breiten Vorhersagegrenzen in Abbildung 1A. Somit scheint das PFS kein nützliches Surrogat zu sein, da man einen großen Effekt auf das PFS benötigen würde, um von einem Effekt auf das OS sprechen zu können. Diese Schlussfolgerung stimmt mit derjenigen überein, die zuvor aus dem Wert von R^2 abgeleitet wurde.

Für das Beispiel des resezierten Magenkarzinoms betrug der STE hingegen 0,92 (siehe Abbildung 1B). Dies ist ein kleiner Behandlungseffekt, der eine Verringerung des Risikos um nur 8 % impliziert. Dieser geringe STE ist eine Folge der engen Vorhersagegrenzen in Abbildung 1B. Somit ist das DFS ein nützliches Surrogat, denn man müsste einen kleinen Effekt auf das DFS haben, um einen Effekt auf das OS behaupten zu können. Diese Schlussfolgerung deckt sich mit derjenigen, die sich aus dem Wert von R^2 ergibt.

Die beiden Beispiele zum Magenkarzinom veranschaulichen, wie eine quantitative Bewertung die Verwendung von Surrogatendpunkten beeinflussen kann; obwohl diese beiden Situationen relativ ähnlich erscheinen, kommen wir zu unterschiedlichen Schlussfolgerungen hinsichtlich der Verwendung von PFS und DFS als Surrogat für das OS. Tat-

sächlich könnte man argumentieren, dass das DFS als angemessenes Surrogat für das OS in der adjuvanten Therapie verwendet werden kann, während das PFS nicht zuverlässig als Surrogat bei fortgeschrittener Krankheit verwendet werden kann.

Die Meta-Analyse von Studien für Patienten mit fortgeschrittenem kolorektalem Karzinom wurde durchgeführt, um das progressionsfreie Überleben (PFS) als Surrogat für das OS zu bewerten [Buyse et al., 2007]. Es lagen Daten von 4.352 Patienten aus 13 Studien vor. Die Studien umfassten 10 historische Studien, in denen Fluorouracil (FU) in Kombination mit Leucovorin entweder mit FU allein (1.744 Patienten) oder mit Raltitrexed (1.345 Patienten) verglichen wurde, sowie drei Validierungsstudien, in denen FU + Leucovorin mit oder ohne Irinotecan oder Oxaliplatin (1.263 Patienten) verglichen wurde. Der Spearman-Rangkorrelationskoeffizient, der die Assoziation zwischen OS und PFS auf individueller Ebene quantifiziert, betrug 0,82 (95 % KI [0,82, 0,83]). Streng genommen liegt der Wert unter 0,85, was darauf hindeuten könnte, dass das PFS kein gültiges Surrogat auf individueller Ebene ist.

Auf Studienebene gab es eine enge Assoziation zwischen den Behandlungseffekten auf das PFS und das OS. Das lineare Regressionsmodell war:

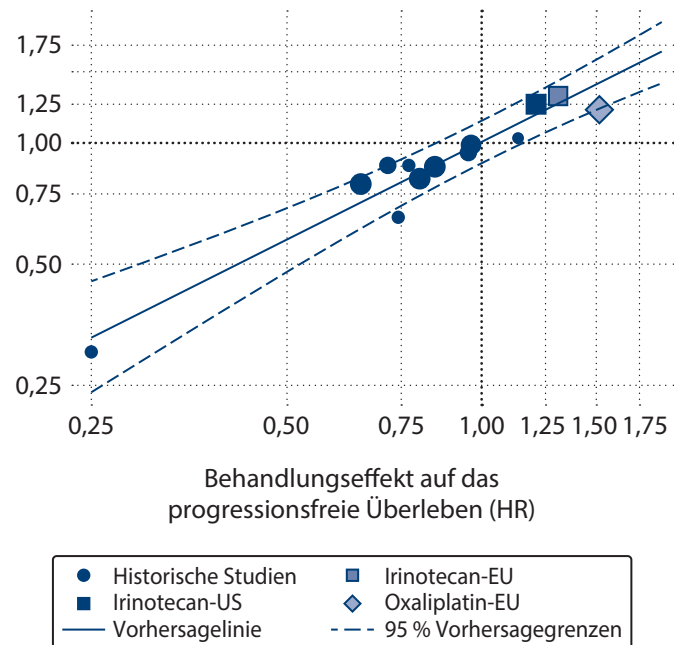
$$\ln(\text{HROS}) = 0,003 + 0,81 \times \ln(\text{HRPFS})$$

Der Wert von R^2 wurde auf 0,98 (95 % KI [0,88, 1,08]) geschätzt, mit einem entsprechenden R-Wert von 0,99. Bei Verwendung des Schwellenwerts von 0,85 ist PFS ein gültiges Surrogat auf Studienebene für das OS.

Die Ergebnisse wurden extern validiert, indem drei (neure) Studien verwendet wurden, die nicht in die Meta-Analyse einbezogen waren. Die vorhergesagten Wirkungen stimmten in den Studien, in denen Irinotecan getestet wurde, sehr gut mit den beobachteten Wirkungen überein, weniger gut jedoch in der Studie, in der Oxaliplatin getes-

Assoziation von progressionsfreiem Überleben beim fortgeschrittenen kolorektalen Karzinom

Behandlungseffekt (HR) auf das Gesamtüberleben



Quelle: Buyse et al. (2007)

Abbildung 3. Assoziation auf Studienebene für das Beispiel des fortgeschrittenen kolorektalen Karzinoms. Die Kreise (deren Größe proportional zum Stichprobenumfang der Studie ist) stellen die Paare der geschätzten Behandlungseffekte auf das Surrogat und das Gesamtüberleben dar. Die durchgezogene gerade Linie ist die lineare Regression, die den vorhergesagten Behandlungseffekt auf das Gesamtüberleben angibt. Die Quadrate entsprechen den Ergebnissen der beiden Validierungsstudien, in denen der Zusatz von Irinotecan untersucht wurde. Die Raute entspricht den Ergebnissen der Validierungsstudie, in der der Zusatz von Oxaliplatin untersucht wurde. Es ist zu beachten, dass beide Achsen auf einer logarithmischen Skala liegen.

tet wurde, wo die vorhergesagte Wirkung die beobachtete Wirkung überschätzte (siehe Abbildung 3). Der Unterschied könnte auf die Wirkung von Zweitlinientherapien und Crossovers zurückzuführen sein, die im Vergleich zu den historischen Studien in größerem Umfang verfügbar waren. In diesem Zusammenhang ist anzumerken, dass das PFS unter der Erstlinientherapie gemessen wird, während das OS unter Folgetherapien gemessen wird, die den Effekt der Erstlinientherapie auf das klinische Ergebnis verzerren können. Dies könnte erklären, warum das PFS in einer Ära marginal wirksamer FU-basierter Therapien (Buyse et al., 2007) ein akzeptables Surrogat für das OS bei fortgeschrittenem kolorektalem Karzinom war, während es heute ein viel weniger überzeugendes Surrogat ist, da Patienten mit fortschreitender Erkrankung viele aktive Therapielinien erhalten können, die ihr Überleben weiter beeinflussen können (Shi et al., 2014).

Das Beispiel des kolorektalen Karzinoms zeigt somit, dass sich die Aussagekraft eines Surrogats im Laufe der Zeit ändern kann. Zudem zeigt der Vergleich mit dem Beispiel des fortgeschrittenen Magenkarzinoms, dass die Gültigkeit desselben Surrogats (PFS) bei verschiedenen Krankheiten unterschiedlich sein kann. Daraus folgt, dass man die Gültigkeit eines Surrogats nicht einfach aus einer Validierungsübung ableiten kann, die bei einer anderen Krankheit durchgeführt wurde (Buyse, 2016). Eine damit verbundene und vielleicht die schwierigste Frage ist, ob ein Surrogat, das für eine bestimmte Medikamentenklasse evaluiert wurde, auch für andere Klassen von Arzneimitteln gültig ist.

In den meisten Fällen ist eine Kombination aus biologischen Überlegungen und statistischem Nachweis erforderlich, um ein Surrogat außerhalb der Bedingungen zu verwenden, unter denen es ursprünglich evaluiert wurde. Statistische Beweise, die für die Verwendung des Surrogats

sprechen, sind am hilfreichsten für neue Medikamente, die den in den Bewertungsdatensätzen verwendeten Arzneimitteln ähnlich sind. Bei neuen Medikamenten mit einem wesentlich anderen Wirkmechanismus bleibt die Frage offen, ob das Surrogat zuverlässig verwendet werden kann, was möglicherweise eine weitere prospektive Bewertung notwendig macht.

Aus praktischer Sicht erfordert der meta-analytische Ansatz zur Validierung von Surrogatendpunkten die Verfügbarkeit von individuellen Patientendaten aus mehreren randomisierten klinischen Studien, in denen Informationen über das Surrogat und die klinischen Ergebnisse verfügbar sind. Trotz bestehender Initiativen zur gemeinsamen Nutzung von Daten stellt die Beschaffung solcher Daten immer noch eine Herausforderung dar. Und selbst wenn der Zugang zu solchen Daten möglich ist, ist nicht gewährleistet, dass die Daten Informationen über beispielsweise von Patienten berichtete Ergebnisse (wie Messungen der Lebensqualität) enthalten, die für das HTA von Interesse sein könnten. Diese praktischen Probleme in Verbindung mit den bereits erwähnten konzeptionellen Fragen machen die Validierung möglicher Surrogatendpunkte immer noch zu einer anspruchsvollen Aufgabe.

Literatur

¹ Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H (2000): The validation of surrogate endpoints in meta-analyses of randomized experiments. (Die Validierung von Surrogat-Endpunkten in Meta-Analysen randomisierter Experimente.) *Biostatistics* 1: 49-68.

² Buyse M, Burzykowski T, Carroll K, Michiels S, Sargent D, Miller LL, Elfring GL, Pignon JP, Piedbois P (2007): Progression-free survival is a surrogate for survival in advanced colorectal cancer. (Progressionsfreies Überleben als Surrogat für das Gesamtüberleben bei fortgeschrittenem kolorektalem Karzinom.) *Journal of Clinical Oncology* 25: 5218-5224.

³ Buyse M (2016) Introduction. In: *Applied Surrogate Endpoint Evaluation*

Methods with SAS and R (Angewandte Methoden der Surrogatendpunktevaluation mit SAS und R) (Alonso A, Bigirimurame T, Burzykowski T, Buyse M, Molenberghs G, Muchene L, Perualila NJ, Shkedy Z, Van der Elst W), Chapman & Hall/CRC (ISBN 978-1-4822-4936-1).

⁴ Buyse M, Saad ED, Burzykowski T, Regan MM, Sweeney CS (2022): Surrogacy beyond prognosis: the importance of „trial level“ surrogacy. (Surrogatparameter jenseits der Prognose: Die Bedeutung der Surrogatvalidität auf Studienebene.) *The Oncologist* 5: 266-271.

⁵ HTA CG (2024) https://health.ec.europa.eu/publications/guidance-outcomes-joint-clinical-assessments_en.

⁶ Oba K, Paoletti X, Alberts S, Bang YJ, Benedetti J, Bleiberg H, Catalano P, Lordick F, Michiels S, Morita S, Ohashi Y, Pignon JP, Rougier P, Sasako M, Sakamoto J, Sargent D, Shitara K, Van Cutsem E, Buyse M, Burzykowski T im Namen der GASTRIC-Gruppe (2013): Disease-free survival as a surrogate for overall survival in adjuvant trials of gastric cancer: a meta-analysis. (Krankheitsfreies Überleben als Surrogat für das Gesamtüberleben in adjuvanten Studien zum Magenkarzinom: Eine Meta-Analyse.) *Journal of the National Cancer Institute* 5: 1600-1607.

⁷ Paoletti X, Oba K, Bang YJ, Bleiberg H, Boku N, Bouche O, Catalano P, Fuse N, Michiels S, Moehler M, Morita S, Ohashi Y, Ohtsu A, Roth A, Rougier P, Sakamoto J, Sargent D, Sasako M, Shitara K, Thuss-Patience E, Van Cutsem E, Burzykowski T, Buyse M im Namen der GASTRIC-Gruppe (2013). Progression-free survival as a surrogate for overall survival in patients with advanced/recurrent gastric cancer: a meta-analysis. (Progressionsfreies Überleben als Surrogat für das Gesamtüberleben bei Patienten mit fortgeschrittenem/rezidiertem Magenkarzinom: Eine Meta-Analyse.) *Journal of the National Cancer Institute* 5: 1608-1612.

⁸ Shi Q, de Gramont A, Grothey A, Zalcborg J, Chibaudel B, Schmoll H-J, Seymour MT, Adams R, Saltz L, Goldberg RM, Punt CJA, Douillard J-Y, Hecht JR, Hurwitz H, Diaz-Rubio E, Prosser R, Tebbutt NC, Fuchs C, Souglakos J, Falcone A, Tournigand C, Kabbinavar FF, Heinemann V, Van Cutsem E, Bokemeyer C, Buyse M, Sargent DJ im Namen der Gruppe für Analyse und Forschung zu Krebserkrankungen des Verdauungssystems (ARCAD-Gruppe) (2014). Individual patient data analysis of progression-free versus overall survival as a first-line endpoint for metastatic colorectal cancer in modern randomized trials: Findings from 16,700 patients from the ARCAD database. (Individuelle Patientendatenanalyse von progressionsfreiem versus Gesamtüberleben als Erstlinienendpunkt für metastasierten kolorektalen Karzinom in modernen randomisierten Studien: Ergebnisse von 16.700 Patienten aus der ARCAD-Datenbank.) *Journal of Clinical Oncology* 33: 22-28.

Was ist ein patientenrelevanter Zusatznutzen? Gesucht wird ein strukturierter Diskurs für ein Konsensverfahren

Von Florian Staeck

Auch nach fast 15 Jahren früher Nutzenbewertung seit Einführung durch das AMNOG steht unter den maßgeblichen Akteuren ein voller Konsens darüber aus, wie ein patientenrelevanter Zusatznutzen im Einzelfall definiert werden kann. So sehr der Paradigmenwechsel grundsätzlich begrüßt wird, dass mit dem AMNOG-Verfahren der höchsten jeweils verfügbaren Evidenz ein hoher Stellenwert eingeräumt wird, ist die Festlegung dieser Evidenz insbesondere in Zusammenhang mit Surrogatparametern noch immer wieder ein Streitpunkt unter den Stakeholdern. Insbesondere dann, wenn versäumt wurde, klinisch relevante Endpunkte zu untersuchen, obwohl dazu die Möglichkeit gegeben war.

Dies wurde im Verlauf der 20. Tagung der Interdisziplinären Plattform zur Nutzenbewertung deutlich, die am 27./28. September 2024 in Berlin unter dem Titel „Welche Zielgrößen sind relevant für Patienten“ abgehalten wurde. Dabei wurde mehrfach betont, wie wünschenswert es wäre, wenn evidenzbasierte Verfahren existierten, durch die Patientenrelevanz im Konsens bestimmt werden könnte.

Allerdings wird durch die Rechtswissenschaft kaum eine Klärung der damit verbundenen Fragen zu erwarten sein, hieß es bei der Tagung. Denn Patientenrelevanz sei ein unbestimmter Rechtsbegriff, der hochgradig auslegungsbedürftig und von geringer inhaltlicher Bestimmtheit sei. Im AMNOG-System sei dem Gemeinsamen Bundesausschuss (G-BA) der relevante Beurteilungsspielraum zugesprochen worden. Nutzenbewertungen sind insofern Akte der untergesetzlichen Normsetzung. Und diese seien nur sehr eingeschränkt der gerichtlichen Kontrolle zugänglich. Diese beschränke sich beispielsweise auf die Frage, ob alle relevanten Zuständigkeits-, Verfahrens- und Formvorschriften eingehalten oder ob sachfremde Erwägungen stattgefunden haben.

Nicht vertretbar wäre eine Bewertung durch den G-BA nur dann, wenn sie beispielsweise dem allgemein anerkannten Stand der HTA-Wissenschaft widerspricht. Dagegen dürften Gerichte nicht ihre eigenen Bewertungen an die Stelle des G-BA setzen. Bisher sei, soweit ersichtlich, in einem Gerichtsverfahren im Zuge einer Klage gegen einen Nutzenbewertungsbeschluss des Bundesausschusses auch nie das Ziel verfolgt worden, einen Endpunkt als patientenrelevant zu erschließen. Mit anderen Worten: Die Jurisprudenz spiele den Ball in den strittigen Fragen an die Wissenschaft und die Akteure im AMNOG-Kosmos zurück.

Gleiches gelte für die Politik, wurde bei der Tagung deutlich. Das deutsche Gesundheitswesen zeichne sich dadurch aus, dass der Gesetzgeber die zentralen Diskurse über medizinischen Fortschritt an die Selbstverwaltung auslagert und sich auf die Setzung eines Bewertungsrahmens – wie beispielsweise im Zuge der Arzneimittelnutzenbewertungs-Verordnung – beschränkt. Dies gelte auch für die Streitfrage, wie ein patientenrelevanter Zusatznutzen festzulegen sei. Eine Intervention des Gesetzgebers sei allenfalls dann denkbar, wenn ersichtlich wird, dass medizinischer Fortschritt nicht mehr bei den GKV-Patienten ankomme, hieß es.

Chronifizierung von Erkrankungen besser abbilden

Würde die Selbstverwaltung selbst nicht mehr zu guten Lösungen gelangen, wäre der Gesetz- oder Ordnungsgeber allein schon vom Volumen der Debatte überfordert. Zudem zeige das Beispiel Großbritannien die Probleme, die mit einer politisierten Entscheidungskultur im Gesundheitswesen einhergingen, hieß es.

In Deutschland ergebe sich mitunter die Situation, dass Surrogatparameter, die im Rahmen der Arzneimittelzulassung akzeptiert wurden, nicht im gleichen Maße im Rahmen der frühen Nutzenbewertung anerkannt werden, hieß

es zur Erläuterung. Allerdings wird die Aufgabe, gemeinsame Standards bei der Bewertung von Surrogatparametern zu entwickeln, allein schon vor dem Hintergrund des medizinischen Fortschritts nicht einfacher – im Gegenteil. Dies gelte insbesondere vor dem Hintergrund, dass ehemals häufig tödlich verlaufende Erkrankungen nunmehr mitunter einen chronischen Verlauf nehmen. Einzelne Teilnehmer plädierten daher dafür, die HTA-Kriterien anzupassen, um zu einer besseren Abbildung der Chronifizierung von Erkrankungen zu kommen.

Was auf jeden Fall vermieden werden müsse, sei, dass es zu Herausbildungen von „Parallelwelten“ in der Bewertung von Arzneimitteln im jeweiligen Fachgebiet komme, getrieben beispielsweise durch die Chronifizierung onkologischer Erkrankungen. An Stelle beispielsweise des Endpunkts Overall Survival (OS) müssten daher – je nach Fachgebiet – zusammengesetzte Endpunkte treten, die nicht nur den Krankheitsprogress messen, sondern auch die Lebensqualität. Teilnehmer stellten die Frage, inwiefern eine indikationsbezogene Gewichtung von Endpunkten hier eine praktikable Strategie sein könnte.

Dies wurde insbesondere von Teilnehmern mit dem Argument unterstrichen, starr festgelegte Patientenendpunkte seien insbesondere in Fällen, in denen ein hoch individuelles Therapieansprechen zu beobachten sei, zu hinterfragen. Ein „one size fits all“ könne es nicht geben, denn Patientenrelevanz sei keine statische Größe, wurde argumentiert. Dies gelte zumal für das Fachgebiet der Psychiatrie. Denn dort seien Patient Reported Outcomes (PRO) zentrale Outcomes. Nur sie könnten Auskunft darüber geben, wie ein Betroffener seine Erkrankung erlebt und ob dadurch die therapeutische Intervention ein patientenrelevanter Nutzen erkennbar wird.

Hierbei gelte es zu beachten, dass erkrankungsbedingte Funktionseinschränkungen von Patienten extrem unter-

schiedlich zurückgespiegelt würden. In der Psychiatrie als einem Fachgebiet, in dem es nur ausnahmsweise Biomarker gibt, sind PROs die zentralen Outcomes. Mortalität eigne sich in der Regel nicht zu Erfassung der Therapieresultate. Valide Surrogatmarker seien hier in vielen Fällen, ob der Patient einer bezahlten Arbeit nachgeht und ob er in einer stabilen Partnerschaft lebt. Als patientenzentriertes Outcome sollte daher ein rückfallfreies und funktionsfähiges Langzeitüberleben der Patienten angegeben werden.

Bei der Tagung wurden unterschiedliche Standpunkte dazu deutlich, warum in den vergangenen Jahren in Deutschland keine oder kaum Fortschritte bei der Interpretation von Surrogatendpunkten erzielt wurden. Dies habe auch mit dem Fehlen einer strukturierten Debatte zu tun, hieß es. Aus Sicht des IQWiG ist es nur in wenigen Fällen bisher zu einer zufriedenstellenden Validierung eines Surrogatendpunktes gekommen. Validierungen würden überhaupt nur sehr selten vorgelegt. Zudem lägen für rund zwei Drittel der Fragestellungen bei Nutzenbewertungen zu chronischen Erkrankungen in klinischen Studien keine geeigneten Daten vor – die Datenlage sei hier seit Jahren unverändert schlecht, wurde beklagt. Inzwischen seien zwar viele generische und krankheitsspezifische Scores etabliert worden. Nur müsse dieses Instrumentarium auch adäquat in klinischen Studien eingesetzt werden – hier gebe es noch „Luft nach oben“, hieß es.

Diskussion um Response-Schwellenwert

Uneins zeigten sich die Teilnehmer auch bei der Interpretation des Schwellenwerts bei Responderanalysen. Darin wird untersucht, ob sich der Anteil an Patienten, die eine spürbare Veränderung im jeweiligen Endpunkt erfahren haben, zwischen den beiden Behandlungsgruppen einer Studie unterscheidet. Das IQWiG vertritt die Position, dass ab einem Response-Schwellenwert von 15 Prozent von

einer zwar noch kleinen, aber hinreichend sicher spürbaren Veränderung ausgegangen werden kann. Dieser Wert habe sich seit seiner Einführung als praxistauglich erwiesen. Zudem hätten sich in der Vergangenheit grundsätzliche Probleme mit der Verwendung von Minimal Important Differences (MID) ergeben, die nach wie vor nicht aufgelöst wurden. Man beobachte aber weiter den Stand der MID-Wissenschaft, hieß es.

Bewertungsrelevanz von PRO-Daten im Fokus

Als Gegenposition wurde formuliert, bei dem Schwellenwert handele es sich um eine „starre deutsche Universalformel“, die beispielsweise besonderen Therapiesituationen nicht Rechnung trage und auch die Problematik einer ausreichenden Studien-Power nicht hinreichend berücksichtige. Zudem liege keine systematisch erhobene quantitative Evidenz vor, die diesen Schwellenwert stütze. Bereits seit dem Jahr 2014 liege der Anteil der Nutzenbewertungsverfahren, bei denen Daten zur Lebensqualität vorgelegt würden, bei rund 70 Prozent. PRO-Daten würden somit erhoben, erlangten bisher aber nicht eine Bedeutung in der Hierarchie der Daten, so dass sie tatsächlich bewertungsrelevant im Hinblick auf den zuerkannten Zusatznutzen werden, lautete die Kritik.

Andere Teilnehmer zeigten sich überzeugt, bei der Auswahl geeigneter Studienendpunkte und Vergleichstherapien sei in den vergangenen Jahren eine „Lernkurve“ festzustellen – dies habe unter anderem damit zu tun, dass mit dem Gesetz für mehr Sicherheit in der Arzneimittelversorgung (GSAV) die Möglichkeit geschaffen wurde, medizinische Fachgesellschaften frühzeitig in Beratungen über die Wahl einer geeigneten Vergleichstherapie einzubinden. Die Zahl der entsprechenden Beratungsanfragen des G-BA habe zuletzt im Jahr bei rund 180 gelegen, hieß es.

Tatsächlich, wurde bei der Tagung deutlich, wird die Vali-

dierung von Surrogatparametern methodisch auch künftig eine hoch anspruchsvolle Aufgabe bleiben. Denn wenn im Rahmen einer Studie klinisch relevante Symptome nicht beobachtet werden können, müsse eine Methodik verfügbar sein, um eine individuelle Beziehung zwischen dem Surrogat und dem Outcome vorhersagen zu können. Voraussetzung für solche Vorhersagen sei eine stabile Test-Umgebung, doch genau dies sei bei den meisten Erkrankungsbildern gerade nicht der Fall.

Erschwerend komme hinzu: Selbst wenn sich ein Surrogat als valide im Rahmen einer bestimmten Behandlung erweise, müsse das längst nicht auch für eine andere Behandlung gelten. Die Varianz in der Vorhersage bleibe in etlichen Fällen so groß, dass die Validierung eines Surrogats selbst dann nicht zuverlässig möglich ist, wenn auf Daten aus großen Patientenkohorten zurückgegriffen wird, hieß es zur Erläuterung.

Allerdings wurden inzwischen unter anderem die Surrogate HBA_{1c} bei Diabetes Typ 2, eGFR oder Creatinine Clearance bei Nierenversagen oder Sustained Virological Response (SVR) bei Viruserkrankungen als klinisch relevante Surrogate akzeptiert.

Unterschiedliche Positionen wurden schließlich bei der Tagung dazu deutlich, welche Bedeutung die im Januar 2025 startende europäische Nutzenbewertung (EU-HTA) haben wird. Einerseits wurde darauf verwiesen, die Definition relevanter Outcomes im Scoping-Verfahren liege auch beim EU-HTA weiterhin bei den Mitgliedsstaaten. Somit gebe es auch kein Ranking und auch keine Kategorisierung von Outcomes im Rahmen des europäischen Assessments. Alle Endpunkte in den künftigen Joint Clinical Assessments (JCA) seien somit gleichberechtigt.

Die entsprechenden Guidance-Dokumente für JCA sehen vor, dass validierte Surrogatendpunkte ergänzend zu patientenrelevanten Endpunkten herangezogen wer-

den können, wenn ein Mitgliedsstaat dies als relevant erachtet. Vor diesem Hintergrund wurde vermutet, es sei unwahrscheinlich, dass die bisher heterogenen nationalen Diskurse über patientenrelevante Endpunkte eine Vereinheitlichung erfahren.

In Absetzung dazu wurden Zweifel laut, ob es sich bei den JCA um völlig wertfreie und ausschließlich „technische“ Berichte handeln werde. Vielmehr sei davon auszugehen, dass es zu einer schrittweisen Konvergenz von bislang unterschiedlichen nationalen Bewertungsstandards kommen wird.

BEIRAT DER INTERDISZIPLINÄREN PLATTFORM

Dr. Jürgen Bausch

Britta Bickel

Wolfgang van den Bergh

Prof. Dr. Wolfgang Greiner

Dr. Ulrike Götting

Dr. Antje Haas

Dr. Harald Herholz

Dr. Ulf Maywald

Dr. Hendrik Pugge

Dr. Daniela Preukschat

Dr. Heinz Riederer

Prof. Dr. Jörg Ruof

Prof. Dr. Bernhard Wörmann

DISKUSSIONSTEILNEHMER

Dr. Jürgen Bausch

Britta Bickel

Prof. Dr. Karl Broich

Prof. Dr. Thomasz Burzykowski

Claus Burghardt

Dr. Jan Daniels-Trautner

Dr. Martin Danner

Prof. Dr. Peter Falkai

Dr. Mathias Flume

Edith Frénoy

Dr. Ulrike Götting

Dr. Rosa Gonzalez-Quevedo

Dr. Harald Herholz

Antje Hoppe

Sabine Jablonka

Dr. Florian Jantschak

Dr. Elaine Julian

Prof. Dr. Sebastian Kluckert

Dr. Werner Kulp

Dr. Norbert Marschner

Dr. Ulf Maywald

Sabine Mentrup

Anja Moeller

Dr. Paula Piechotta

Dr. Daniela Preukschat

Dr. Hendrik Pugge

Dr. Andrey Rasch

Dr. Heinz Riederer

Prof. Dr. Jörg Ruof

Dr. Christian Rybak

Dr. Felix Schönfeldt

Stefan Schwartz

Dr. Johanna Seeger

Eva Stumpe

Wolfgang van den Bergh

IMPRESSUM

HERAUSGEBER

Redaktionsbeirat der
Interdisziplinären Plattform:
Dr. Harald Herholz
Kassenärztliche Vereinigung Hessen
Europa-Allee 90
60486 Frankfurt am Main

VERLAG

Springer Medizin Verlag GmbH
Am Forsthaus Gravenbruch 5
63263 Neu-Isenburg
Handelsregister: Amtsgericht Berlin
Charlottenburg
HRB: 167094 B
Umsatzsteuer-ID: DE 230026696
Telefon: +49 6102 5060
Mail-Adresse: info@aerztezeitung.de

REDAKTIONELLE BEARBEITUNG

Dr. Florian Staechel

AUTOREN

Stefan Schwartz
Dr. Martin Danner
Eva Stumpe
Prof. Dr. Sebastian Kluckert
Dr. Jutta Wendel-Schrief
Prof. Dr. Bernhard Wörmann
Dr. Daniela Preukschat
Dr. Sebastian Meller
Dr. Andreas Rhode
Prof. Dr. Peter Falkai
Prof. Tomasz Burzykowski
Dr. Andrej Rasch
Dr. Johanna Seeger
Prof. Dr. Jörg Ruof

BILDNACHWEIS

Titelbild:
The Little Hut / stock.adobe.com

LAYOUT / GRAFIK

Sandra Bahr
Oliver Hippmann

DRUCK

F&W Druck- und Mediacenter GmbH
Holzhauser Feld 2, 83361 Kienberg
© Springer Medizin Verlag GmbH,
Berlin, März 2025
ISSN 2364-91X



In Zusammenarbeit und mit
freundlicher Unterstützung der
AbbVie Deutschland GmbH & Co. KG,
DAK Gesundheit, MSD Sharp &
Dohme GmbH, Novo Nordisk Pharma
GmbH, Roche Pharma AG, Verband
Forschender Arzneimittelhersteller
e.V., Xcenda

Quellenhinweise zu den Referenten-Portraits

Stefan Schwartz: Jan Pauls | Martin Danner: BAG Selbsthilfe | Eva Stumpe: Privat | Bernhard Wörmann: DGHO |
Sebastian Kluckert: Ralf Baumgarten | Jutta Wendel-Schrief: MSD | Andreas Rhode: Olaf Plotke/MDWL |
Peter Falkei: Fotostudio Sauter | Jörg Ruof: European Access Academy | Tomasz Burzykowski: International Drug
Development Institute | Andrej Rasch: vfa / B. Brundert | Johanna Seeger: Studioline Photography

INTERDISZIPLINÄRE PLATTFORM ZUR NUTZENBEWERTUNG

Welche Endpunkte sind patientenrelevant?

Heft 20
März 2025
ISSN 2364-916X