



Schriftenreihe

INTERDISZIPLINÄRE PLATTFORM ZUR NUTZENBEWERTUNG

Kontextuelle Evidenz – Wege zur gezielten Therapie

Ausgabe 9
Oktober 2019
ISBN 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

ALLE HEFTE SIND ABRUFBAR UNTER:

[HTTPS://WWW.AERZTEZEITUNG.DE/POLITIK_GESELLSCHAFT/GP_SPECIALS/PLATTFORM_ZUR_NUTZENBEWERTUNG/](https://www.aerztezeitung.de/politik_geellschaft/gp_specials/plattform_zur_nutzenbewertung/)
ODER [HTTPS://TINYURL.COM/ZGFEFLK](https://tinyurl.com/zgfeflk)

Inhalt

EDITORIAL

Kontextuelle Evidenz: vom Megatrial zu gezielteren Therapieverfahren 6

HOLGER SCHÜNEMANN

Kontextbezug: Welche Evidenz wird für die jeweilige Forschungsfrage benötigt? 8

BERNHARD WÖRMANN

Evidenzgewinnung und -beurteilung am Beispiel Hämatologie und Onkologie 26

ELENA WOLFF-HOLZ / KLAUS CICHUTEK

Evidenz in der Nutzen-Risiko-Bewertung von Biopharmaka 36

THOMAS KAISER

Eignung einarmiger Studien für die Bewertung des Zusatznutzens 46

DIRK MÜLLER-WIELAND

Evidenzgewinnung und -beurteilung bei Volkskrankheiten am Beispiel Diabetologie 56

GIOVANNI TAFURI

EUnetHTAs Perspektive im Hinblick auf Evidenzanforderungen 64

MICHAEL HENNRICH

Wie umgehen mit Real-World-Data und Registern? Eine politische Betrachtung 70

FLORIAN STAECK

Diskurs: Generierung zusätzlicher Evidenz – nötig sind konsenterte Spielregeln 82

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, Verband Forschender Arzneimittelhersteller e.V. und die Xcenda GmbH an.

Der Beirat der Interdisziplinären Plattform zur Nutzenbewertung

Kontextuelle Evidenz: vom Megatrial zu gezielteren Therapieverfahren

Von Professor Dr. Jörg Ruof

Die Präzision medikamentös-therapeutischer Interventionen hat sich auf Basis verbesserter pathophysiologischer Erkenntnisse, der Verfügbarkeit vielfältiger Biomarker und neuen, zum Teil gentechnisch basierten Therapiemöglichkeiten bei vielen Krankheitsbildern deutlich verbessert. Dieser Trend erhöht die Komplexität in der medizinischen Versorgung, kommt aber letztendlich fraglos den betroffenen Patienten zu Gute.

Im Zusammenhang mit der vergleichenden Bewertung des Zusatznutzens gemäß Paragraf 35a SGB V ergeben sich dadurch verschiedene Herausforderungen:

- Die Anzahl von Patienten, die für eine spezifische klinische Studie zur Verfügung stehen, wird geringer, sie umfasst nur noch die diagnostisch präspezifizierte Teilpopulation sowie die entsprechende Therapielinie;
- Vergleichstherapien basieren z.T. noch auf einem generischen Zulassungsstatus und bilden den aktuellen Zuzugewinn an pathophysiologischen Erkenntnissen nicht oder nur unzureichend ab;
- Die zum Zeitpunkt der Zulassung vorhandene Evidenz erlaubt ggf. die zulassungsrelevante Beurteilung des Nutzen-Risiko Profils, für eine Quantifizierung des Zusatznutzens reichen die Daten jedoch nicht oder noch nicht aus.

Die Arzneimittelnutzenverordnung trägt diesem Trend dem Grunde nach Rechnung. So heißt es dort in Paragraf 5, Absatz 3: „Sofern es unmöglich oder unangemessen ist, Studien höchster Evidenzstufe durchzuführen oder zu fordern, sind Nachweise der bestverfügbaren Evidenzstufe einzureichen.“ Wie dann mit der eingereichten ‚bestverfügbaren‘ Evidenz umzugehen ist, spezifiziert die Verordnung nicht weiter und es bleibt letztendlich dem G-BA bzw. der Selbstverwaltung überlassen, über den Zusatznutzen, dessen Quantifizierbarkeit oder auch eine mögliche Befristung der Entscheidung zu befinden.

Das Berichtsheft der Plattform zur Nutzenbewertung bietet aus verschiedenen Blickwinkeln einen Diskussionsbeitrag zu der zunehmenden Kontextualisierung des bislang überwiegend auf den ‚Goldstandard‘ der randomisierten Vergleichsstudie ausgerichteten Evidenzbegriffs:

- Holger Schünemann beschreibt am Beispiel von zwei kontrovers diskutierten medikamentösen Innovationen die unter anderem in McMaster entwickelte ‚GRADE‘-Methodologie. Diese bietet ein ‚Evidence to Decision‘-Rahmenwerk, welches die immer erforderlichen Wertentscheidungen bei regulatorischen oder HTA-Prozessen nachvollziehbar und transparent machen kann.
- Im Bericht vom Paul-Ehrlich-Institut (Frau Wolff-Holz & Herr Cichutek) werden zunächst die Eckpunkte des deutschen und europäischen Zulassungsprozesses und entsprechende Varianten geschildert. Im Anschluss folgt die Darstellung mehrerer Beispiele, bei denen unter anderem einarmigen Studien (Ceritinib, Axicabtagene-Ciloleucel), Intra-Patient Comparisons (Emicizumab) und komplexen Studiendesigns (Canakinumab) erheblicher Anteil an der regulatorischen Entscheidungsfindung zukam.
- Aus der Perspektive des IQWiG beschreibt Herr Kaiser Verfahren, bei denen Evidenz aus einarmigen Studien für die Nutzenbewertung vorgelegt wurde. Vor allem innovative Verfahren bei Hepatitis C, aber auch in der Onkologie, finden Erwähnung. Zusammenfassend wird betont, dass die Anwendung einarmiger Studien in der Nutzenbewertung auf Ausnahmefälle beschränkt bleiben sollte.
- Die versorgungsrelevante Perspektive der Kliniker wird in den Beiträgen von Herr Wörmann (Hämatologie und Onkologie) und Herr Müller-Wieland (Diabetologie) adressiert. Rasche Identifizierung wirksamer Arzneimittel; die Berücksichtigung von Biomarkern und die Auswahl patienten-orientierter Endpunkte sind zentrale Herausforderungen für die künftige Nutzenbewertung in der Onkologie.

Die Verfügbarkeit und die Interpretation von kardiovaskulären Sicherheitsstudien unter anderem auf Subgruppenebene sowie die Erfassung der Patientenperspektive auf qualitativ hochwertigem Niveau werden in der Diabetologie als wesentliche Faktoren genannt.

- Anschließend vermittelt Giovanni Tafuri, der wissenschaftliche Direktor von EUnetHTA, den Lesern einen Einblick in den gegenwärtigen Diskussionsstand und das laufende Gesetzgebungsverfahren auf europäischer Ebene. Die Erarbeitung eines nachhaltigen Modells für die Zusammenarbeit der 83 nationalen und regionalen europäischen HTA-Institutionen steht hierbei im Vordergrund.
- Die ‚Knackpunkte‘ aus Sicht der Politik werden im Beitrag von Herrn Hennrich dargestellt. So autorisiert die Neufassung des Paragraphen 35a Absatz 3b den G-BA, bei bestimmten Arzneimitteln eine anwendungsbegleitende Datenerhebung zu verlangen. In diesem Zusammenhang werden bisherige Erfahrungen mit Orphan Diseases und Arzneimitteln für neuartige Therapien im Arzneimittelgesetz behandelt sowie Erfahrungen mit den Krebsregistern, dem Rheumaregister ‚RABBIT‘ oder der sogenannten ‚CRISP‘-Plattform beim Lungenkarzinom. Diskutiert wird auch die Frage, wie künftige Datenerhebungen auf europäischer Ebene verzahnt werden sollten. Einen möglichen weiteren Anpassungsbedarf seitens der Politik sieht Herr Hennrich dabei gegenwärtig allenfalls auf der Ebene der Arzneimittelverordnung.

Die Leser seien insbesondere hingewiesen auf die vielfältigen Beispiele, die sich in allen Beiträgen finden und anhand derer letztendlich über die Angemessenheit oder den Anpassungsbedarf bei den aktuellen gesetzlichen und methodischen Rahmenbedingungen zu befinden ist.

Kontakt:
joerg.ruof@r-connect.org

Kontextbezug: Welche Evidenz wird für die jeweilige Forschungsfrage benötigt?

Prof. Dr. Dr. Holger J. Schünemann | Abteilungen für Forschungsmethoden, Evidenz und Wirkung und für Medizin, Direktor des McGRADE und Michael G. DeGroot Cochrane Canada Centre, McMaster University Health Sciences Centre

Um das Ziel der kontextbezogenen Evidenz im Rahmen von Entscheidungen für Regulierungen, Technologiebeurteilungen und Leitlinien zu erreichen, werden gut strukturierte Entscheidungsrahmen benötigt, die austauschbar sind und dennoch an die jeweilige Art der Entscheidungsfindung angepasst werden können. Entscheidungen hängen stets von der gewählten Perspektive ab. Mit Hilfe von strukturierten Verfahren, die nachvollziehbar sind und in den richtigen Kontext gesetzt werden, wird die Transparenz verbessert und Effizienzgewinne auf mikro- und makroökonomischer Ebene geschaffen. Die Organisatoren der Konferenz schlugen ein Beispiel vor, das sich auf die Behandlung von Hämophilie A konzentriert, und ich habe einen zweiten Vorschlag ergänzt, bei dem der EtD-Ansatz (engl. Evidence to Decision, EtD) der GRADE-Methodik (engl. Grading of Recommendations Assessment, Development and Evaluation) anhand der multiresistenten Tuberkulose (MDR-TB) vorgestellt wird. Bei den beiden beschriebenen Beispielen können Entscheidungsträger beim EtD-Ansatz zuversichtlich sein, dass trotz Unsicherheit bezüglich der Effekte der Interventionen sie zum Zeitpunkt der Entscheidung eine informierte Entscheidung treffen.

Einleitung
Es sollte stets das erklärte Ziel von Entscheidern im Gesundheitswesen sein, unter dem Strich keinen Schaden¹ anzurichten. Um das Ziel der kontextbezogenen Evidenz im Rahmen von Entscheidungen für Regulierungen, Technologiebeurteilungen und Leitlinien zu erreichen, werden gut strukturierte Entscheidungsrahmen benötigt, die austauschbar sind und dennoch an die jeweilige Art der Entscheidungsfindung angepasst werden können.

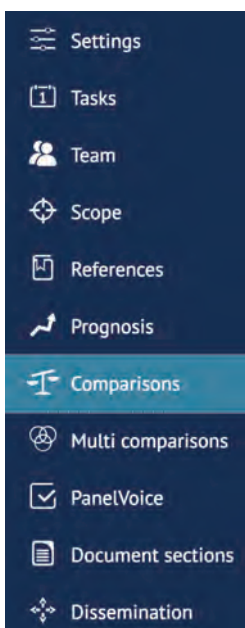
Entscheidungen hängen stets von der gewählten Perspektive ab, wie z.B. des Einzelnen, der Bevölkerung oder des Gesundheitssystems. Mit Hilfe von strukturierten Verfahren, die nachvollziehbar sind und in den richtigen Kontext gesetzt werden, wird die Transparenz verbessert und Effizienzgewinne auf mikro- und makroökonomischer Ebene geschaffen.

Es wurden viele Kriterien vorgeschlagen, um umfangreiche Rahmenbedingungen hierfür zu schaffen. Die EtD-Ansätze der GRADE-Methodik werden nun zunehmend genutzt, um die Ziele der Vollständigkeit, Transparenz und Benutzerfreundlichkeit in unterschiedlichen Kontexten und Rechtsprechungen zu erreichen.²⁻⁶ Der EtD-Ansatz besteht aus vier Abschnitten: Fragestellung und Hintergrund, Bewertung, Schlussfolgerungen und Präsentation.

Zum Abschnitt Fragestellung und Hintergrund gehört die detaillierte Beschreibung der PICO (engl. Population, Intervention, Comparators and Outcomes) und möglicher Untergruppen sowie eventueller Interessenskonflikte der Gruppe, die die Fragestellung aufwirft bzw. beantwortet.

Die Bewertung beschreibt die Kriterien, die für die Entscheidung maßgeblich sind. Gruppen oder Organisationen bewerten bis zu elf Kriterien (die Kriterien werden flexibel auf der Basis der Bedürfnisse und Anforderungen des Entscheidungsträgers ausgewählt), einschließlich eines Be-

Interaktive Evidence-to-Decision-Ansätze



Frage

- Details – PICO-Subgruppen
- Hintergrund und Interessenskonflikte

Bewertung

- Kriterien
- Entscheidungen
- Forschungsevidenz (HTA und systematische Überprüfungen)
- Weitere Überlegungen

Schlussfolgerungen

- Art der Entscheidung – Empfehlung
- Entscheidung
- Überlegungen zur Implementierung – Überwachung und Beurteilung
- Forschungsüberlegungen

Präsentation

- Treffen der Richtliniengruppe & Mitteilung von Kostenübernahmeentscheidungen
- Datenbank mit Entscheidungsrahmen
- Interactive Decision Aids (iDeAs), Apps



Perspektiven

- Klinische – persönliche
- Klinische – Bevölkerung
- Gesundheitssysteme & Öffentliche Gesundheit
- Gesundheitssysteme & Politik

Art von Entscheidungen

- Empfehlung
- Richtlinie
- Kostenübernahme

Verwendung

- Gruppenentscheidungsfindung
- Persönlich / Online

Quelle: gradepro.org



Prof. Dr. Dr. Holger Schünemann arbeitet als Professor in den Abteilungen für Forschungsmethoden, Evidenz und Wirkung (Health Research Methods, Evidence and Impact) und für Medizin an der McMaster University und leitete erstere Abteilung von 2009 bis 2019. Nach Medizinstudium an der Medizinischen Hochschule Hannover (Promotion 1994)

absolvierte er Weiterbildungen in den Bereichen Lungenbiologie, Epidemiologie, Innere Medizin und Präventivmedizin/Öffentliche Gesundheit an der Universität von Buffalo, State of New York. Er ist Ko-Vorsitzender der GRADE-Arbeitsgruppe und Direktor der Cochrane Canada und McMaster GRADE-Zentren. Er betreibt eine Praxis für Innere Medizin.

Kriterien, Evidenzlage und zusätzliche Überlegungen, die Stärke und Ausrichtung der GRADE-Empfehlung beeinflussen

1. Problem: Ist das Problem eine Priorität?
2. Erwünschte Effekte: Wie groß sind die zu erwartenden erwünschten Effekte?
3. Unerwünschte Effekte: Wie groß sind die zu erwartenden unerwünschten Effekte?
4. Vertrauenswürdigkeit der Evidenz: Wie ist das gesamte Vertrauen in die Evidenz der Effekte?
5. Wertvorstellungen: Gibt es wesentliche Unsicherheit darüber, welche Wichtigkeit Betroffene den Hauptendpunkten beimessen oder variieren die Einschätzungen?
6. Abwägen der Effekte: Spricht ein Abwägen von erwünschten und unerwünschten Wirkungen für die Intervention oder die Vergleichsintervention?
7. Benötigte Ressourcen: Wie hoch ist der Ressourcenbedarf (Kosten)?
8. Vertrauenswürdigkeit der Evidenz der benötigten Ressourcen: Wie groß ist das Vertrauen in die Evidenz zum Ressourcenbedarf (Kosten)?
9. Kostenwirksamkeit: Spricht die Kostenwirksamkeit für die Intervention oder die Vergleichsintervention?
10. Gleichstellung: Was wären die Auswirkungen auf die Gerechtigkeit im Gesundheitswesen?
11. Akzeptanz: Ist die Intervention für die wichtigsten Interessengruppen akzeptabel?
12. Machbarkeit: Ist es möglich die Intervention umzusetzen?

Quelle: GRADE EtD framework

Abbildung 2: Kriterien, Forschungsevidenz, Entscheidungen und weitere Überlegungen, die die Ausprägung und Ausrichtung bei den EtD-Ansätzen der GRADE-Methodik beeinflussen.

reichs über Forschungsevidenz – idealerweise aus systematischen Übersichtsarbeiten oder HTA stammen – sowie weiteren Überlegungen und Entscheidungen.

Der Abschnitt der Schlussfolgerungen beginnt mit einer Entscheidung oder Empfehlung, die sich aus der Überprüfung ergibt, und gegebenenfalls der Gewichtung, d. h. der Aufschlüsselung der Entscheidungen für die einzelnen Kriterien. Je nachdem, welcher Ansatz verwendet und welche Perspektive eingenommen wird, können dies Empfehlungen aus klinischen Leitlinien oder Entscheidungen zur Kostenübernahme oder zum Gesundheitssystem sein. Schließ-

lich können diese mit dem offiziellen GRADE-Tool und der GRADEpro-App auf unterschiedliche Art und Weise präsentiert und verbreitet werden (www.grade.pro.org) (Abbildung 1).

Um Entscheidungen transparent zu machen, sollten Entscheidungsträger ihre Perspektive (individuelle Perspektive, Bevölkerungsperspektive, Perspektive des Gesundheitssystems und der Politik) und die Art der Entscheidung (Empfehlung, Richtlinie oder Kostenübernahme) erläutern. Obwohl viele dieser Konzepte von Entscheidungsträgern in ganz Europa – also auch in Deutschland – verwendet

werden, fehlen bisher umfassende Strukturen und Kooperationen, um die gemeinsamen Ziele zu erreichen und einen Austausch zwischen den Organisationen zu ermöglichen.

Mein Ziel bei dieser Präsentation war es – auf der Basis von Beispielen, die Debatten zwischen Entscheidungsträgern ausgelöst haben – zu beschreiben, welche Art von Forschung benötigt wird, um Vertrauen in die Ergebnisse dieser Forschungsaktivitäten und Vertrauen in die daraus resultierende Entscheidung zu schaffen, selbst wenn die Forschung keine vertrauenswürdigen Entscheidungen im Hinblick auf den Effekt einer Intervention zulässt. Außerdem wollte ich aufzeigen, dass die Verwendung solcher Entscheidungsrahmen den Austausch und die Nutzung zwischen unterschiedlichen Agenturen bzw. Ressorts erleichtern kann.

Zur Erläuterung dieser Konzepte werde ich in diesem Artikel erörtern, welche Kriterien sich beim EtD-Ansatz auf dieses Vertrauen auswirken und wie bzw. in welchen Bereichen innerhalb dieser Kriterien sie dies tun.

2. Beispiele

Die Organisatoren der Konferenz schlugen ein Beispiel vor, das sich auf die Behandlung von Hämophilie A konzentriert, und ich habe einen zweiten Vorschlag ergänzt, bei dem der EtD-Ansatz vorgestellt wird.⁵ Die EtD-Ansätze der GRADE-Methodik sind das Ergebnis eines auf fünf Jahre angelegten Projekts, das von der Europäischen Kommission unterstützt wurde (DECIDE-Projekt) und seitdem von zahlreichen Organisationen für Entscheidungen und Empfehlungen im Gesundheitssystem verwendet wird.⁷

Hämophilie A

Ich wurde gebeten, ein Beispiel in Zusammenhang mit Hämophilie A anzusprechen. Hämophilie A ist eine Erbkrank-

heit, bei der die Blutgerinnung gestört ist, und bei der das Blut des Betroffenen aufgrund eines angeborenen Faktor-VIII-Mangels nicht oder nur langsam gerinnt. Dadurch kann es zu unkontrollierten Blutungen kommen, die schon bei leichten Verletzungen oder sogar spontan auftreten. Hämophilie A kann die Lebensqualität der Betroffenen stark beeinträchtigen. Sie ist eine seltene Erkrankung, die nur bei ca. 1 von 5.000 Menschen auftritt. Bei ungefähr 1 von 10.000 Patienten ist sie stark ausgeprägt.

Bei der Hämophilie bezieht sich der Begriff Inhibitor auf einen Autoantikörper, der normalerweise als Reaktion auf den infundierten Faktor VIII gebildet wird. Inhibitoren finden sich am häufigsten bei Menschen mit sehr niedrigen Faktor VIII-Ausgangskonzentrationen. Emicizumab ist ein rekombinanter, humanisierter, bispezifischer, therapeutischer, monoklonaler Antikörper, der die hämostatische Funktion des FVIIIa (im menschlichen Körper durch Überbrückung des aktivierten Faktor IX (FIXa) bzw. Faktor X (FX)) ersetzen soll.^{8,9} Eine der wichtigsten klinischen PICO-Fragen, die sich hier stellt, um zu beurteilen, ob es einen Netto-Gesundheitsnutzen gibt, ist:

Population: Bei Menschen mit einem Faktor VIII-Mangel mit Inhibitoren, welche Auswirkungen hat
Intervention: Emicizumab im Vergleich zu
Komparator: Ohne Emicizumab auf die
Endpunkte: Blutungsraten, unerwünschten Nebenwirkungen der Behandlung, Lebensqualität

Multiresistente Tuberkulose

Anhand eines zweiten Beispiels möchte ich einige der wichtigsten Punkte in diesem Artikel erläutern. Tuberkulose (TB) gehört zu den ältesten bekannten Erkrankungen

der Menschheit. Sie ist heute noch eine der weltweit zehn häufigsten Todesursachen und die weltweit führende tödliche Infektionskrankheit. Bei einer multiresistenten Tuberkulose (MDR-TB) handelt es sich um eine TB, die aufgrund einer Resistenz des Organismus gegen die wichtigsten Antibiotika schwer zu behandeln ist. Nach Schätzungen der WHO tritt sie bei mehr als 600.000 Menschen auf, und jedes Jahr sterben mehr als 240.000 Menschen daran. Die Behandlung der MDR-TB besteht normalerweise aus einer Kombination von mehreren Antibiotika, um Synergieeffekte zu erzielen und die Patienten so heilen zu können.

Bedaquilin ist ein recht neuer antibiotischer Wirkstoff, den die WHO für die Behandlung der MDR-TB empfohlen hat.¹⁰ Eine entscheidende Frage ist nun, ob Bedaquilin auf der Basis der WHO-Empfehlungen bei der Behandlung von MDR-TB Patienten als Ergänzung zur Basisbehandlung eingesetzt werden sollte? Folgende PICO-Gesundheitsfrage haben wir in einem WHO-Gremium gestellt, das Empfehlungen abgibt, um einen möglichen Netto-Gesundheits-

Population: Bei Patienten mit einer multiresistenten Tuberkulose (MDR-TB), welche Auswirkungen hat

Intervention: Bedaquilin plus MDR-TB Hintergrund-Behandlung im Vergleich zu

Komparator: MDR-TB Hintergrund-Behandlung alleine auf die

Endpunkte: Heilung nach 120 Wochen, unerwünschten Arzneimittelwirkungen (klinisch und biologisch schwerwiegende Nebenwirkungen), Mortalität, Zeit bis zur Kulturkonversion, Kulturkonversion nach 24 Wochen, erworbene Resistenz gegen Fluoroquinolon und injizierbare Arzneimittel

nutzen zu beurteilen: Es werden Entscheidungen benötigt, und Fragen beantworten sich auch bei vorliegender Evidenz nicht von selbst, sodass eine gewisse Unsicherheit bleibt bei den Entscheidungsträgern zurückbleibt.

Nachfolgend werde ich beschreiben, welche Art von Forschung benötigt wird, um Vertrauen in die Forschungsergebnisse zu schaffen, und wie man Vertrauen in die Entscheidung bekommen kann, selbst wenn die Forschungsergebnisse keine fundierten Entscheidungen hinsichtlich der Effekte einer Intervention und anderer Arten von Forschungsevidenz zulassen. Dies ergibt sich aus den folgenden strukturierten Prozessen und dem Verständnis und der Akzeptanz der Lücken von Forschungsevidenz.

3. Vertrauenswürdigkeit der Evidenz im Vergleich zu Vertrauen in eine Empfehlung oder Entscheidung

Was versteht man unter Vertrauenswürdigkeit der Evidenz?

Die Gewissheit, dass ein geschätzter Zusammenhang bzw. Effekt richtig ist oder - was noch besser ist – dass ein tatsächlicher Effekt auf einer Seite eines zuvor festgelegten Schwellenwertes bzw. gewählten Bereichs liegt.¹¹

Was versteht man unter Vertrauen in eine Entscheidung?

Vertrauen in eine Entscheidung entsteht, wenn ein Entscheidungsträger das Gefühl hat, dass die besten verfügbare Evidenz – unabhängig davon, ob diese eine hohe oder sehr niedrige Vertrauenswürdigkeit hat – mit Hilfe von systematischen Überprüfungen oder HTA und im Kontext eines transparenten und umfassenden Ansatzes bei der Beurteilung dieser Evidenz zusammengetragen wurde. Welche Entscheidung auch immer getroffen wird, z. B. ob die Kosten für eine Intervention übernommen werden

oder nicht, oder ob angesichts der Unsicherheit der Evidenz zu möglichen Gesundheitseffekten eine bedingte Empfehlung für eine Intervention ausgesprochen wird, für einen Entscheidungsträger kann seine Arbeit zuverlässig und vertretbar sein, selbst wenn die Forschungsevidenz lücken- oder fehlerhaft ist.

Beim Hämophilie A-Beispiel stammte die Evidenz hauptsächlich aus zwei randomisierten Studien, einer mit Patienten mit Inhibitoren,⁸ und zwei jüngeren repräsentativen HTA, die ich mit Hilfe einer pragmatischen Literatursuche ausgewählt habe.^{12,13} Es gab keine formelle Einstufung der Vertrauenswürdigkeit der Evidenzlage, sondern eine Beurteilung des Verzerrungsrisikos nur bei den Berichten, auf die ich später noch einmal eingehen werde. Ich stelle ein hypothetisches Szenario für eine Entscheidung dar, die auf der Basis der überprüften Evidenz sicher getroffen werden könnte.

In den WHO MDR-TB Richtlinien stammen die wichtigsten Daten aus einer randomisierten Studie mit bis zu 160 Probanden. Die Vertrauenswürdigkeit der Evidenz wurde hinsichtlich der zu erwartenden Heilung, Mortalität und unerwünschten Nebenwirkungen als die wichtigsten Endpunkte als sehr gering eingestuft (siehe Abschnitt 4). Trotz dieser Ungewissheit hinsichtlich der geschätzten wichtigsten Hauptendpunkte könnte das Gremium davon überzeugt sein, mit seiner bedingten Empfehlung der Intervention mit der Maßgabe, unter welchen Umständen die Intervention eine sinnvolle Option wäre, die richtige Entscheidung getroffen zu haben.

Dieser Unterschied zwischen der (sehr geringen) Vertrauenswürdigkeit der Evidenz und dem (hohen) Vertrauen in eine Entscheidung ergibt sich daraus, dass alle Aspekte berücksichtigt werden, die in offener und transparenter Art und Weise Informationen für die Empfehlung liefern, und dann auf die Unsicherheit bezüglich der Effekte von Bedä-

quilin hingewiesen wird. Die Erkenntnis, dass weitere Forschungsaktivitäten die Vertrauenswürdigkeit der Evidenz erhöhen könnten und dies zu einer anderen Empfehlung führen könnten, verbesserten das Vertrauen in die Entscheidung zusätzlich. Nachstehend werde ich die Einzelheiten und die Grundprinzipien erklären, die sich dahinter verbergen.

4. Kriterien beim EtD-Ansatz

Abbildung 2 zeigt die Kriterien des EtD-Ansatzes der GRADE-Methodik, aus denen die Personen, die die Beurteilungen durchführen und Entscheidungen treffen, auswählen können.

4.1. Ist das Problem prioritär?

Die Beurteilung, ob das Problem prioritär ist, richtet sich nach der Bedeutung und Häufigkeit des entsprechenden medizinischen Problems (Krankheitslast, Prävalenz, Kosten, Ausgangsrisiko, politische Frage etc.).

Benötigte Evidenz und Beispiele

Wenn die Prävalenz und das Risiko zu Beginn der Erkrankung die relevanten Faktoren sind, stammt die Evidenz, die benötigt wird, um herauszufinden, ob ein Problem prioritär ist, normalerweise aus Kohortenstudien. Leitliniengruppen verwenden häufig Umfragen, um herauszufinden, ob es Abweichungen in der Praxis, neue wichtige Evidenz, unklare Leitlinien oder fehlende Empfehlungen gibt, um die Wichtigkeit eines Problems zu beurteilen.

Während eine hohe Prävalenz einer Erkrankung oder eine große Anzahl an Betroffenen die Wahrscheinlichkeit erhöht, dass eine Intervention, die das Problem beseitigt, prioritär sein sollte, können für seltene Erkrankungen andere Prioritätsgründe gelten. So schaffen beispielsweise die Einführung oder Zulassung neuer Interventionen, bei denen

man davon ausgeht, dass sie eine hohe Wirksamkeit haben, oder hohe Medikamentenkosten zur Behandlung von seltenen Erkrankungen, für die es bislang nur wenige Optionen gab, Prioritäten für Patienten, Kliniker und politische Entscheidungsträger. In ähnlicher Weise schaffen Anordnungen von Arzneimittelbehörden und Verordnungsgebern Prioritäten. Dementsprechend wird hierfür eine andere Art von Forschungsevidenz benötigt, und es werden häufig Erläuterungen im Abschnitt der zusätzlichen Überlegungen zur Priorität gemacht.

Bei den Beispielen 1 und 2 waren neue Evidenz für Interventionen mit einem potenziell großen Behandlungserfolg bei seltenen Erkrankungen bzw. die Verfügbarkeit und Zulassung eines neuen Antibiotikums für eine tödliche Erkrankung auf der Basis von Beobachtungsstudien bzw. Studien die Gründe dafür, dass ein Problem als prioritär angesehen wird.

4.2. und 4.3. Wie groß sind die Auswirkungen einer Intervention auf die erwünschten und unerwünschten Gesundheitseffekte (Vorteile und Schaden und Last)?

Hierfür muss das Ausmaß der absoluten Effekte einer Intervention auf die erwünschten und unerwünschten Effekte beurteilt werden. Es müssen alle entscheidenden und manchmal auch die wichtigen Endpunkte berücksichtigt werden, um zu entscheiden und festzustellen, ob diese Effekte unbedeutend, gering, moderat oder groß sind.

Benötigte Evidenz und Beispiele

Für Beispiel 1 benötigen wir Interventionsstudien, und zwar im Idealfall methodisch präzise durchgeführte randomisierte kontrollierte Studien, in denen die Behandlung mit Emicizumab mit der ohne Emicizumab verglichen und die Auswirkungen auf die relevanten Endpunkte beurteilt wird: Blutungsraten, unerwünschte Nebenwirkungen und

Lebensqualität. Die beobachteten Effekte waren eine relative Reduktion der Blutungsepisoden um 87 Prozent, was eine deutliche Verbesserung der Lebensqualität darstellt, sowie eine Steigerung der unerwünschten Nebenwirkungen.^{8,12}

In Beispiel 2 wären randomisierte Studien, bei denen das bestehende WHO-Behandlungsprotokoll um Bedaquilin ergänzt wurde, die Art von Studien, die eine Schätzung des zu erwartenden absoluten Effekts bei repräsentativen Patienten hinsichtlich aller relevanten Endpunkte ermöglicht. Die Ergebnisse ließen einen großen Effekt (> 20 Prozent) auf die Heilungsraten vermuten, die einem potenziell großen Anstieg der Mortalität (bis zu 10 Prozent) und unerwünschten Nebenwirkungen gegenüber stand (in den Richtlinien von 2013).

4.4. Vertrauenswürdigkeit der Evidenz über die erwünschten und unerwünschten Effekte

Die Vertrauenswürdigkeit der Evidenz kann für die einzelnen Kriterien beurteilt werden, die als Forschungsevidenz in einem EtD-Ansatz enthalten sind, spielt jedoch eine wesentliche Rolle bei der Beurteilung der Effekte einer Intervention auf die Gesundheitseffekte. Die acht Bereiche des GRADE-Systems werden erst individuell und dann im Verhältnis zueinander beurteilt, um einen Gesamtüberblick über die Vertrauenswürdigkeit der Evidenz zu bekommen, idealerweise auf der Basis von systematischen Überprüfungen in Zusammenhang mit der PICO-Frage. Diese acht Bereiche sind: Verzerrungspotenzial, Inkonsistenz, Indirektheit, unzureichende Präzision und Publikationsbias, starke Assoziation, Dosis-/Wirkungs-Verhältnis und plausible Confounder oder Verzerrungen, die den gezeigten Behandlungseffekt eher reduzieren bzw. einen scheinbaren Effekt erzeugen würden.

4.4.1. Beurteilung des Verzerrungspotenzials (Grenzen bei Design und Durchführung von Studien) bei Studien

Sowohl bei randomisierten kontrollierten Studien (engl. randomized controlled trials, RCT), als auch bei (nicht randomisierten) Beobachtungsstudien kann ein zusätzliches Risiko bestehen, irreführende Ergebnisse zu erzielen, wenn deren Design oder Durchführung fehlerhaft sind. Die entscheidende Frage für diesen Bereich ist „Sind die Forschungsstudien gut geplant und durchgeführt?“. Hierbei spricht man häufig von Problemen mit der „Validität“ oder „internen Validität“, was bei GRADE als „Studienlimitationen“ oder „Verzerrungsrisiko“ eingestuft wird.

In randomisierten kontrollierten Studien und Beobachtungsstudien werden verschiedene Tools verwendet, um das Verzerrungsrisiko zu beurteilen. In den meisten Fällen werden jedoch ähnliche Punkte verwendet, z.B. Verzerrung aufgrund von Confoundern, Verzerrung durch die Auswahl der Teilnehmer an der Studie, Verzerrung bei der Klassifizierung von Interventionen, Verzerrungen aufgrund von Abweichungen von geplanten Interventionen, Verzerrungen aufgrund von fehlenden Daten, Verzerrungen bei den Messung von Endpunkten und Verzerrung bei der Auswahl des berichteten Ergebnisses.¹⁴ GRADE schlägt vor, eine Gesamtbeurteilung des Verzerrungsrisikos nach Studie und Endpunkt und dann zwischen den einzelnen Studien nach Endpunkt vorzunehmen.¹⁵

4.4.2. Beurteilung der Inkonsistenz der Ergebnisse

Die entscheidende Frage für diesen Bereich ist: „Sind die Ergebnisse zwischen den einzelnen Studien konsistent, wenn sie konsistent sein sollten, da sie ähnliche PICO-Fragen behandeln?“ Sehr unterschiedliche Schätzungen des Behandlungseffekts (d.h. Heterogenität oder Variabilität der Ergebnisse) zwischen den einzelnen Studien lassen auf tatsächliche Unterschiede beim zugrunde liegenden Be-

handlungseffekt schließen. Bei einer Heterogenität, für die die Untersucher keine plausible Erklärung finden können, sollte die Vertrauenswürdigkeit der Evidenz je nach Ausprägung der Inkonsistenz bei den Ergebnissen um eine oder zwei Stufen herabgestuft werden. Diese Inkonsistenz kann durch unterschiedliche Patientenpopulationen bedingt sein (z.B. Medikamente können größere relative Effekte bei kränkeren Populationen haben), Interventionen & Komparatoren (z.B. größere Effekte bei höheren Medikamentendosen, leicht veränderte Interventionen oder unterschiedliche Komparatoren) oder Endpunkte (z.B. nachlassender Behandlungseffekt im Zeitverlauf).

4.4.3. Beurteilung der Indirektheit der Ergebnisse

Die entscheidende Frage für diesen Bereich ist „Wie direkt stehen die Ergebnisse mit unserer PICO-Frage in Zusammenhang?“ Evidenz ist stets etwas indirekt. Der Begriff Indirektheit beschreibt, wie direkt die identifizierte Evidenz mit der Forschungsfrage in Zusammenhang steht. Es gibt zwei Arten von verbundener Indirektheit (man spricht hier auch von Verallgemeinerbarkeit, Übertragbarkeit, externer Validität, Relevanz, Anwendbarkeit, Umsetzbarkeit). Erstens kann die Evidenz von einer indirekten Population, Intervention, Komparator oder Endpunkt stammen. Zweitens kann sie aus einem indirekten Vergleich stammen, sofern Vergleiche zwischen Intervention A und B nicht verfügbar sind, aber Studien, in denen A mit C und B mit C verglichen wurden. Solche Studien ermöglichen indirekte Vergleiche der Ausprägung des Effekts von A im Vergleich zu B. Diese Evidenz hat eine geringere Vertrauenswürdigkeit als ein Head-to-Head-Vergleich von A und B.

4.4.4. Beurteilung der unzureichenden Präzision der Ergebnisse

Die entscheidende Frage für diesen Bereich ist „Ist die Effektgröße präzise – aufgrund zufälliger Messabweichung?“ Der Begriff unzureichende Präzision beschreibt den Grad der zufälligen Messabweichung, die die Beurteilung der Ergebnisse und die Interpretation beeinflussen kann. Ergebnisse sind dann ungenau, wenn Studien relativ wenige Patienten und wenige Ereignisse beinhalten und somit breite Konfidenzintervalle um den geschätzten Effekt bzw. die Assoziation haben. Es gibt bestimmte Grundprinzipien, nach denen man Ergebnisse als unpräzise bezeichnet. Wenn man die Vertrauenswürdigkeit der Evidenz berücksichtigt, ist die Frage, ob das Konfidenzintervall um den geschätzten Effekt bzw. die Assoziation eng genug ist.

4.4.5. Beurteilung des Risikos des Publikationsbias

Die entscheidende Frage für diesen Bereich ist „Stellen diese Studien die Gesamtheit aller Untersuchungen dar, die für diesen Endpunkt durchgeführt wurden?“ Publikationsbias ist eine systematische Unter- oder Überschätzung der zugrunde liegenden vorteilhaften oder schädlichen Effekte aufgrund einer selektiven Publikation von Studien (Publikationsbias). Untersucher veröffentlichen ihre eigenen Studien nicht (normalerweise solche, die keinen Effekt zeigen) oder Studien, die keinen Effekt ergeben, werden seltener von Fachzeitschriften akzeptiert. Das passiert häufig, wenn es sich um kommerzielle Interessen handelt und es nur kleine Studien gibt, die positive Effekte zeigen.

4.4.6. Beurteilung einer möglichen starken Assoziation

Die entscheidende Frage für diesen Bereich ist „Ist der Effekt groß oder sehr groß und (relativ) wertneutral?“ Wenn methodisch präzise durchgeführte Beobachtungsstudien (mit einem niedrigen Verzerrungsrisiko des relevanten

Tools) große oder sehr große und konsistente Schätzungen des Ausmaßes eines Behandlungs- bzw. Expositionseffekts ergeben, können wir den Ergebnissen vertrauen. In diesen Situationen liefert das schwache Studiendesign wahrscheinlich keine Erklärung für den gesamten erkennbaren Nutzen oder Schaden, selbst wenn Beobachtungsstudien vermutlich eine Überschätzung des tatsächlichen Effekts ergeben. Je größer die Ausprägung des Effekts, umso stärker wird die Evidenz. Als Grenzwerte für die Beurteilung schlägt die GRADE-Methodik relative Risiken (RR) von >2 oder $<0,5$ für Fälle ohne plausible wichtige Confounder bzw. sehr große RR von >5 oder $<0,2$ für Fälle ohne plausible wichtige Confounder vor.

4.4.7. Beurteilung des Dosis-Wirkungs-Verhältnisses

Die entscheidende Frage für diesen Bereich ist: „Bestehen Zusammenhänge zwischen Expositionsdosis und Effekten, die uns mehr Sicherheit geben?“ Eine Dosis-Wirkungs-Beziehung kann unser Vertrauen in die Ergebnisse von Beobachtungsstudien erhöhen und somit zu einer größeren Vertrauenswürdigkeit der Evidenz führen. Nur Studien, bei denen die Validität nicht gefährdet ist (die aus irgendeinem Grund nicht herabgestuft wurden), sollten uns mehr Sicherheit geben.

4.4.8. Beurteilung der Auswirkungen von plausiblen Confoundern oder Verzerrungen, die den gezeigten Behandlungseffekt eher reduzieren bzw. einen scheinbaren Effekt erzeugen würden

Die entscheidende Frage für diesen Bereich ist: „Lassen die Ergebnisse trotz schlimmstmöglicher Prädiktoren dennoch Schlussfolgerungen zum Effekt zu?“ Manchmal können alle plausiblen Confounder aus Beobachtungsstudien dazu beitragen, den gezeigten Effekt zu reduzieren oder ihn zu verstärken, falls kein Effekt beobachtet wurde. Wenn bei-

spielsweise kränkere Patienten eine experimentelle Intervention oder Exposition erhalten, kann der tatsächliche Interventions- oder Expositionseffekt größer sein, als die Daten nahe legen, wenn es ihnen besser ergeht.

4.4.9. Gesamtvertrauenswürdigkeit der Evidenz

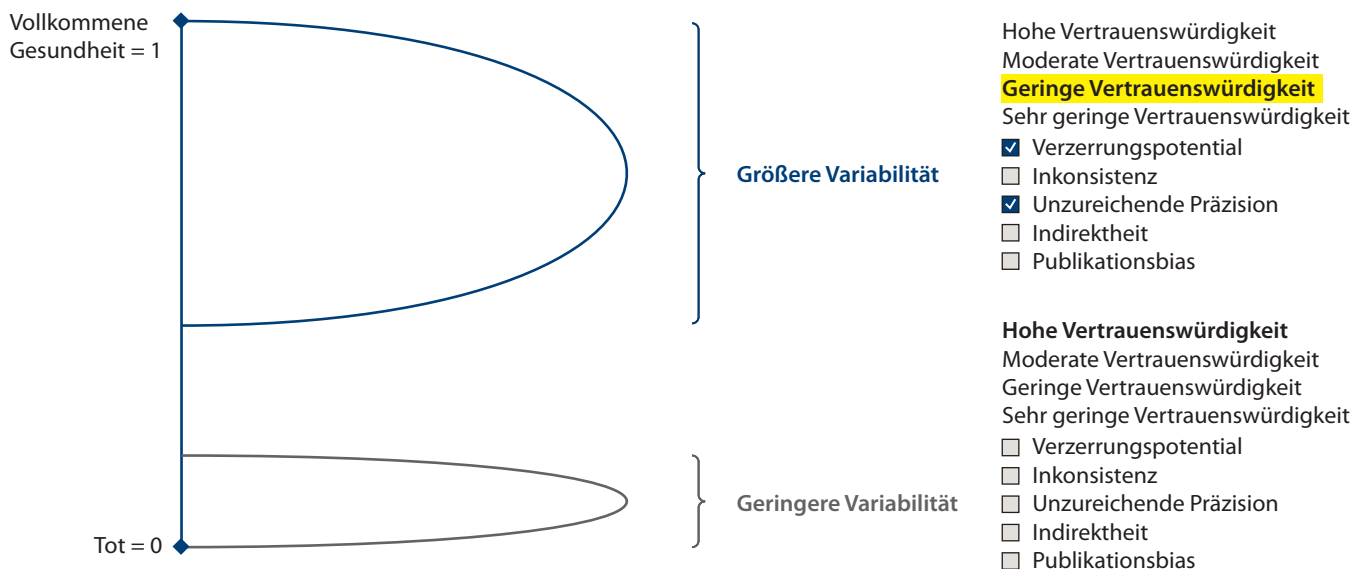
Die Vertrauenswürdigkeit der Evidenz wird dann nach der Beurteilung der Evidenz für alle Bereiche basierend auf der Herabstufung der Vertrauenswürdigkeit innerhalb und zwischen den Bereichen um eine bis drei Stufen (je nach verwendeten Instrumenten), zunächst nach Endpunkten in hoch, moderat, niedrig und sehr niedrig eingestuft.^{15, 16} Für Entscheidungen wird dann die niedrigste Vertrauenswürdigkeit der Evidenz für alle wichtigen Endpunkte – egal ob

Nutzen oder Schaden – verwendet, um die Gesamtvertrauenswürdigkeit der Evidenz wiederzugeben.

4.4.10. Benötigte Evidenz und Beispiele

Bei beiden Beispielen ergäbe die Verfügbarkeit aller randomisierten kontrollierten Studien, die konsistent, präzise und direkt sind und ein niedriges Verzerrungsrisiko haben, eine hohe Vertrauenswürdigkeit der Evidenz für die Behandlungseffekte. Beim Hämophilie A-Beispiel gab es in den beiden HTA-Berichten beispielsweise Bedenken hinsichtlich des Verzerrungsrisikos und der Präzision der Schätzungen. Dies würde bei der geringen Anzahl an Patienten, die an der Studie teilnahmen, zu einer niedrigen oder sehr niedrigen Vertrauenswürdigkeit der Evidenz füh-

Variabilität und Vertrauenswürdigkeit der Evidenz über die relative Bedeutung von Endpunkten



Quelle: Prof. Dr. Dr. Schünemann

Abbildung 3: Variabilität und Vertrauenswürdigkeit der Evidenz über die relative Bedeutung von Ergebnissen.

ren. In seltenen Fällen haben wir großes Vertrauen in Behandlungseffekte, wenn diese sehr ausgeprägt und präzise sind. Das kann in Situationen passieren, in denen alle Patienten vor der Intervention unter schwerwiegenden Konsequenzen litten und nach Beginn der Intervention diese Endpunkte fehlen bzw. nahezu fehlen.

Der Effekt von Emicizumab auf die Blutungsraten erfüllt derzeit noch nicht das Grundprinzip für eine hohe Vertrauenswürdigkeit auf der Basis von ausgeprägten Effekten, weil der Effekt auf einer relativ niedrigen Anzahl an Patienten und Ereignissen beruht. Methodisch präzise durchgeführte nicht randomisierte Studien können die Vertrauenswürdigkeit erhöhen – zumindest für den Endpunkt Blutungsrate, sofern diese methodisch präzise sind und keine anderen wichtigen Einschränkungen haben, die Effekten zeigen würden, die denen ähneln, die in der kleinen randomisierten Studie beobachtet wurden.

Beim MDR-TB-Beispiel wurde die Vertrauenswürdigkeit der Evidenz für die meisten Endpunkte vom WHO-Gremium im Jahr 2013 als niedrig bzw. sehr niedrig eingestuft – hauptsächlich aufgrund der Bedenken hinsichtlich der unzureichenden Präzision und Indirektheit.

4.5. Wertvorstellungen und Präferenzen bzw. die Bedeutung von Endpunkten

Dieses Kriterium beschreibt, wie wichtig Gesundheitsergebnisse für die Betroffenen sind, wie variabel sie sind, und ob es Unsicherheiten im Zusammenhang mit ihnen gibt. Diese relative Bedeutung der Endpunkte entspricht den Wertvorstellungen und Präferenzen, die Menschen mit Gesundheitsergebnissen in Verbindung bringen. Eine hohe Variabilität, die Betroffene den Endpunkten beimessen, oder eine größere Unsicherheit darüber, wie wichtig die Endpunkte sind, führte zu einer größeren Unsicherheit der Evidenz (Abbildung 3).

Benötigte Evidenz und Beispiele

Als geeignete Evidenz würden Studien benötigt, in denen die Entscheidungen über die Verwendung von Emicizumab und Bedaquilin bei den entsprechenden Patientenpopulationen auf der Basis von fundierten Entscheidungen und eine Erklärung der erwarteten Ergebnisse beurteilt würden.^{17,18} Diese Arten von Studien, denen eine informierte Entscheidungsfindung zugrunde liegt, so genannte Informed-Choice-Studien, liefern Informationen über die zugrunde liegenden Wertvorstellungen und Präferenzen. Alternativ könnten Beobachtungsstudien eine hohe Vertrauenswürdigkeit der Evidenz in diesen Patientenpopulationen schaffen, die die relative Bedeutung der Ergebnisse direkt erfassen.^{19,20} Solche Studien gibt es zwar, aber normalerweise wird in diesem Bereich mehr Evidenz benötigt,^{19,20} die explizit auch im ICER-Bericht verwendet wurde.¹³

In Situationen, in denen keine Evidenz verfügbar ist, können mehrere Strategien verwendet werden, um die Evidenz zu beschaffen, oder man kann den Input des Entscheidungsgremiums für weitere Überlegungen verwenden, obwohl diese nur in seltenen Fällen zu einer hohen Vertrauenswürdigkeit der Evidenz führen würden. In der WHO MDR-TB Richtlinie für Bedaquilin hatte die Richtliniengruppe „das Gefühl, dass es potenziell große Variationen bei den Wertvorstellungen und Präferenzen der Patienten bei den einzelnen Endpunkten gab. Die meisten Mitglieder hatten das Gefühl, dass die Patienten großen Wert auf das Überleben legen würden, es jedoch weniger eindeutig sei, dass die Patienten eine mikrobiologische Kulturkonversion gleichermaßen wertschätzten.“

4.6. Abwägen der erwünschten und unerwünschten Effekte

Je größer der Netto-Gesundheitsnutzen bzw. -schaden ist,

umso wahrscheinlicher ist eine starke Empfehlung für oder gegen die entsprechende Option bzw. Intervention. Für diese Beurteilung müssen die erwünschten und unerwünschten Gesundheitseffekte unter Berücksichtigung ihrer Summe, die unterschiedliche Gewichtung deren Bedeutung sowie die Vertrauenswürdigkeit der Evidenz zu den Gesundheitseffekten einer Intervention und der relativen Bedeutung der Endpunkte gegeneinander abgewogen werden.

Benötigte Evidenz und Beispiele

Die Evidenz für dieses Kriterium stammt von der Beurteilung von Nutzen und Schaden, der Wertvorstellung und Vertrauenswürdigkeit der Evidenz. Somit ist das Gleichgewicht zwischen Nutzen und Schaden eine Kombination der numerischen (absolut) Effektgrößen und der entsprechenden relevanten Bedeutung der Endpunkte. Obwohl dies nur selten gemacht wird, sollte die Gesamtvertrauenswürdigkeit der Evidenz das Ergebnis der Beurteilung der Vertrauenswürdigkeit der Evidenz für die Effekte einer Intervention und der zugrunde liegenden Wertvorstellungen und Präferenzen sein. Im oben beschriebenen Fall wäre die Vertrauenswürdigkeit bei beiden Beispielen niedrig bis sehr niedrig, da die Evidenz bei einigen wichtigen Endpunkten (Blutungen und seltene unerwünschte Nebenwirkungen in Beispiel 1 bzw. Mortalität und Heilung in Beispiel 2) niedrig oder sehr niedrig eingestuft wurde.

4.7.-4.9. Ressourcenbedarf

Dieser Punkt beschreibt, wie Ressourcen-intensiv eine Option ist, ob sie kosteneffizient ist, und ob es hinsichtlich der Kosteneffektivität einen inkrementellen Nutzen gibt. Je vorteilhafter oder eindeutig ungünstig dieser Ressourcenbedarf ist, umso wahrscheinlicher ist eine starke Empfehlung.

Benötigte Evidenz und Beispiele

Für die drei Kriterien würden erste Beobachtungsstudien oder – noch besser – Studien benötigt, in denen die Kosten in der tatsächlichen Interventionsstudie ermittelt werden.²¹⁻²³ Die Beurteilung der Vertrauenswürdigkeit erfolgt nach den GRADE-Prinzipien, die auch für die Verwendung und Kosten von Ressourcen gelten.^{22,23} Für Kosten-Nutzen-Studien sollten geeignete, glaubwürdige und vollständige Modelle entwickelt werden, die die geschätzten Effekte der systematischen Überprüfungen nutzen, die bei den EtD-Ansätzen verwendet wurden, die entsprechenden Schätzungen der relativen Bedeutung der Endpunkte. Der HTA-Bericht vom Institut für klinische und wirtschaftliche Überprüfung (engl. Institute for Clinical and Economic Review, ICER)¹³ enthält auch Kosten-Nutzen-Analysen. Die Gesamtvertrauenswürdigkeit bei diesem Modell wäre in der Regel nicht höher als die seiner Input-Variablen, die aus der Evidenz des geschätzten Effekts und der relativen Bedeutung der Endpunkte abgeleitet werden. Das ICER schlussfolgerte jedoch, dass die „Ergebnisse aus unserer Kosten-Nutzen-Analyse zeigen, dass Emicizumab eine vorherrschende Strategie ist“.

Für das Bedaquilin-Beispiel wurde eine Kosten-Nutzen-Analyse durchgeführt, und die Richtliniengruppe schlussfolgerte, dass „während das Kosten-Nutzen-Modell einen Gesamtnutzen ergab, bestanden Bedenken bezüglich der verwendeten vereinfachenden Annahmen (z.B. keine Berücksichtigung des Unterschieds bei den schwerwiegenden Nebenwirkungen, keine Berücksichtigung des Effekts auf die Übertragung, Unsicherheit hinsichtlich der Anwendung von Studienendpunkten – einschließlich Todesfällen – auf routinemäßige programmatische Bedingungen etc.). Die Richtliniengruppe war zudem der Auffassung, dass das Kosten-Nutzen-Verhältnis nicht automatisch dazu führt, dass Bedaquilin bei den potenziell hohen Kosten bezahl-

bar ist bzw. ein Land bereit ist, die Kosten zu übernehmen.“

4.10. Zugangsgerechtigkeit

Interventionen sollten einen ähnlichen Nutzen für alle Populationen haben, für die eine Entscheidung oder Empfehlung gilt. Wenn die Effekte für Populationsgruppen unterschiedlich sind, insbesondere bei solchen, die in manchen Gruppen der Population benachteiligt sind oder Nachteile mit sich bringen, kann dies zu Benachteiligungen kommen. Je größer die Wahrscheinlichkeit ist, Benachteiligungen zu reduzieren und eine größere Gleichbehandlung zu schaffen und je leichter eine Option zugänglich ist, umso wahrscheinlicher ist es, dass eine starke Empfehlung ausgesprochen wird.

Benötigte Evidenz und Beispiele

Die GRADE-Methodik schlägt vor, die Auswirkungen von Interventionen auf die Zugangsgerechtigkeit im Gesundheitswesen anhand von fünf Verfahren explizit zu ermitteln, wenn hier seitens der Interessengruppen Bedenken bestehen: (1) Berücksichtigung der Zugangsgerechtigkeit im Gesundheitswesen als Endpunkt; (2) Berücksichtigung der wichtigen Patientenendpunkte als relevante Faktoren für die Zugangsgerechtigkeit im Gesundheitswesen; (3) Beurteilung der Unterschiede bei der relativen Effektgröße einer Behandlung; (4) Beurteilung der Unterschiede beim Ausgangsrisiko und der unterschiedlichen Auswirkungen auf die absoluten Effekte; (5) die Indirektheit der Evidenz für benachteiligte Populationen und/oder Settings.²⁴

Häufig stammt diese Art der Evidenz aus Beobachtungsstudien, in denen die Bedeutung von Ergebnissen beurteilt wird, oder aus randomisierten und Beobachtungsstudien, bei denen die Effektveränderung beurteilt wird. In beiden Beispielen ist nicht klar, welche Auswirkung die Interventionen auf die Gleichbehandlung haben. Obwohl die rele-

vanten Endpunkte in beiden Szenarien erscheinen, so bleiben einige Fragen offen. So kann es bei einer seltenen Erkrankung aufgrund der geringen Anzahl an Patienten, die behandelt werden müssen, beispielsweise einen leichteren Zugang für alle Patienten geben; der Zugriff könnte jedoch auch erschwert sein, wenn nur Patienten mit einer Anbindung an spezialisierte Zentren oder solche, bei denen eine Kostenübernahme besteht, Zugang haben. Die WHO-Richtliniengruppe hatte das Gefühl, dass „die Effekte auf die Zugangsgerechtigkeit durch die Ergänzung von Bedaquilin zur von der WHO empfohlenen MDR-TB Behandlung aufgrund der Unsicherheit bezüglich der Bezahlbarkeit und der Bereitschaft des jeweiligen Landes zur Kostenübernahme sowie der oben erwähnten unterschiedlichen Einschätzung von Nutzen und Schaden schwierig zu beurteilen sind“. Das könnte bedeuten, dass nicht alle Patienten die Behandlung bekommen, wenn eine Empfehlung ausgesprochen wird.²⁵

4.11. Akzeptierbarkeit

Beim Begriff Akzeptierbarkeit geht es darum, ob es wichtige Interessengruppen gibt, die die Verteilung von Nutzen und Schaden, Kosten und zugeordneten Wertvorstellungen wahrscheinlich nicht akzeptieren würden. Weitere Überlegungen hierzu sind, ob die Intervention die Autonomie der Betroffenen beeinträchtigen würde oder ob wichtige Interessengruppen die Intervention aus moralischen Gründen ablehnen würden (z. B. aufgrund ethischer Prinzipien wie beispielsweise den Grundsätzen zur Schadensvermeidung, Wohltätigkeit oder Gerechtigkeit). Je höher die Akzeptierbarkeit einer Option bei allen oder den meisten Interessengruppen ist, umso wahrscheinlicher ist es, dass eine starke Empfehlung ausgesprochen wird.

Benötigte Evidenz und Beispiele

Die Art der benötigten Evidenz sind Interventionsstudien, bei denen die Akzeptierbarkeit gemessen wird, oder Beobachtungsstudien zur praktischen Anwendung der Interventionen. Umfragen bei den relevanten Interessengruppen (Patienten, Gesundheitsdienstleistern, politischen Entscheidungsträgern etc.) könnten informativ sein und Forschungsevidenz liefern, die wiederum genutzt werden könnte. Studien zur Einnahmedisziplin der Patienten können indirekte Evidenz dafür liefern, dass die Interventionen akzeptiert werden. Die GRADE-Bereiche für die Vertrauenswürdigkeit der Evidenz sollten für dieses Kriterium gelten, aber GRADE hat bislang noch keine expliziten Anwendungsleitlinien für deren Beurteilung geliefert.

Derzeit wird im Rahmen von Umfragen Evidenz für Emicizumab erfasst. Die WHO-Richtliniengruppe hat dieses Kriterium in der Richtlinie von 2013 nicht explizit berücksichtigt, hatte jedoch in der Fassung von 2017 einige Evidenz zusammengetragen und erläutert, dass „die Effekte auf die Zugangsgerechtigkeit durch die Ergänzung von Bedaquilin zur von der WHO empfohlenen MDR-TB Behandlung aufgrund der Unsicherheit bezüglich der Bezahlbarkeit und der Bereitschaft des jeweiligen Landes zur Kostenübernahme sowie der oben erwähnten unterschiedlichen Einschätzung von Nutzen und Schaden schwierig zu beurteilen sind“. Die angegebene Evidenz stammt aus Vietnam und Weißrussland.²⁶

4.12. Machbarkeit

Der Begriff Machbarkeit („feasibility“) bezieht sich darauf, ob eine Intervention bzw. Option tragfähig ist, oder ob es wesentliche Hemmnisse gibt, die die praktische Umsetzbarkeit der Intervention (oder Option) wahrscheinlich einschränken oder bei der Implementierung berücksichtigt werden müssen. Je höher die Machbarkeit einer Option bei

allen oder den meisten Interessengruppen ist, umso wahrscheinlicher ist es, dass eine starke Empfehlung ausgesprochen wird.

Benötigte Evidenz und Beispiele

Ähnlich wie bei der Evidenz zur Akzeptierbarkeit kann die Evidenz hier aus Beobachtungsstudien und Umfragen zur praktischen Anwendung der Interventionen abgeleitet werden. Weitere Überlegungen beziehen sich auf logische Konsequenzen (z. B. wenn eine neue Intervention nicht zugelassen ist, ist es eventuell nicht sinnvoll, sie anzuwenden). Für viele medikamentöse Interventionen lassen sich die Überlegungen zur Machbarkeit von indirekter Evidenz zu ähnlichen Interventionen oder aus diesen logischen Überlegungen ableiten. Diese Evidenz muss noch zusammengetragen werden, es ist jedoch wahrscheinlich, dass die Verwendung von Emicizumab praktikabel ist. Die WHO-Richtliniengruppe, die Bedaquilin beurteilte, brachte in der Richtlinie von 2013 ihr Gefühl zum Ausdruck, dass es aufgrund der Beobachtungen und Berichte aus nationalen Tuberkuloseprogrammen, die die Machbarkeit bestätigt haben, sinnvoll sei, die Verwendung von Bedaquilin einzuführen; die Gruppe traf jedoch keine einstimmige Entscheidung.²⁶

5. Lösung der Beispiele

Beispiel 1 – Hämophilie A

Ich habe zwei HTA-Berichte über zwei wichtige Studien evaluiert, in denen Emicizumab zur Behandlung von Hämophilie A bei Patienten mit Inhibitoren beurteilt wurde.^{12,}

^{13,27} Hierbei ergaben sich unterschiedliche Schlussfolgerungen. Der Gemeinsame Bundesausschuss (G-BA) schlussfolgerte, dass es Hinweise auf einen nicht quantifizierbaren Netto-Zusatznutzen von Emicizumab für Patienten gibt, für die keine Alternativen verfügbar sind. Der ICER schlussfol-

gerte, dass „die Ergebnisse unserer Analyse darauf schließen lassen, dass eine Emicizumab-Prophylaxe Vorteile hinsichtlich der qualitätsadjustierten Lebensjahre bietet und über die Lebenszeit des Patienten mit wesentlich geringeren Kosten verbunden ist, wobei diese Ergebnisse bei unterschiedlichen Sensitivitäts- und Szenarioanalysen stabil blieben.“¹³

Bei der Empfehlung eines Gremiums in einer Richtlinie kann es sich um eine bedingte Empfehlung auf der Basis einer sehr niedrigen bis niedrigen Vertrauenswürdigkeit der Evidenz für die Behandlungseffekte zur Verwendung von Emicizumab bei Hämophilie A-Patienten mit Inhibitoren handeln. Dies wäre primär durch die fehlende höhere Vertrauenswürdigkeit der Evidenz für die Effektschätzer, Akzeptierbarkeit für manche Interessengruppen und Kosten in unterschiedlichen Settings bedingt.

Beispiel 2 – Multiresistente Tuberkulose

Die WHO-Richtliniengruppe sprach 2013 und 2017 eine bedingte Empfehlung für die Verwendung von Bedaquilin aus.^{25,26} Die primären Gründe waren erste Bedenken hinsichtlich der unerwünschten Nebenwirkungen in den randomisierten Studien, die laut einer Metaanalyse von individuellen Patientendaten in der aktualisierten Fassung von 2017 etwas niedriger waren.²⁸ Kosten und Machbarkeit waren weitere Gründe, auf die in den Leitlinienberichten detailliert eingegangen wurde. Ein interaktiver EtD-Ansatz für das Bedaquilin-Beispiel findet sich hier: <http://bit.ly/2GPgze7> (siehe Tabelle mit interaktiver Zusammenfassung der Ergebnisse). Ähnliche Präsentationen wie beim Bedaquilin-Beispiel würden ein besseres Verständnis für die Informationen in HTA-Berichten schaffen, insbesondere für die Tabelle mit der Zusammenfassung der Ergebnisse, mit der manche Fachzeitschriften die Ergebnisse von systematischen Übersichtsarbeiten vorstellen.^{29,30}

6. Auswirkungen für die Praxis und Entscheidungsfindung

Um informierte Entscheidungen über die EtD-Kriterien treffen zu können, müssen Einschätzungen vorgenommen werden. Selbst bei den strengsten methodischen Ansätzen treffen sich Entscheidungen nicht von alleine, sie können aber transparent gemacht werden. Bei den beiden beschriebenen Beispielen wären Entscheidungsträger – trotz der Bedenken hinsichtlich der Evidenz, die die Auswirkungen auf die Ergebnisse zeigen, beim EtD-Ansatz – zuversichtlich bzw. könnten zuversichtlich sein, dass sie zum Zeitpunkt der Entscheidung eine informierte Entscheidung treffen. Angesichts der Forschungslücken, die sich im EtD-Ansatz ergäben, ist klar, dass sich Entscheidungen zwar ändern können, aber nicht das Vertrauen in diese Entscheidungen – das groß sein kann, wenn ein Entscheidungsträger das Gefühl hat, dass die gesamte relevante Forschungsevidenz, weitere Überlegungen und Entscheidungen transparent verwendet und beschrieben wurden, obwohl sich die Art der Entscheidung (Kostenübernahme oder starke statt bedingte Entscheidungen) nach Überprüfung von neuer oder zusätzlicher Evidenz ändern kann.

7. Fazit

Empfänger von medizinischen Interventionen verdienen einen transparenten Ansatz bei der Entscheidungsfindung, die das Ergebnis von zweckdienlichen medizinischen Fragestellungen ist. Durch die Verwendung von transparenten Rahmenbedingungen werden sichere Entscheidungen ermöglicht. Für die Kriterien in den EtD-Ansätzen werden unterschiedliche Arten an Evidenz benötigt, um die relevante Fragestellung zu beantworten, wie z. B. Evidenz aus Beobachtungsstudien und aus Umfragen und Aufgaben zur Prioritätensetzung. Zur Beurteilung der Effekte von Interventionen liefern randomisierte kontrollierte Studien

²⁰Naraine VS, Risebrough NA, Oh P, Blanchette VS, Lee S, Stain AM, et al. Health-related quality-of-life treatments for severe haemophilia: utility measurements using the Standard Gamble technique. *Haemophilia*. 2002;8(2):112-20.

²¹Parmelli E, Amato L, Oxman AD, Alonso-Coello P, Brunetti M, Moberg J, et al. GRADE EVIDENCE TO DECISION (EtD) FRAMEWORK FOR COVERAGE DECISIONS. *Int J Technol Assess Health Care*. 2017:1-7.

²²Perleth M, Matthias K, Langer G, Meerpohl JJ, Gartlehner G, Kaminski-Hartenthaler A, et al. [GRADE guidelines: 10. Considering resource use and rating the quality of economic evidence]. *Z Evid Fortbild Qual Gesundheitsw*. 2012;107(3):256-68.

²³Guyatt GH, Oxman AD, Kunz R, Jaeschke R, Helfand M, Liberati A, et al. Incorporating considerations of resources use into grading recommendations. *BMJ*. 2008;336(7654):1170-3.

²⁴Welch VA, Akl EA, Pottie K, Ansari MT, Briel M, Christensen R, et al. GRADE equity guidelines 3: health equity considerations in rating the certainty of synthesized evidence. *J Clin Epidemiol*. 2017.

²⁵World-Health-Organization. The use of bedaquiline in the treatment of multidrug-resistant tuberculosis. Interim policy guidance http://apps.who.int/iris/bitstream/10665/75146/1/9789241548441_eng.pdf. Accessed 8 Dec 2014. 2013.

²⁶World-Health-Organization. Report of the Guideline Development Group Meeting on the use of bedaquiline in the treatment of multidrug-resistant tuberculosis. Report of the Guideline Development Group Meeting on the use of bedaquiline in the treatment of multidrug-resistant tuberculosis, A review of available evidence (2016). Geneva: World Health Organization. Licence: CC BY-NC-SA 3.0 IGO Accessed July 15, 2017. 2017 [

²⁷Gesundheitswesen IGWiG. Efficizumab (Hämophilie A) – Nutzenbewertung gemäß § 35a SGB V [Available from: <https://www.iqwig.de/en/projects-results/projects/drug-assessment/a18-20-emicizumab-haemophilia-a-benefit-assessment-according-to-35a-social-code-book-v-9393.html>.

²⁸Mbuagbaw L, Guglielmetti L, Hewison C, Bakare N, Bastard M, Caumes E, et al. Outcomes of Bedaquiline Treatment in Patients with Multidrug-Resistant Tuberculosis. *Emerg Infect Dis*. 2019;25(5):936-43.

²⁹Schunemann HJ, Santesso N, Brozek JL. Interactive Summary of Findings tables: the way to present and understand results of systematic reviews. *JBI Database System Rev Implement Rep*. 2019;17(3):259-60.

³⁰Chu DK, Wood RA, French S, Fiocchi A, Jordana M, Wasserman S, et al. Oral immunotherapy for peanut allergy (PACE): a systematic review and meta-analysis of efficacy and safety. *Lancet*. 2019;393(10187):2222-32.

³¹Schünemann HJ, Lerda D, Dimitrova N, Alonso-Coello P, Gräwingholt A, Quinn C, et al. On behalf of the ECIBC contributor group. *Annals of Internal Medicine*. 2019 (epub).10.7326/M18-3445.

³²Schunemann HJ, Cushman M, Burnett AE, Kahn SR, Beyer-Westendorf J, Spencer FA, et al. American Society of Hematology 2018 guidelines for management of venous thromboembolism: prophylaxis for hospitalized and nonhospitalized medical patients. *Blood Adv*. 2018;2(22):3198-225.

Evidenzgewinnung und -beurteilung am Beispiel Hämatologie und Onkologie

Prof. Dr. Bernhard Wörmann | Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie (DGHO) und Charité Universitätsmedizin Berlin, Klinik für Hämatologie, Onkologie und Tumorimmunologie

Die Bewertung der Wirksamkeit und des Nutzens neuer Arzneimittel geschieht auf dem Boden der evidenzbasierten Medizin. Die Welle neuer Arzneimittel in der Hämatologie und in der Onkologie kann zu einer Verbesserung von Prognose und Lebensqualität der Patienten führen, erfordert aber auch eine kritische Überprüfung der bisherigen Instrumente von Evidenzgewinnung und -beurteilung. Besondere Herausforderungen sind die Durchführung qualitativ hochwertiger randomisierter klinischer Studien (Randomized Clinical Trial, RCT) unter Berücksichtigung aller patientenrelevanten Endpunkte auch bei seltenen Erkrankungen (Orphan Diseases), der Verzicht auf RCT bei Vorliegen überzeugender Daten aus einarmigen Studien, die kritische Überprüfung von Testverfahren einschließlich der Nutzung neuer Methoden zur Analyse von Biomarkern sowie die Sicherung des Bezugs von Studienendpunkten in frühen Phase I/II-Studien zum langfristigen Therapieziel.

Einleitung
Die frühe Nutzenbewertung neuer Arzneimittel nach dem Arzneimittelmarktneuordnungsgesetz (AMNOG) ist inzwischen ein etabliertes Verfahren.¹ Die wissenschaftlichen medizinischen Fachgesellschaften begleiten den Prozess sowohl aus der methodisch-wissenschaftlichen Sicht als auch aus der ärztlichen Perspektive.^{2,3} Hämatologie und Onkologie haben allein aufgrund der hohen Anzahl von Verfahren eine Vorreiterrolle in der Beurteilung und Weiterentwicklung des Prozesses der frühen Nutzenbewertung übernommen (siehe Abbildung 1).

Dazu kommt, dass die Onkologie in der Medizin zurzeit ebenfalls eine zentrale Rolle in der Translation des Verständnisses pathophysiologischer Krankheitsgrundlagen in Diagnostik und Therapie einnimmt. Krebs ist heute nicht mehr eine Krankheit, sondern der Sammelbegriff für eine Vielzahl verschiedener Erkrankungen, basierend auf biologischen Einteilungen. Das beeinflusst die Therapie und auch die Strukturen klinischer Studien.

Im Folgenden wird zunächst kurz die Evidenzbeurteilung von neuen Arzneimitteln anhand der aktuellen Ergebnisse der frühen Nutzenbewertung zusammengefasst, dann ein Überblick über die bisherigen Formen der Evidenzgewinnung im Rahmen klinischer Studien gegeben und abschließend die Perspektive von Evidenzgewinnung und -bewertung angesichts der zurzeit in Entwicklung befindlichen Arzneimittel aufgezeigt.

Evidenzbeurteilung im Rahmen der frühen Nutzenbewertung

Die Mitglieder der Kommission „Nutzenbewertung neuer Arzneimittel“ bei der Arbeitsgemeinschaft der Wissenschaftlichen Medizinischen Fachgesellschaften (AWMF) haben gemeinsam mit der Deutschen Gesellschaft für Häma-

tologie und Medizinische Onkologie (DGHO) die Ergebnisse der frühen Nutzenbewertung neuer Arzneimittel aus den Jahren 2011 bis 2018 ausgewertet.⁴ Eine Besonderheit des deutschen AMNOG-Verfahrens ist die Bewertung von Subgruppen, die der Gemeinsame Bundesausschuss (G-BA) vor Beginn des Verfahrens festlegen kann. Da diese Subgruppen sowohl einen Einfluss auf die Preisbildung als auch potenziell auf die wirtschaftliche Verordnung von Arzneimitteln haben, erfolgt die Auswertung der Festlegungen entsprechend den Subgruppen. Abbildung 2 vergleicht die Gesamtheit der Ergebnisse aller abgeschlossenen Verfahren (Abbildung 2 A) mit den spezifischen Ergebnissen der Onkologie (Abbildung 2 B) und der Hämatologie (Abbildung 2 C).

Bis Ende 2018 wurden vom G-BA 344 Verfahren vollständig durchgeführt und abgeschlossen. Unter Berücksichtigung der Bildung von Subgruppen/Subpopulationen ergeben sich 684 Festlegungen. In 61 Prozent aller Subgruppen/Subpopulationen wurde der Zusatznutzen als „nicht belegt“ bewertet. 14,2 Prozent der positiven Festlegungen

zum Zusatznutzen wurden als „gering“, 11,7 Prozent als „beträchtlich“, 0,4 Prozent als „erheblich“ und 12 Prozent als „nicht quantifizierbar“ eingestuft. In 0,7 Prozent der Subgruppen/Subpopulationen wurde der Schaden höher als der Nutzen eingestuft und in die Kategorie „geringerer Zusatznutzen“ eingeordnet. In der Kategorie „nicht quantifizierbar“ finden sich vor allem Arzneimittel mit Orphan-Drug-Status.

Bei Onkologika liegt der Anteil von Verfahren mit der Festlegung „Zusatznutzen nicht belegt“ etwas unter 50 Prozent, der Anteil von Subgruppen mit der Festlegung „beträchtlicher Zusatznutzen“ bei fast 20 Prozent. Die Hämatologie zeigt ein völlig anderes Muster. Den relativ größten Anteil haben Verfahren mit der Festlegung „nicht quantifizierbarer Zusatznutzen“. Die umfassende Darstellung der großen Unterschiede zwischen den Fachgebieten ist Gegenstand einer eigenen Schrift⁴.

Evidenzgewinnung:

Qualität von Zulassungsstudien

Grundlage der frühen Nutzenbewertung sind die Zulassungsstudie(n) und ergänzende Daten. Akzeptierte Formen klinischer Studien in absteigender Wertigkeit sind⁵

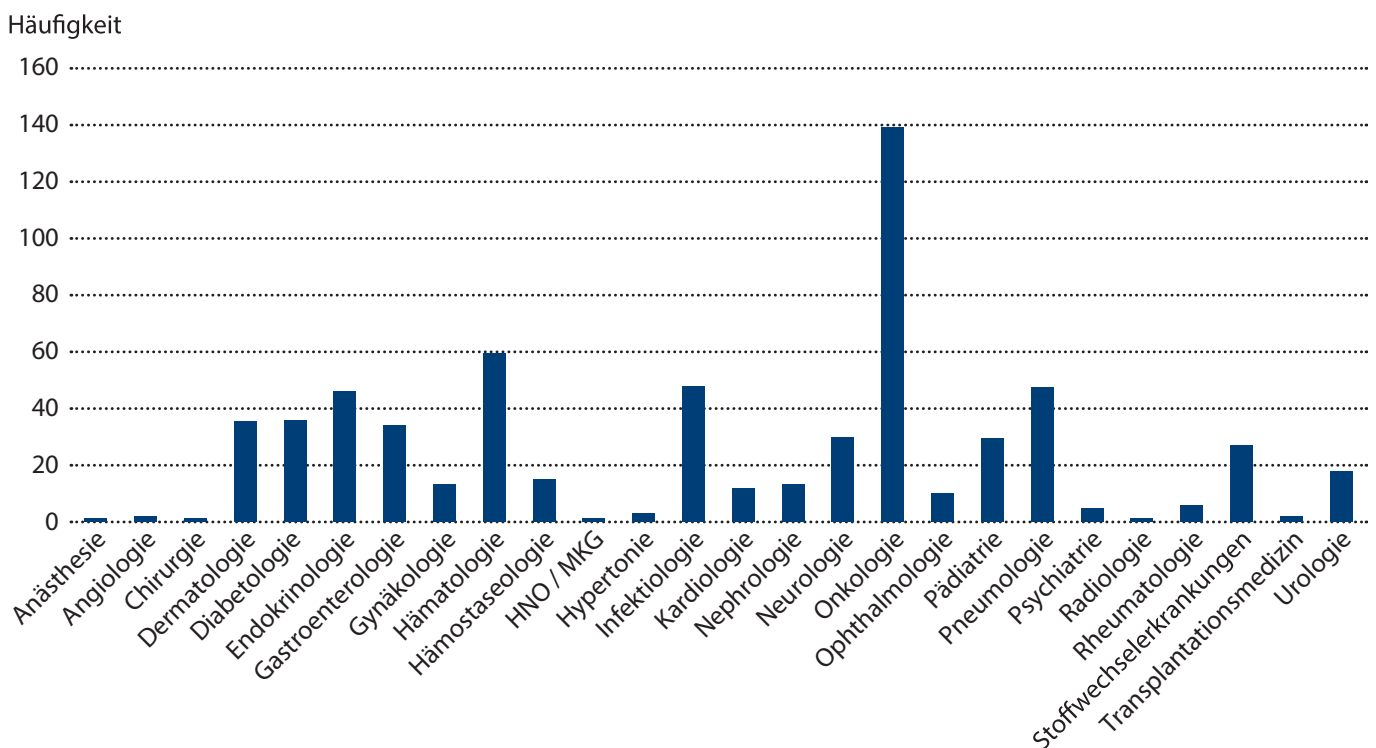
- Metaanalysen
- > zwei randomisierte klinische Studien mit konkordanten Ergebnissen
- eine randomisierte klinische Studie
- Nicht randomisierte Interventionsstudie
- Prospektive Beobachtungsstudie
- Retrospektive Beobachtungsstudie
- Fallserien / Fallberichte.

Goldstandard in der evidenzbasierten Medizin ist die randomisierte klinische Studie (RCT). Wünschenswert sind mehr als ein RCT zur selben Fragestellung. Aufgrund einer Heterogenität des Patientenkollektivs und anderer nicht



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. Ärztlich ist er in der Charité Universitätsmedizin Berlin am Campus Virchow tätig.

Medizinische Fachgebiete in der frühen Nutzenbewertung neuer Arzneimittel



Quelle: Prof. Dr. Bernhard Wörmann

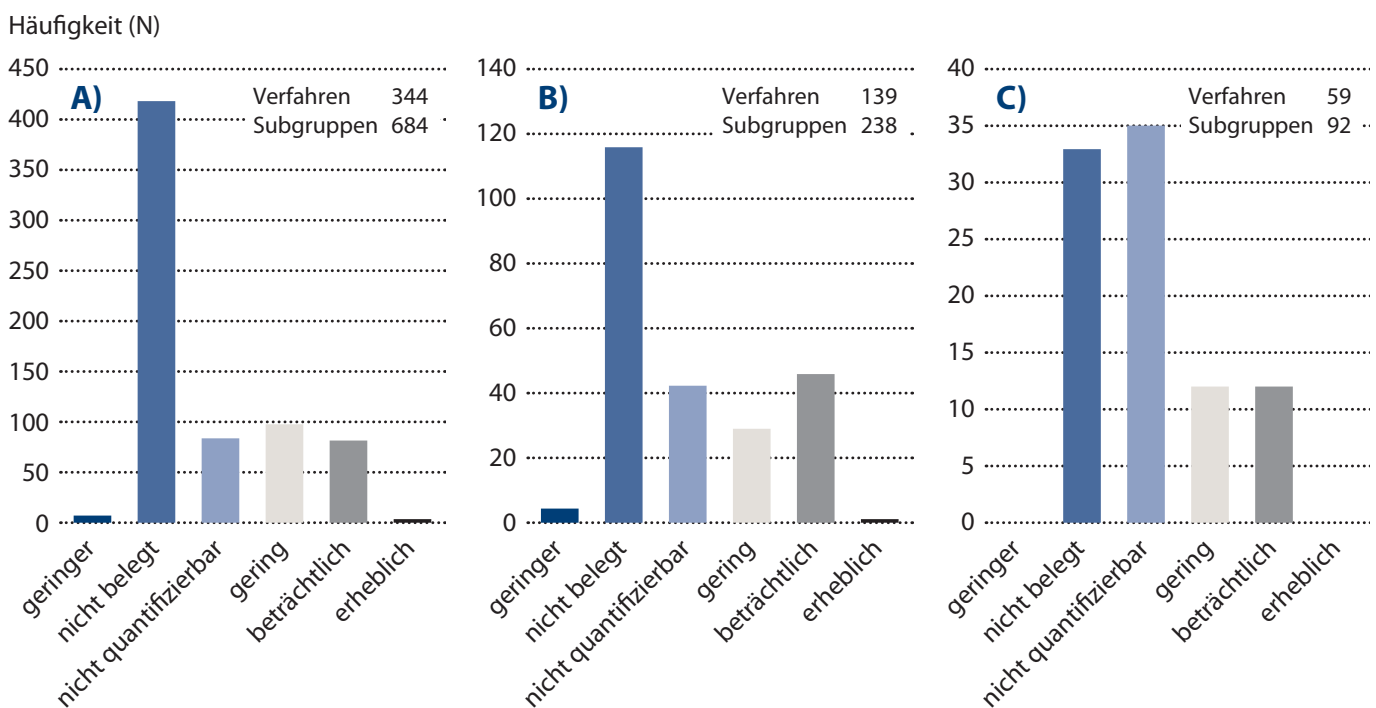
Abbildung 1: Hämatologie und Onkologie haben allein aufgrund der hohen Zahl von Verfahren eine Vorreiterrolle in der Beurteilung und Weiterentwicklung des Prozesses der frühen Nutzenbewertung übernommen.

vorgesehener Unterschiede können Studien mit identischem Design zu unterschiedlichen Ergebnissen führen. Aus finanziellen und logistischen Gründen ist die Durchführung von Studien mit identischem Design bei der Testung neuer Arzneimittel in der Onkologie eher zur Ausnahme als zur Regel geworden.

Der Anteil von RCT in der Hämatologie und in der Onkologie ist deutlich unterschiedlich. Während er in der Onkologie bei etwa 80 Prozent liegt, beträgt der Anteil in der

Hämatologie lediglich 66 Prozent. Hauptursache ist die höhere Rate von Arzneimitteln mit Orphan-Drug-Status in der Hämatologie. Allerdings ist bemerkenswert, dass auch in der Onkologie der Anteil von Arzneimitteln mit einem Orphan-Drug-Status immerhin bei 32 Prozent liegt. Das zeigt, dass auch bei Arzneimitteln für seltene Erkrankungen oft die Durchführung von randomisierten klinischen Studien möglich ist.

Festlegungen des Zusatznutzens neuer Arzneimittel in allen Fachgebieten (A), in der Onkologie (B) und der Hämatologie (C)



Quelle: Prof. Dr. Bernhard Wörmann

Abbildung 2: Berücksichtigt man alle Fachgebiete, wurde in 61 Prozent aller Subgruppen/Subpopulationen der Zusatznutzen als „nicht belegt“ bewertet. Bei Onkologika liegt der Anteil der Bewertungen mit dieser Festlegung unter 50 Prozent.

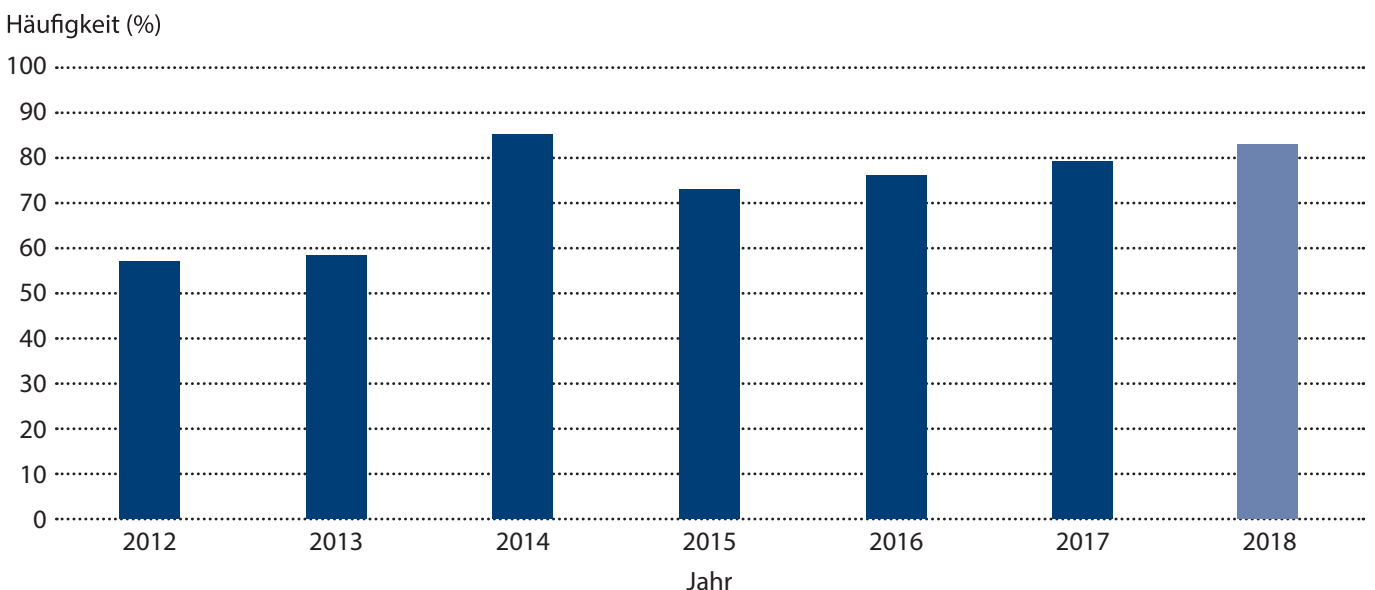
Erfassung der Lebensqualität

Grundlage der frühen Nutzenbewertung sind die Zulassungsstudie(n) und ergänzende Daten. Patienten-relevante Endpunkte bei der Arzneimitteltherapie orientieren sich an den vier zentralen inhaltlichen Begriffen

- Mortalität / Letalität
- Morbidität
- Nebenwirkungen
- Lebensqualität.

Endpunkte der Mortalität sind Gesamtüberleben oder – bei Einbeziehung von Parametern der Morbidität – krankheitsfreies bzw. ereignisfreies Überleben. Endpunkte der Morbidität sind sehr vielfältig. Dazu gehören in der Onkologie neben krankheitsfreiem, ereignisfreiem und progressionsfreiem Überleben auch Remissionsrate, Rezidivrate, Symptomatik, Vermeidung belastender Symptome sowie bildgebende, laborchemische und zytologische/histologische Parameter.

Erfassung der Lebensqualität in den Verfahren der Onkologie



Quelle: Prof. Dr. Bernhard Wörmann

Abbildung 3: Der Anteil von Verfahren mit standardisierter Erfassung der Lebensqualität ist seit Beginn des AMNOG-Prozesses gestiegen und liegt in der Onkologie bei etwa 80 Prozent.

Lebensqualität wurde früher nicht durchgehend in allen klinischen Studien erfasst. Das hat sich in den vergangenen Jahren geändert. Als Begriff wurde „Quality of Life“ Ende der 60er Jahre des letzten Jahrhunderts aus den USA übernommen. Lebensqualität ist in höchstem Maße subjektiv. In der Medizin stellt die Erfassung von Lebensqualität den Versuch dar, das subjektive Erleben von Gesundheit und Krankheit messbar zu machen. Aus der Vielzahl von regional und national entwickelten Fragebögen haben sich inzwischen einige Instrumente durchgesetzt, wurden validiert und werden flächendeckend in internationalen Studien eingesetzt.

Der Anteil von Verfahren mit standardisierter Erfassung

der Lebensqualität ist seit Beginn des AMNOG-Prozesses gestiegen und liegt in der Onkologie bei etwa 80 Prozent (siehe Abbildung 3). Weiterhin ist leider die Erfassung der Lebensqualität in nicht-randomisierten Studien eine Ausnahme.

Vergleich der Ergebnisse von Phase-II- und Phase-III-Studien

Auch wenn die Phase-III-Studie als Goldstandard zur Erfassung des (Zusatz-) Nutzens neuer Arzneimittel akzeptiert ist, stellt sich bei hochwirksamen, neuen Arzneimitteln auf der individuellen Patientenebene die ethische Frage nach der Berechtigung einer Randomisierung. Patienten neh-

men nicht primär aus altruistischen Gründen an einer Studie teil. Sie hoffen auf das neue Medikament und darauf, nicht in den Placebo-Arm randomisiert zu werden.

In der Onkologie gab es in den vergangenen Jahren drei Beispiele, in denen die Ergebnisse einer initialen, zur Zulassung führenden Phase-II-Studie mit den Ergebnissen einer konfirmatorischen Phase-III-Studie verglichen werden konnten. Ergebnisse sind in Tabelle 1 zusammengefasst.

In den Studien zu Blinatumomab und zu Osimertinib bestätigen die Ergebnisse der Phase-III- die Ergebnisse der Phase-II-Studie, während sie bei Olaratumab divergieren. Eine mögliche Erklärung ist, dass in der Studie zu Blinatumomab bei Patienten mit rezidivierender/refraktärer Akuter Lymphatischer Leukämie (ALL) ein klinisch relativ homogenes Patientenkollektiv untersucht wurde, ebenso wie bei Osimertinib bei Patienten mit EGFR T790M + nichtkleinzelligem Lungenkarzinom (NSCLC). Dagegen wurden in die Studien zu Olaratumab bei fortgeschrittenen Weichteilgewebssarkomen mehr als 20 unterschiedliche histologische Subtypen aufgenommen.

Zu diskutieren ist, ob bei homogenen Patientenkollektiven die Durchführung von konfirmatorischen Phase-III-Studien sinnvoll und erforderlich ist.

Herausforderungen:

Rasche Identifizierung wirksamer Arzneimittel

Die Fülle von Substanzen zum möglichen Einsatz in der Onkologie und Hämatologie ist fast unüberschaubar. Abbildung 4 gibt einen Überblick über Arzneimittel, die im MD Anderson Cancer Center (Houston, Texas, USA) zurzeit für klinische Studien bei Patienten mit Akuter Myeloischer Leukämie (AML) zur Verfügung stehen:¹²

Biomarker

Der Begriff „Biomarker“ ist der Oberbegriff für sehr unter-

schiedliche Parameter, die zur Klassifikation und zur Stratifizierung von Patientengruppen, vor allem in der Onkologie, verwendet werden. Besonders relevant sind Biomarker, die in direktem Zusammenhang mit dem Wirkmechanismus des Arzneimittels stehen, zum Beispiel Kinase-Inhibitoren. Allerdings haben wir in den letzten Jahren gelernt, dass ein einzelner Biomarker, auch eine sogenannte Treiber-Mutation, sich im Kontext unterschiedlicher Malignome differenziert auswirkt. So reichen die Remissionsraten von BRAF-Inhibitoren wie Dabrafenib, Encorafenib und Vemurafenib bei Patienten mit einer BRAF V600E Mutation von <10 Prozent beim kolorektalen Karzinom¹⁵ über 50 Prozent beim Melanom¹⁶ bis zu >90 Prozent bei Patienten mit Haarzell-Leukämie (HZL).¹⁷

Darüber hinaus müssen die bisher als Standard akzeptierten DNA-basierten Analysen mittels Sequenzierung, FISH o.ä. bei einigen Aberrationen möglicherweise kritisch ergänzt werden. Kürzlich wurde zum Beispiel gezeigt, dass ALK-Translokationen in der Regel, aber nicht bei allen Patienten, zu einer Proteinexpression führen. Hier kann eine RNA-basierte ALK-Analyse die Patienten besser identifizieren, die von einer ALK-gezielten Therapie nicht profitieren können.¹⁸

Besonders schwierig sind Biomarker, die nicht direkt mit dem Pathomechanismus der gezielten Therapie assoziiert sind, sondern als Surrogatparameter fungieren. Dazu gehören zum Beispiel bei den Immuncheckpoint-Inhibitoren die Analysen von PD-L1 oder der Mutationslast als Tumor Mutational Burden (TMB).

Patienten-orientierte Endpunkte

Eine besonders große Herausforderung bei der Identifikation neuer Arzneimittel ist die Wahl des richtigen Endpunktes. Viele gezielte Arzneimittel werden erstmals bei Patienten im Rahmen von Phase-I/II-Studien allein oder als Teil

Studien zu neuen Arzneimitteln bei der Akuten Myeloischen Leukämie^[12]

Newly Diagnosed

- Fludarabine + Ara-C + G-CSF + Gemtuzumab
- ATRA and arsenic +/- Gemtuzumab
- Nivolumab + Ida + Ara-C
- Cladribine + Ida + Ara-C
- FLAG-Ida + Venetoclax
- CPX-351 + Venetoclax
- BP1001 + LD Ara-C
- Pracinostat + AZA
- Chemo +/- Uproleselan
- Venetoclax + Cladribine + LDAC + LDAC w/ AZA
- Cladribine + LD Ara-C alternating DAC
- Nivolumab + AZA
- Dexrazoxane
- MCLA-117
- AML123
- DCLL9718S + AZA
- PCM-075 + LDAC
- PDR001/MBG453 + Decitabine
- SL401 + AZA
- Venetoclax + Decitabine
- Ulocuplumab + LD Ara-C

FLT3 Mutated Only

- ASP2215 vs ASP2215 + AZA vs AZA
- AC220 + AZA or LD Ara-C
- Quinzartinib + DAC

IDH Mutated only

- FT-2102
- AG120
- AG221
- Enasidenib + AZA

Frontline AML Post-Hypomethylating Therapy for MDS

- Nivolumab + AZA
- MCLA-117
- DCLL9718S + AZA
- BP1001 + LD Ara-C
- Ulocuplumab + LD Ara-C

Secondary Leukemia

- Ruxolitinib + DAC

Salvage

FLT3 Positive at relapse

- AlloSCT Initial Salvage Therapy
- AC220 + LD Ara-C
- Quinzartinib + DAC
- DS3032b + Quizartinib
- SKI-G-801
- Venetoclax + Gilteritinib
- CPX-351 + Venetoclax
- Palbociclib
- CD33-CAR-T
- Venetoclax + Decitabine
- Cladribine + Ida + Ara-C + Sorafenib
- SY-1425
- MCLA-117
- AMG 330
- XmAb 14045
- FLAG-Ida + Venetoclax
- AMG673
- DCLL9718S + AZA
- PDR001/MBG453 + Decitabine
- Daratumumab
- DS-3032b + LD Ara-C

RAS Mutated

- AlloSCT Initial Salvage Therapy
- FLAG-Ida + Venetoclax
- Venetoclax + Decitabine
- CD33-CAR-T
- SY-1425
- MCLA-117
- AMG 330
- AMG 673
- DCLL9718S + AZA
- PDR001/MBG453 + Decitabine
- Daratumumab
- CPX-351 + Venetoclax
- DS-3032b + LD Ara-C

AML with MLL gene at 11q23 Translocations

- SY-1425
- DS-3032b + LD Ara-C

IDH Mutated

- FT-2102
- AG120
- Enasidenib + AZA
- Venetoclax + Decitabine
- CD33-CAR-T
- SY-1425
- AMG 330
- XmAb 14045
- PDR001/MBG453 + Decitabine
- Daratumumab
- AG221
- DS-3032b + LD Ara-C

All others (Regardless of mutation status)

- Alvocidib (Flavopiridol)/Ara-C/Mitoxantrone (FLAM)
- CPX-351 + Venetoclax
- AlloSCT Initial Salvage Therapy
- CD33-CAR-T
- FLAG-Ida + Venetoclax
- Pilot Study CPX-351 w/ Gemtuzumab
- Cladribine + Ida + Ara-C + Sorafenib
- Idasanutlin + Ara-C vs Ara-C + Placebo
- SY-1425
- Nivolumab + 5-AZA
- LY2606368 + Ara-C + Flu
- Venetoclax + Decitabine
- MCLA-117
- DS-3032B
- AMG 330
- XmAb 14045
- AML123
- ADCT-301
- DCLL9718S + AZA
- OX40
- GO + Glasdegib
- AMG673
- Daratumumab
- IACS-010759
- SL401 + AZA
- MGD006
- PDR001/MBG453 + Decitabine
- DS-3032b + LD Ara-C

Quelle: ^[12]

Abbildung 4: Die Fülle von Substanzen zum möglichen Einsatz in der Onkologie und Hämatologie ist fast unüberschaubar, hier dargestellt am Beispiel der Akuten Myeloischen Leukämie.

Vergleich der Ergebnisse aus Phase-II- versus Phase-III-Studien

Studie	Patienten	Kontrolle	neue Therapie	N ¹	RR ²	PFÜ/RFÜ ³ (HR) ⁴	ÜL ⁵
MT103-211 [6]	ALL ⁶ , Ph-, Refraktärität oder Rezidiv innerhalb von 12 Monaten	–	Blinatumomab	189	43	–	6,1
TOWER [7]	Ph-, Refraktärität oder Rezidiv innerhalb von 12 Monaten, zweites oder drittes Rezidiv	Standard-Chemotherapie	Blinatumomab	405	16 vs. 34 ⁷ p < 0,001	4,6 vs. 6,3	4,0 vs. 7,7 0,71 ⁸ p = 0,01
Jänne, 2015 [8] Dossier	NSCLC, EGFR T790M, nach EGFR-TKI Therapie	–	Osimertinib	400	66,1	9,7	85 % ¹⁰
Mok, 2016 [9] Dossier	EGFR T790M, nach EGFR-TKI Therapie	Cisplatin/Carboplatin + Pemetrexed	Osimertinib	419	31 vs. 71 ⁷ p < 0,001	4,4 vs. 10,1 0,37 ⁸ p < 0,001	n.b. vs n.b. ⁹ n. s. ¹¹
Tap, 2016 [10]	Weichgewebssarkome, keine Vorbehandlung mit Anthrazyklinen	Doxorubicin	Olaratumab	143	11,9 vs 18,2 n. s.	4,1 vs 6,6 0,67 p = 0,06	14,7 vs 26,5 0,46 p = 0,0003
ANNOUNCE, 2019 [11]	keine Vorbehandlung mit Anthrazyklinen	Doxorubicin	Olaratumab	460	–	6,8 vs 5,4 1,23 p = 0,042	19,7 vs 20,4 0,95 n. s.

¹ N – Anzahl Patienten; ² RR – Remissionsrate; ³ PFÜ - progressionsfreies Überleben in Monaten, RFÜ – rezidivfreies Überleben in Monaten; ⁴ HR – Hazard Ratio; ⁵ ÜL – Gesamtüberleben in Monaten oder %; ⁶ ALL – Akute Lymphatische Leukämie; EGFR T790M – Mutation im Gen für den EGF Rezeptor, NSCLC – nichtkleinzelliges Lungenkarzinom, Ph – Philadelphia Chromosom; ⁷ Rate nach 24 Wochen; ⁸ Ergebnis für Kontrolle, Ergebnis für Neue Therapie; ⁹ Hazard Ratio für Neue Therapie; ¹⁰ n.b.- nicht berechenbar; ¹¹ Überlebensrate in % nach 9 Monaten; ¹¹ n.s.- nicht signifikant;

Quelle: Prof. Dr. Bernhard Wörmann

Tabelle 1: In den Studien zu Blinatumomab und zu Osimertinib bestätigen die Ergebnisse der Phase-III- die Ergebnisse der Phase-II-Studie, während sie bei Olaratumab divergieren.

von Basket-Trials analysiert. Kurzfristiger Endpunkt ist oft die Remissionsrate. Bei aggressiv verlaufenden Malignomen wie akuten Leukämien ist eine Verlängerung der Überlebenszeit nur bei Erreichen einer Remission, in der Regel einer kompletten Remission, möglich. Bei indolent verlaufenden Malignomen ist diese Korrelation weniger eindeutig. Da eine sehr kurzzeitige Remission, wenn auch bildgebend und/oder histologisch/zytologisch bestätigt, nur von fragwürdigem Nutzen ist, wird in der großen Mehrzahl von Studien die Remission nicht als primärer Endpunkt angenommen. Dies trifft durchgehend auch auf die Festlegungen der frühen Nutzenbewertung onkologischer Arzneimittel zu.

Auch sind Endpunkte wie Verbesserung der Lebensqualität in der Regel nicht entscheidend in frühen Phase I/II-Studien. Damit werden möglicherweise Substanzen aus der weiteren klinischen Forschung frühzeitig herausgenommen, die langfristig von Nutzen für die Patienten sein könnten.

Literatur:

¹ Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz – AMNOG), Bundesgesetzblatt Teil I vom 27. 12. 2010. [http://www.bgbl.de/Xaver/start.xav?startbk=Bundesanzeiger_BGBI&bk=Bundesanzeiger_BGBI&start=/*\[@attr_id=%27bgbl-110s2262.pdf%27#_Bundesanzeiger_BGBI__%2F%2F*%5B%40attr_id%3D'bgbl110s2262.pdf'%5D__1374326855765](http://www.bgbl.de/Xaver/start.xav?startbk=Bundesanzeiger_BGBI&bk=Bundesanzeiger_BGBI&start=/*[@attr_id=%27bgbl-110s2262.pdf%27#_Bundesanzeiger_BGBI__%2F%2F*%5B%40attr_id%3D'bgbl110s2262.pdf'%5D__1374326855765)

² Frühe Nutzenbewertung neuer Arzneimittel in Deutschland – Subgruppen Definition, Analyse und Kriterienkatalog, 2016. https://www.dgho.de/informationen/gesundheitspolitische-schriftenreihe/band-8-fruehe-nutzenbewertung-subgruppen/dgho_schriftenreihe_Bd8-2016_final.pdf

³ Frühe Nutzenbewertung neuer Arzneimittel in Deutschland 2011-2017, April 2018. https://www.awmf.org/fileadmin/user_upload/Service/Publikationen/AWMF_AMNOG_210x297_365_f_web_ok.pdf

⁴ Frühe Nutzenbewertung von Arzneimittel in Deutschland 2011-2018,

Gerechtigkeit und Nachhaltigkeit, Mai 2019. https://www.awmf.org/fileadmin/user_upload/Service/Publikationen/AWMF_AMNOG_210x297_365_f_web_ok.pdf

⁵ Nutzenbewertung neuer Arzneimittel der Onkologie und Hämatologie. Eine Standortbestimmung. https://www.dgho.de/publikationen/schriftenreihen/fruehe-nutzenbewertung/dgho_gpsr_arzneimittelnutzenbewertung.pdf

⁶ Topp MS, Gökbüget N, Stein AS et al.: Safety and activity of blinatumomab for adult patients with relapsed or refractory B-precursor acute lymphoblastic leukaemia: a multicentre, single-arm, phase 2 study. *Lancet Oncol* 16:57-66, 2015. DOI: 10.1016/S1470-2045(14)71170-2

⁷ Kantarjian H, Stein A, Gökbüget N et al.: Blinatumomab versus Chemotherapy for Advanced Acute Lymphoblastic Leukemia. *N Engl J Med* 376:836-847, 2017. DOI: 10.1056/NEJMoa1609783

⁸ Jänne PA, Yang JC, Kim DW et al.: AZD9291 in EGFR inhibitor-resistant non-small-cell lung cancer. *N Engl J Med* 372:1689-1699, 2015. DOI: 10.1056/NEJMoa1411817

⁹ Mok TS, Wu YL, Ahn MJ et al.: Osimertinib or Platinum–Pemetrexed in EGFR T790M-positive lung cancer. *N Engl J Med* 376:629-640, 2017. DOI: 10.1056/NEJMoa1612674

¹⁰ Tap WD, Jones RL, van Tine BA et al.: Olaratumab and doxorubicin versus doxorubicin alone for treatment of soft-tissue sarcoma: an open-label phase 1b and randomised phase 2 trial. *Lancet* 388:488-497, 2016. DOI: 10.1016/S0140-6736(16)30587-6

¹¹ Olaratumumab beim Weichgewebssarkom – negative Phase-III-Studie. <https://www.dgho.de/aktuelles/news/news/2019/olaratumab-beim-weichgewebssarkom-negative-phase-iii-studie>

¹² <https://www.mdanderson.org/research/departments-labs-institutes/departments-divisions/leukemia/clinical-trials.html>

¹³ Woodcock J, LaVange LM: Master Protocols to Study Multiple Therapies, Multiple Diseases, or Both. *N Engl J Med* 377:62-70, 2017. DOI: 10.1056/NEJMra1510062

¹⁴ Cecchini M, Rubin EH, Blumenhthal GM et al.: Challenges with Novel Clinical Trial Designs: Master Protocols. *Clin Cancer Res* Jan 29, 2019. DOI: 10.1158/1078-0432.CCR-18-3544

¹⁵ Kopetz S, Desai J, Chan E et al.: Phase II Pilot Study of Vemurafenib in Patients With Metastatic BRAF-Mutated Colorectal Cancer. *J Clin Oncol* 33:4032-4038, 2015. DOI: 10.1200/JCO.2015.63.2497

¹⁶ Sullivan R, LoRusso P, Boerner S, Dummer R: Achievements and challenges of molecular targeted therapy in melanoma. *Am Soc Clin Oncol Educ Book* 2015:177-186. DOI: 10.14694/EdBook_AM.2015.35.177

¹⁷ Falini B, Martelli MP, Tiacci E: BRAF V600E mutation in hairy cell leukemia: from bench to bedside. *Blood* 128:1918-1927, 2016. DOI: 10.1182/blood-2016-07-418434

¹⁸ Vollbrecht C, Lenze D, Hummel M et al.: RNA-based analysis of ALK fusions in non-small cell lung cancer cases showing IHC/FISH discordance. *BMC Cancer* 18:118, 2018. DOI: 10.1186/s12885-018-5070-6

Evidenz in der Nutzen-Risiko-Bewertung von Biopharmaka

Dr. Elena Wolff-Holz und Prof. Dr. Klaus Cichutek | Paul-Ehrlich-Institut, Langen

Die Marktzulassung (MZ) eines Medikaments erfolgt auf der Basis einer sorgfältigen Bewertung von Daten, die eine qualitativ hochwertige Evidenz für ein günstiges Nutzen-Risiko-Verhältnis liefert. Diese Nutzen-Risiko-Bewertung für eine MZ wird von unabhängigen und erfahrenen Regulatoren und wissenschaftlichen Experten bei den zuständigen nationalen oder europäischen Behörden im Rahmen unterschiedlicher Verfahren durchgeführt. Die eingereichten Daten beschreiben die Zusammensetzung und Qualität des Medikaments, wobei für Biopharmaka eine detaillierte Beschreibung des Herstellungsprozesses einschließlich In-Prozess-Kontrollen benötigt wird, da die Qualität eines Biopharmakons nicht ausschließlich durch Analyse des Endprodukts geprüft werden kann. Neben nicht-klinischen und pharmakologischen-toxikologischen Ergebnissen werden Daten und Erkenntnisse aus klinischen Phase I-III Studien präsentiert. Eine Vielzahl unterschiedlicher Verfahren zur Marktzulassung bietet Flexibilität für die Zulassung von Arzneimitteln bei einem ungedeckten medizinischen Bedarf und / oder für Medikamente für besonders seltene Krankheiten (Orphan Drugs).

Marktzulassungsverfahren auf der Basis der Nutzen-Risiko-Bewertung
Die Marktzulassung (MZ) eines Medikaments erfolgt auf der Basis einer sorgfältigen Bewertung von Daten, die qualitativ hochwertige Evidenz für ein günstiges Nutzen-Risiko-Verhältnis liefert. Diese Nutzen-Risiko-Bewertung für eine MZ wird von unabhängigen und erfahrenen Regulatoren und wissenschaftlichen Experten bei den zuständigen nationalen oder europäischen Behörden im Rahmen unterschiedlicher Verfahren durchgeführt. Regulatoren führen die Nutzen-Risiko-Bewertung für neue Medikamente auf der Basis der verfügbaren Evidenz durch. In Artikel 26 der Richtlinie 2001/83/EG heißt es, dass eine Marktzulassung abzulehnen ist, wenn entweder das Nutzen-Risiko-Verhältnis des Arzneimittels nicht als günstig betrachtet wird, seine therapeutische Wirksamkeit unzureichend begründet ist, oder seine qualitative und quantitative Zusammensetzung nicht den Angaben entspricht.

Ferner heißt es in Paragraph 25 des Arzneimittelgesetzes (AMG), dass die Marktzulassung versagt werden sollte, wenn das Nutzen-Risiko-Verhältnis des Arzneimittels als ungünstig betrachtet wird. Die eingereichten Daten beschreiben die Inhalte und Qualität des Medikaments, wobei für Biopharmaka eine detaillierte Beschreibung des Herstellungsprozesses einschließlich In-Prozess-Kontrollen erwartet wird, da die Qualität eines Biopharmakons nicht ausschließlich durch Analyse des Endprodukts geprüft werden kann. Neben nicht-klinischen und pharmakologisch-toxikologischen Daten werden auch Daten und Erkenntnisse aus klinischen Phase I-III Studien präsentiert.

Für Entwickler von Arzneimitteln und Regulatoren sind wissenschaftliche Richtlinien (<https://www.ema.europa.eu/en/human-regulatory/research-development>) wichtig; sie legen fest, welche wissenschaftliche Evidenz für die

klinische Zielindikation und die jeweilige Patientenpopulation benötigt wird. Dies umfasst auch Überlegungen zu den relevanten klinischen Endpunkten, mit denen der therapeutische Nutzen und die Risiken gemessen werden. Der Ausschuss für Humanarzneimittel (engl. Committee for Human Medicinal Products, CHMP) bei der Europäischen Arzneimittelagentur (engl. European Medicines Agency, EMA) gibt nach der wissenschaftlichen Bewertung der Daten über Qualität, Sicherheit und Wirksamkeit, die im Dossier des Zulassungsantrags (engl. Marketing Authorisation Application, MAA) vorgestellt werden, durch das Mitglied der zuständigen nationalen Behörde für Arzneimittel eine Stellungnahme ab. Basis hierfür ist die Mitbewertung durch den Ausschuss für Risikobewertung der Pharmakovigilanz (engl. Pharmacovigilance Risk Assessment Committee, PRAC) und bei Arzneimitteln für neuartige Therapien

(engl. Advanced Therapy Medicinal Product, ATMP) durch den Ausschuss für neuartige Therapien (engl. Committee for Advanced Therapy Medicinal Products, CAT). Nur wenn die Vorteile des Prüfpräparats die damit verbundenen Risiken überwiegen, wird positiv über den Zulassungsantrag entschieden und eine Zulassung erteilt.

Bei der Beurteilung von Nutzen und Risiken eines bestimmten Prüfpräparats ist ein gewisser Grad an Unsicherheit hinsichtlich der gesamten Datenlage nichts Ungewöhnliches für Regulatoren. So kann es je nach den statistischen Annahmen, die der entsprechenden klinischen Studie zugrunde liegen, beispielsweise eine große Herausforderung sein, seltene unerwünschte Arzneimittelwirkungen gegenüber einer bestimmten Hintergrundinzidenz bei einer begrenzten Patientenprobengröße festzustellen (siehe Eichler et al.).



Dr. Elena Wolff-Holz ist Ärztin und Fachreferentin für Medizinische Grundsatzfragen am Paul Ehrlich Institut mit Schwerpunkt Autoimmunerkrankungen, Onkologie und Biosimilars. Zusätzlich ist sie Lehrbeauftragte für Arzneimittelforschung der J. W. Goethe-Universität Frankfurt. Sie ist Mitglied der Scientific Advice Working Party (SAWP) der European Medicines Agency (EMA) und leitet die Biosimilar Medicinal Products Working Party der EMA. Vorher war sie 15 Jahre in leitenden Positionen in Forschungs- und Entwicklungsabteilungen von Biotechnologieunternehmen beschäftigt.



Prof. Dr. Klaus Cichutek ist Präsident des Paul-Ehrlich-Instituts, Bundesinstitut für Impfstoffe und biomedizinische Arzneimittel, in Langen bei Frankfurt/Main und außerplanmäßiger Professor für Biochemie an der Johann Wolfgang Goethe-Universität Frankfurt/Main. Seit 2012 ist er Mitglied des „Expert Advisory Panel on Biological Standardization (ECBS)“ und seit 2014 im „Product Development for Vaccines Advisory Committee (PD-Vac)“ der WHO. Er ist Gründungsmitglied des Deutschen Zentrums für Infektionsforschung DZIF.

Varianten von Marktzulassungsverfahren

Standard-Marktzulassung KEINE Verpflichtung, nach der Marktzulassung Daten zur Wirksamkeit / Sicherheit (W/S) zu erzeugen	Standard-Marktzulassung MIT Verpflichtung, nach der Marktzulassung Daten zur Wirksamkeit / Sicherheit (W/S) zu erzeugen	Bedingte Marktzulassung (engl.: conditional approval)	Marktzulassung unter außergewöhnlichen Umständen (engl.: approval under exceptional circumstance)
<p><i>positives Nutzen-Risiko Verhältnis</i></p> <p>Umfangreiche Daten, die im Zulassungsdossier zur Erlangung der Marktzulassung eingereicht wurden</p> <p>Anpassung des Inhalts entsprechend der Nutzen-Risiko-Analyse möglich</p> <p>Verlängerung nach 5 Jahren</p>	<p><i>positives Nutzen-Risiko Verhältnis</i></p> <p>Umfangreiche Daten im Zulassungsdossier plus Daten, die auf Anfrage während der Beurteilungsphase geliefert wurden</p> <p>Anpassung des Inhalts entsprechend der Nutzen-Risiko-Analyse möglich</p> <p>Aufgrund der Verpflichtungen aus der Marktzulassung muss der Zulassungsinhaber eventuell weitere Daten generieren.</p> <p>Maßnahmen nach der Marktzulassung können sein</p> <ul style="list-style-type: none"> • PASS (Post Authorization Safety Study) • PAES (Post Authorization Efficacy Study) <p>Zulassungsinhaber ist verpflichtet, Obligationen einzuhalten</p> <p>Verlängerung nach 5 Jahren</p>	<p><i>positives Nutzen-Risiko Verhältnis</i></p> <p>Unzureichende Daten über W/S, aber</p> <ul style="list-style-type: none"> • ungedeckter medizinischer Bedarf • Nutzen für öffentliche Gesundheit durch sofortige Verfügbarkeit > Risiko der Unsicherheit <p>Nur möglich, wenn Arzneimittel vorgesehen ist für Behandlung von</p> <ul style="list-style-type: none"> • zu schwerer Invalidität führenden oder lebensbedrohlichen Zuständen • seltenen Erkrankungen • Notfallsituationen <p>Vom Zulassungsinhaber wird erwartet, dass er für eine Umstellung auf eine vollständige Zulassung umfangreiche Daten nachliefert</p> <p>Zulassungsinhaber ist verpflichtet, spezifische Obligationen einzuhalten</p> <p>Gültigkeit 1 Jahr, danach jährlich verlängerbar, Umstellung auf Standard-Marktzulassung nach 5 Jahren</p>	<p><i>positives Nutzen-Risiko Verhältnis</i></p> <p>Unzureichende Daten</p> <p>Nur möglich, wenn</p> <ul style="list-style-type: none"> • umfangreiche Daten aufgrund der Seltenheit der Erkrankung nicht erhoben werden können • begrenzte wissenschaftliche Erkenntnisse vorhanden sind • Erhebung fehlender Informationen unethisch ist <p>Vom Zulassungsinhaber wird NICHT erwartet, dass er fehlende Daten erzeugen kann</p> <p>Zulassungsinhaber ist verpflichtet, spezifische Obligationen einzuhalten</p> <p>Gültigkeit 5 Jahre mit jährlicher Neubeurteilung; Umstellung auf Standard-Marktzulassung nicht vorgesehen</p>

Quelle: Eigene Zusammenstellung

Bei der Nutzen-Risiko-Bewertung werden verschiedene Aspekte berücksichtigt: Jede Möglichkeit eines etwaigen Mangels hinsichtlich der pharmazeutischen Qualität des Arzneimittels wird eruiert, zusätzlich alle Erkenntnisse über die Pharmakologie und Toxikologie des Produkts, sowie die Gesamtheit der Evidenz über die Wirksamkeit und das Sicherheitsrisiko betrachtet. Schließlich wird die Möglichkeit der Risikominimierung nach MZ beurteilt, und es kann ein konkreter Risikomanagementplan (engl. Risk Management Plan, RMP) verbindlich vorgeschrieben werden.

Es gibt unterschiedliche Verfahren, die zur MZ führen können – nationale, nationale europäisierte wie z. B. gegenseitige Anerkennung (engl: Mutual Recognition) und dezentrales Verfahren und schließlich das zentralisierte Verfahren. In allen Fällen muss ein positives Nutzen-Risiko-Verhältnis festgestellt worden sein, bevor die MZ gewährt wird (siehe Abbildung).

Zudem gibt es verschiedene Arten der Marktzulassung. Bei der Standard-MZ muss dem Regulator ein umfangreicher Datensatz vorgelegt werden, und es gibt keine weiteren Anforderungen. Dem Zulassungsinhaber kann jedoch eine Verpflichtung auferlegt werden, nach erfolgter MZ weitere Daten zu erheben (Maßnahmen nach der Zulassung, engl. Post Authorization Measures, PAM), entweder hinsichtlich der Wirksamkeit (Wirksamkeitsstudie nach der Zulassung (engl. Post Authorization Efficacy Study, PAES) und/oder der Sicherheit (Sicherheitsstudie nach der Zulassung, engl. Post Authorization Safety Study, PASS). In beiden Fällen muss nach fünf Jahren die Verlängerung der Zulassung beantragt werden; hierbei kann der Inhalt der Zulassung an die aktualisierte Nutzen-Risiko-Analyse angepasst werden.

Darüber hinaus sind zwei weitere unterschiedliche regulatorische Zulassungswege vorgesehen, bei denen der Antragsteller Daten über die Wirksamkeit und Sicherheit ein-

reichen kann, die für eine Standard-Marktzulassung unzureichend wären:

(1) Bedingte Zulassung für Medikamente, die bei einem großen ungedeckten medizinischen Bedarf Einsatz finden und bei denen der Nutzen für die öffentliche Gesundheit durch eine sofortige Verfügbarkeit größer ist, als das Risiko der Unsicherheit. Das ist nur möglich, wenn damit Erkrankungen behandelt werden sollen, die ansonsten zu schwerer Invalidität führen, lebensbedrohlich sind oder bei Orphan-Produkten bzw. in Notfallsituationen. In diesen Fällen wird erwartet, dass zu einem späteren Zeitpunkt umfangreiche Daten nachgereicht werden, um eine bedingte in eine vollständige Marktzulassung umzuwandeln. Der Zulassungsinhaber (engl. Marketing Authorisation Holder, MAH) muss besondere Verpflichtungen befolgen, die bedingte Zulassung ist ein Jahr lang gültig, und man geht normalerweise davon aus, dass nach fünf Jahren die bedingte in eine Standard-Marktzulassung umgewandelt wird.

Einem aktuellen Bericht über bedingte Zulassungen in der EU zufolge, wird dieser Weg nur in zehn Prozent der Fälle genutzt, aber die Bedingungen werden meistens, d.h. in etwa 70 Prozent der Fälle innerhalb des vorgegebenen Zeitraums, erfüllt, wenn sie erteilt werden. Von 2006 bis 2016 wurden lediglich 30 bedingte Marktzulassungen erteilt (<https://www.ema.europa.eu/en/human-regulatory/marketing-authorisation/conditional-marketing-authorisation>) für Arzneimittel, die bei lebensbedrohlichen Erkrankungen (24), seltenen Erkrankungen (14) und Notfallsituationen (3) eingesetzt werden sollen.

(2) Eine andere Option, bei der der Antragsteller für die Zulassung unzureichende Daten zur Wirksamkeit und Sicherheit einreichen darf, ist die Marktzulassung unter außergewöhnlichen Umständen. Dies ist jedoch nur möglich, wenn umfangreiche Daten zum Medikament aufgrund der Seltenheit der Erkrankung nicht verfügbar sind, der wis-

Beispiele für Möglichkeiten, positive Evidenz aus regulatorischer Perspektive zu erzeugen

Produkt	(Initiale) Indikation	Kontext	Grundlage für pos. W/S	Ergebnis
Keytruda (Pembrolizumab) Imfinzi (Durvalumab)	Behandlung von fortg./unresezierbarem Melanom; Fortg./ unresezierbares NSCLC	Rolle der Biomarker definieren	USA: beschleunigtes Zulassungsverfahren (einarmige Kohorte) EU: 2 randomisierte Studien	USA: Beschleunigte Zulassung EU: Standard-Marktzulassung = Vollständige Zulassung
Zykadia (Ceritinib) Alk-Inhibitor	2 ^o L ALK-positives Fortgeschrittenes NSCLC	Ausgewählter Biomarker	Einarmige Studie	Bedingte Marktzulassung mit spezifischen Obligationen, definiert im Annex II
Yescarta (Axicabtagen-Ciloleucel)	Vorbehandeltes R/R aggressives B-Zell Non-Hodgkin Lymphom (NHL)	Klarer wissenschaftlicher Kontext	Einarmige Studie	Standard-Marktzulassung mit Verpflichtung; vorgegebene Rate; historischer Kontext
Hemlibra (Emicizumab)	Hämophilie A	Klarer wissenschaftlicher Kontext	Randomisierte Studie, wenige Patienten; intraindividuelle Kontrolle (Non-Intervention Studie, NIS)	Standard-Marktzulassung mit Verpflichtung; vorgegebene Rate; historischer Kontext
Ilaris (Canakinumab)	Cryopyrin-assoziiertes periodisches Syndrom (CAPS; IL-1 β -Überproduktion)	Seltene, chronische Erkrankung, langsam fortschreitend	Randomisierte Studie	Marktzulassung unter außergewöhnlichen Bedingungen

Quelle: www.ema.europa.eu

In der letzten Phase der klinischen Entwicklung sind doppelblinde randomisierte Studien der Goldstandard. Sie sind jedoch nicht in allen Fällen sinnvoll.

wissenschaftliche Kontext noch recht unerforscht ist oder die Erhebung der fehlenden Informationen unethisch wäre. In jedem Fall geht man aber davon aus, dass der Zulassungsinhaber die fehlenden Daten nicht erheben kann.

Auch hier ist der Zulassungsinhaber verpflichtet, spezifische Obligationen zu befolgen, und in der Regel gilt die Gültigkeitsdauer von fünf Jahren mit einer jährlichen Neubewertung. Ein Wechsel zu einer vollständigen Zulassung ist nicht vorgesehen. Die Zulassung unter außergewöhnlichen Umständen wird recht selten erteilt (zwischen 2002

und 2017 wurden nur 32 Zulassungen unter außergewöhnlichen Umständen erteilt, d.h. < zwei Prozent aller MZ); die Gründe für diese Art der Zulassung waren hier hauptsächlich die Seltenheit einer Erkrankung (78 Prozent) bzw. ethische (16 Prozent) oder sonstige Gründe (sechs Prozent). Ca. 62 Prozent der Antragsteller, die einen positiven Bescheid für ihr Medikament erhalten haben, wurden in der Entwicklungsphase ihres Produktes wissenschaftlich von der EMA beraten. Bei diesen wissenschaftlichen Beratungen können die Regulatoren in einer frühen Phase

rückmelden, welche Evidenz für die Marktzulassung benötigt wird, und so dazu beitragen, dass Patienten nicht an unnötigen oder schlecht konzipierten klinischen Studien teilnehmen.

Zusammenfassend kann man sagen, dass die regulatorischen Rahmenbedingungen viel Flexibilität bei Zulassungsentscheidungen ermöglichen. Die Vollständigkeit der Datenpakete und die Art der Erkrankung bestimmen die Art der Zulassung, und die Maßnahmen nach der Zulassung (PAM) stehen im Zusammenhang mit den Unsicherheiten, die nach der ersten Beurteilung des Nutzen-Risiko-Profiles des neuen Arzneiprodukts bestehen blieben.

Bei der Entwicklung eines neuen Arzneimittels sind in der letzten Phase der klinischen Entwicklung doppelblinde randomisierte kontrollierte klinische Studien (engl. Randomized Controlled Clinical Trial, RCT) die bevorzugte Option, die man auch als Gold-Standard bezeichnet. Sie sind jedoch unter Umständen nicht in allen Fällen notwendig oder sinnvoll, und es gibt auch andere Möglichkeiten, um positive Evidenz aus regulatorischer Perspektive zu erzeugen, was nachstehend anhand von Beispielen näher erläutert wird.

Während die US-amerikanische Zulassungsbehörde (Food and Drugs Administration, FDA) verschiedene beschleunigte Zulassungswege vorsieht, gibt es in der EU nur zwei Möglichkeiten, nämlich die bedingte Zulassung (siehe oben) und die beschleunigte Beurteilung, bei der der Zeitrahmen von 210 auf 150 Tage verkürzt ist, wenn der CHMP entscheidet, dass das Produkt von hohem Interesse für die öffentliche Gesundheit ist und eine therapeutische Innovation darstellt (Ref. J. Martinalbo et al.).

Keytruda (Pembrolizumab)

Pembrolizumab ist ein humanisierter monoklonaler Antikörper, der an den PD1-Rezeptor bindet und die T-Zell-Re-

aktion und Immunreaktion gegen den Tumor durch Hemmung der Bindung an seine Liganden PD-L1 und PD-L2 erhöht. Er wurde als erster Checkpoint-Inhibitor seiner Klasse von der FDA auf der Basis von Daten aus zwei Patientenkohorten im Rahmen einer Kohortenvergleichsstudie zugelassen (Keynote 001, Kohorte B2, B3), die aus einer Phase-I-Studie mit Patienten mit unresezierbarem oder metastasiertem Melanom hervorgegangen ist. Pembrolizumab wurde zur Primär- oder Sekundärtherapie in unterschiedlichen Stärken und Dosierungen eingesetzt. Hierbei wurden noch nie da gewesene, sehr hohe Gesamtansprechraten erzielt (ORR >20 Prozent versus <10 Prozent, engl.: Overall Response Rate). Obwohl kein prospektiver Vergleich zur verfügbaren Standardtherapie durchgeführt wurde, genügten der FDA diese ungewöhnlich hohen Ansprechraten für eine frühe MZ. Im Gegensatz dazu wartete die EU noch ab, bis weitere Ergebnisse aus zwei großen randomisierten kontrollierten klinischen Studien verfügbar waren. Bei einer Studie wurde eine Ipilimumab (IPI)-naive Population untersucht (Keynote 006) und bei der anderen Studie Patienten, die zuvor mit IPI behandelt wurden (Keynote 002). Nach Beurteilung aller Daten wurde die vollständige MZ und Kennzeichnung für die Behandlung von Patienten mit fortgeschrittenem Melanom erteilt.

Imfinzi (Durvalumab)

Die Europäische Kommission erteilt eine Zulassung für Arzneimittel auf der Basis von positiven Ergebnissen einer Untergruppe von einer größeren Patientenpopulation nur zögerlich. Wenn die Gesamtergebnisse der Studie positiv sind, aber herausgearbeitet werden kann, dass eine klar definierte Untergruppe von Patienten bei der Medikamentengabe einem inakzeptablen Risiko ausgesetzt werden würde, kann im umgekehrten Fall genau diese Untergruppe von der Marktzulassung ausgeschlossen werden. Bei-

spielsweise hat der vierte zugelassene PDL-1/PD1 Checkpoint-Inhibitor Durvalumab von der FDA aufgrund der positiven Evidenz für die Wirksamkeit bei der gesamten Patientenpopulation die Marktzulassung zur Behandlung aller Patienten mit unresezierbaren nicht-kleinzelligen Lungenkarzinomen (engl. Non-Small Cell Lung Cancer, NSCLC) Stadium III erhalten. Die europäischen Experten haben allerdings nach gründlichen Subgruppenanalysen ein negatives Risikoprofil (Hazard Ratio) bei Patienten mit einer sehr niedrigen PD-L1 Expression auf Tumorzellen (TC) von TC <1 Prozent festgestellt und diese Untergruppe wurde folglich bei der EU-Marktzulassung ausdrücklich ausgeschlossen.

Zykadia (Ceritinib)

In seltenen Fällen sind Daten aus einarmigen Studien als hinreichende Evidenz für den klinischen Nutzen ausreichend. Ceritinib ist ein ALK-positiver Inhibitor, der hauptsächlich zur Behandlung von metastasiertem NSCLC verwendet wird. Ceritinib erhielt die bedingte Marktzulassung auf der Basis von zwei einarmigen klinischen Studien, die eine Gesamtansprechrates (ORR) von 40 bis 57 Prozent (im Vergleich zu bestehender Evidenz von ca. 25 Prozent), ein progressionsfreies Überleben (engl. Progression Free Survival, PFS) von sechs bis sieben Monaten (im Vergleich zu bestehender Evidenz von ca. drei bis vier Monaten) und ein Gesamtüberleben (engl. Overall Survival, OS) von 15 bis 16 Monaten (im Vergleich zu bestehender Evidenz von weniger als 12 Monaten) gezeigt hatten. Bei den mehr als 300 untersuchten Patienten wurden zudem keine inakzeptablen Risiken beobachtet. Zum Zeitpunkt der ersten bedingten Marktzulassung im Jahr 2015 wurden zwei spezielle Verpflichtungen auferlegt, wovon eine die Beendigung einer laufenden randomisierten kontrollierten klinischen Vergleichsstudie von Ceritinib und Chemotherapie für die

Zielindikation war. Beide Verpflichtungen wurden 2017 erfüllt, sodass die bedingte MZ in eine vollständige MZ umgewandelt werden konnte.

Yescarta (Axicabtagen-Ciloleucel)

Axicabtagen-Ciloleucel ist eine gegen CD19-gerichtete Immuntherapie mit autologen T-Zellen. Das Arzneimittel wird als Gentherapeutikum eingestuft. Zielgruppe sind Patienten mit vorbehandeltem aggressiven B-Zell Non-Hodgkin Lymphom (NHL). Axicabtagen-Ciloleucel wurde auf Basis einer multizentrischen einarmigen Phase I/II Open-Label Studie (ZUMA-1) zugelassen, die mit 108 Patienten durchgeführt wurde. Der primäre Endpunkt war die ORR, die bei 62 Prozent der Patienten beobachtet wurde (versus den zu erwartenden 20 Prozent). Sekundäre Endpunkte der Studie waren die Dauer des Ansprechens (engl. Duration of Response, DOR), die 14 Monate dauerte, das Gesamtüberleben (OS), das bei mehr als 17 Monaten lag, sowie die Ausprägung der unerwünschten Ereignisse.

Bei fast allen Patienten wurde ein Zytokin-Freisetzungssyndrom (engl. Cytokine Release Syndrome, CRS) in unterschiedlicher Ausprägung festgestellt, das sich mit Tocilizumab-Infusionen behandeln lässt, sowie Neurotoxizitäten, die intensiv und engmaschig überwacht und gelegentlich behandelt werden müssen. Ein positives Nutzen-Risiko-Verhältnis konnte festgestellt werden.

Allerdings musste die Einschätzung der ORR des Kontrollarms von 20 Prozent im Rahmen des regulatorischen Marktzulassungsverfahrens durch Daten einer neuen historischen Kontrollkohorte (noch vor der Entscheidung einer potenziellen Zulassung) untermauert werden.

Hierfür wurde seitens des Antragstellers eine retrospektive gepoolte Analyse der Ergebnisse von Patienten mit therapieresistentem aggressivem NHL (n=636) durchgeführt (Crump et al., 2017), um die vorgegebene mutmaßli-

che Kontrollansprechrates von 20 Prozent zu bestätigen und einen historischen Kontext für die Interpretation der ZUMA-1 Ergebnisse zu liefern. Bei der Beurteilung von Ansprechen und Überlebensrate nach der Behandlung mit dem verfügbaren Behandlungsstandard zeigte sich, dass sich diese bei allen Evidenzparametern in der gleichen Größenordnung wie die mutmaßliche Rate bewegen, d. h. die ORR war 26 Prozent und die vollständige Ansprechrates (engl.: complete response, CR) sieben Prozent mit einem mittleren OS von 6,3 Monaten, was deutlich unterhalb dem therapeutischen Nutzen von Axicabtagen-Ciloleucel liegt. Auf der Basis dieser positiven Wirksamkeitsdaten und vorliegender günstiger Sicherheitsdaten kam es zu der Entscheidung einer positiven Nutzen-Risikoabwägung. Allerdings blieb noch eine gewisse Restunsicherheit zum Sicherheitsprofil und deshalb wurde dem Antragsteller eine nicht-interventionale Sicherheitsstudie nach der Zulassung (PASS) auferlegt. Diese Ermittlung von zusätzlichen Sicherheits- und Anwendungsdaten erfolgt im Rahmen eines Registers.

Zusammenfassend zeigt dieser Fall, dass Daten aus nur einer einarmigen Studie hinreichende Evidenz für eine MZ liefern können, wenn in der klinischen Indikation ein ungedeckter medizinischer Bedarf vorliegt, es sich um eine lebensbedrohliche Erkrankung handelt, eine Klassifizierung der Erkrankung im Vorfeld möglich ist, und der Wirkmechanismus durch eine fundierte wissenschaftliche Hypothese und/oder präklinische Daten untermauert werden kann. Darüber hinaus waren die Biomarker für die Wirksamkeit (hier: ORR) akzeptierte klinische Endpunkte und das Medikament führte im Vergleich zu historischen Daten zu beispiellos hohen Ansprechrates.

Hemlibra (Emicizumab)

Emicizumab ist ein bispezifischer monoklonaler Antikör-

per, der den aktivierten Faktor IX und Faktor X verbindet, um die fehlende Funktion des Faktor VIII (in seiner aktivierten Form) wiederherzustellen, die für eine wirksame Hämostase erforderlich ist. Emicizumab ist zur routinemäßigen Prophylaxe von Blutungsepisoden bei Patienten mit Hämophilie A mit Faktor VIII-Inhibitoren indiziert. Emicizumab hat keine strukturelle Beziehung oder Sequenzhomologie zu Faktor VIII, sodass es die Entwicklung direkter Inhibitoren gegen Faktor VIII als humanisierter, monoklonaler Antikörper weder auslöst noch verstärkt.

Die Prophylaxe mit Emicizumab wurde in einer randomisierten, multizentrischen, offenen klinischen Open-Label Studie an 109 männlichen Jugendlichen und Erwachsenen mit Hämophilie A mit Faktor-VIII-Inhibitoren geprüft, die zuvor entweder episodisch (nach Bedarf) oder prophylaktisch mit Bypassing-Präparaten (aPCC und rFVIIa) behandelt worden waren.

53 Patienten, die zuvor nach Bedarf mit Überbrückungswirkstoffen (sogenannte Bypassing-Agents, kurz BPA) behandelt worden waren, wurden im Verhältnis 2:1 in die Gruppe mit Emicizumab-Prophylaxe (n=35 Patienten) bzw. in die Gruppe ohne Prophylaxe (n=18 Patienten) randomisiert. Die beobachtete Anzahl an Blutungen, die im Laufe der Zeit mit einem Koagulationsfaktor behandelt werden mussten, war signifikant geringer, und die absolute Blutungsrate (engl. Absolute Bleeding Rate, ABR) lag bei 2,3 Prozent im Vergleich zu 23 Prozent, was einer Reduktion von 87 Prozent entspricht ($p < 0,0001$). Weitere sekundäre Endpunkte wiesen in die gleiche Richtung. So zeigte sich beispielsweise eine Reduktion der Anzahl aller Blutungen (80 Prozent Reduktion; $p < 0,0001$), spontaner Blutungen (92 Prozent, $< 0,0001$) und in ähnlicher Weise auch bei Blutungen in Gelenken und Zielgelenken, sowie bei der Beurteilung der gesundheitsbezogenen Lebensqualität und Gesundheitsstatus der Patienten.

Zudem wurden 24 Patienten, die zuvor mit prophylaktischen Bypassing-Präparaten behandelt worden waren und dann eine Emicizumab-Prophylaxe erhalten hatten, mit Patienten verglichen, die zuvor episodisch (nach Bedarf) mit Bypassing-Präparaten behandelt worden waren und vor der Aufnahme in die Studie an einer nicht-interventionellen Studie (engl. Non-Interventional Study, NIS) teilgenommen hatten. Bei der NIS handelte es sich in diesem Fall um eine Beobachtungsstudie mit dem Hauptziel, detaillierte klinische Daten über die Blutungsepisoden und die Verwendung von Hämophilie-Medikamenten bei Patienten mit Hämophilie A außerhalb einer interventionellen Studiumgebung zu sammeln. In dieser intraindividuellen Analyse führte die Prophylaxe mit Emicizumab im Vergleich zur Prophylaxe mit Bypassing-Präparaten bei der NIS zu einer statistisch signifikanten ($p=0,0003$) und klinisch relevanten Reduktion der Blutungsrate (79 Prozent) bei behandelten Blutungen.

Zusammenfassend lässt sich sagen, dass selbst wenn nur sehr wenige Patienten tatsächlich mit Emicizumab behandelt worden waren, die verfügbare Evidenz ausreichte, um eine positive Entscheidung für ein günstiges Nutzen-Risiko-Verhältnis zu treffen, insbesondere da Hämophilie als seltene, stabile Erkrankung gilt und eine Klassifizierung der Patientengruppe, die in der klinischen Studie untersucht wurde, im Vorfeld möglich war. Der Wirkmechanismus von Emicizumab wird durch eine fundierte wissenschaftliche Hypothese und/oder präklinische Daten untermauert.

Die verwendeten klinischen Endpunkte zur Bestimmung der Wirksamkeit (z. B. Annualisierte Blutungsrate etc.) sind allgemein anerkannt, und die Ergebnisse aus den Biomarker-Analysen untermauern dies. Mit Emicizumab wurden im Vergleich zur Nichtbehandlung (randomisierter Vergleich) und zu anderen prophylaktischen Behandlungen

(intraindividuellem Vergleich in einer nicht-interventionellen Beobachtungsstudie) beachtliche und beispiellose Ansprechraten erzielt, was zu einem positiven Nutzen-Risiko-Verhältnis für die vollständige Standard-Marktzulassung ohne Verpflichtung führte.

Ilaris (Canakinumab)

Canakinumab ist ein humanisierter monoklonaler Antikörper, der für die Behandlung des Cryopyrin-assoziierten periodischen Syndroms (engl. cryopyrin-associated periodic syndrome, CAPS) bei Erwachsenen und Kindern ab vier Jahren zugelassen ist. Es bindet an menschliches IL-1 β und neutralisiert dessen Aktivität, indem es seine Interaktion mit den IL-1 Rezeptoren blockiert.

Die Wirksamkeit und Sicherheit von Canakinumab zur Behandlung von CAPS wurde in einer dreiteiligen randomisierten klinischen Studie mit Canakinumab ($n=15$ Patienten) im Vergleich zum Placebo ($n=16$ Patienten) gezeigt.

In Teil 1, eine Woche nach Behandlungsbeginn, wurde bei 71 Prozent bzw. nach acht Wochen bei 97 Prozent der Patienten ein vollständiges klinisches Ansprechen erreicht. In der randomisierten Absetzungsphase (Teil 2) wurde der primäre Endpunkt definiert als der Anteil an Patienten mit einem Rezidiv bzw. Schub der Erkrankung: insgesamt zeigten 81 Prozent der Patienten, die in die Placebo-Gruppe randomisiert wurden, einen Krankheitsschub im Vergleich zu keinem Patienten (0 Prozent). Somit wurde ein enormer Behandlungserfolg erzielt, und es wurden nur wenigen schwerwiegende Nebenwirkungen (engl. serious adverse events, AEs) beobachtet. Als häufigste schwerwiegende Nebenwirkungen traten bei CAPS-Patienten Entzündungen der oberen Atemwege, Durchfall, Influenza, Kopfschmerzen und Übelkeit auf, und diese wurden im Vergleich zum Nutzen für die Patienten als akzeptabel betrachtet.

Zusammenfassend kann man sagen, dass CAPS eine extrem seltene Erkrankung mit gut vorhersehbarem Verlauf ist. Der Wirkmechanismus von Canakinumab ist hinreichend bekannt und wird durch eine fundierte wissenschaftliche Hypothese und präklinische Daten untermauert. Die zur Messung der Autoimmunaktivität verwendeten Endpunkte sind wissenschaftlich und klinisch allgemein anerkannt, und die Behandlungseffekte sind überwältigend gut; dies zeigte sich sowohl beim primären Endpunkt als auch bei den Entzündungsmarkern.

Dies ist ein Beispiel, wie man mit einem kreativen Studiendesign trotz sehr geringer Patientenzahl eine aussagekräftige randomisierte Studie durchführen kann. Aufgrund der vorgelegten Evidenz erhielt das Arzneimittel die Marktzulassung unter außergewöhnlichen Umständen, da es unethisch wäre, bei solch hoher Wirksamkeit und ohne verfügbare Behandlungsalternativen, weitere Studien durchzuführen.

Zusammenfassung

Um eine Marktzulassung zu erhalten, muss die Herstellung eines Biopharmakons gut kontrolliert und einheitlich sein, seine Qualität muss hoch sein, nicht-klinische Daten müssen den Nachweis für das Wirkprinzip (proof-of-principle), sowie pharmakologische Wirksamkeit und eine geringe bzw. gar keine Toxizität belegt haben. Auf der Basis von klinischen Daten, die üblicherweise in klinischen Studien der Phase I, II und später III erhoben werden, führen die Regulatoren eine sorgfältige Nutzen-Risiko-Bewertung durch. Vorzugsweise stammt die klinische Evidenz aus einer randomisierten klinischen Studie, aber es gibt auch Ausnahmen, insbesondere bei seltenen Erkrankungen und in Fällen, in denen es nicht sinnvoll ist, viele bzw. große klinische Studien durchzuführen. Mit Hilfe von einarmigen Studien, historischen Kontrollen und inter- und intraindividuellen

Vergleichen und oft mit einem kreativen klinischen Studiendesign können auch Daten von nur wenigen Patienten effizient genutzt werden.

Sehr häufig lassen sich mit Hilfe von Biomarkern die Patienten auswählen, die am meisten profitieren bzw. am wenigsten Schaden durch eine Behandlung erfahren würden, so dass ein positives Nutzen-Risiko-Verhältnis erzielt werden kann. Dadurch werden insgesamt weniger Ressourcen verbraucht. Gewisse Herausforderungen bleiben jedoch bestehen, wie beispielsweise die Unsicherheit hinsichtlich der Nachhaltigkeit der beobachteten Ergebnisse im klinischen Alltag sowie das Risiko, positive Effekte zu überschätzen. In manchen Fällen werden nach der Zulassung eventuell weitere Studien benötigt, und manchmal wird dies als Verpflichtung auferlegt (PASS und/oder PAES), um bestehende Unsicherheiten zu beseitigen.

Das verfügbare Regulierungssystem ist flexibel genug und bietet eine ausreichende Anzahl von Werkzeugen, mit denen eine sorgfältige wissenschaftliche Beurteilung eines Antrags zur Marktzulassung von innovativen, qualitativ hochwertigen und dringend benötigten Arzneimitteln möglich ist.

Literatur:

¹ Eichler, H.G., Aronsson, B., Abadie, E., Salmonson, T. (2010). New drug approval success rate in Europe in 2009; *Nature Reviews Drug Discovery*; Heft 9, Seiten 355–356

² EMA Geschäftsbericht 2017: https://www.ema.europa.eu/en/documents/annual-report/2017-annual-report-european-medicines-agency_en.pdf

³ Martinalbo, J., Bowen, D., Camarero, J., Chapelin, M., Démolis, P., Foggi, P., Jonsson, B., Llinares, J., Moreau, A., O'Connor, D., Oliveira, J., Vamvakas, S., Pignatti, F. (2016). Early market access of cancer drugs in the EU; *Ann. Oncol.* 27: 96-105.

⁴ Crump, M., Neelapu, S.S., Farooq, U., Van Den Neste, E., Kuruvilla, J., Westin, J., Link, B.K., Hay, A., Cerhan, J.R., Zhu, L., Boussetta, S., Feng, L., Maurer, M., Navale, L., Wiezorek, J., Go, W.Y., Gisselbrecht, C. (2017). Outcomes in refractory diffuse large B-cell lymphoma: results from the international SCHOLAR-1 study; *Blood*, 130(16):1800-1808

Eignung einarmiger Studien für die Bewertung des Zusatznutzens

Dr. Thomas Kaiser | Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG), Köln

Direkt vergleichende RCTs sind für die vergleichende Nutzenbewertung in der Regel die bestgeeignete Evidenz. Für die frühe Nutzenbewertung können allerdings auch einarmige Studien vorgelegt werden. Solche Studien wurden in der Vergangenheit indikationsabhängig in unterschiedlicher Häufigkeit vorgelegt (Hepatitis C: 78 Prozent aller Fragestellungen; Onkologie: 32 Prozent, Sonstige: 13 Prozent). Für die Nutzenbewertung ist nicht entscheidend, ob eine einzelne einarmige Studie relevant ist, sondern ob der Vergleich auf Basis einarmiger Studien zur Intervention und Vergleichstherapie relevant ist. Die Eignung eines solchen Vergleichs hängt daher wesentlich von den Daten zur Vergleichstherapie ab. Ein Zusatznutzen lässt sich aus einem solchen Vergleich zudem nur bei großen beobachteten Unterschieden ableiten. Die Durchführung einarmiger Studien zum Zwecke der Nutzenbewertung sollte auf diese wenigen Ausnahmefälle beschränkt bleiben.



intergrund

Gegenstand der frühen Nutzenbewertung ist die Bewertung des Zusatznutzens neuer Wirkstoffe gegenüber der bisherigen Standardtherapie (der zweckmäßigen Vergleichstherapie)¹. Direkt vergleichende randomisierte kontrollierte Studien (RCTs) stellen hierfür den Goldstandard dar.² Dieser Standard der evidenzbasierten Medizin ist auch in der Rechtsverordnung zur frühen Nutzenbewertung, der Arzneimittel-Nutzenbewertungsverordnung (AMNutenz-V), abgebildet.³ Sofern allerdings keine direkt vergleichenden RCTs vorliegen, kann der pharmazeutische Unternehmer (pU) auch Studien geringerer Evidenzstufen zum Nachweis des Zusatznutzens vorlegen.³

Neben nicht-randomisierten, direkt vergleichenden Studien, kann es sich hierbei z. B. auch um Studien zu einem neuen Wirkstoff handeln, die für sich genommen nicht auf den Vergleich mit der zweckmäßigen Vergleichstherapie ausgelegt sind (z. B. einarmige Studien oder einzelne Studienarme aus Studien ohne Vergleichsarm mit der zweckmäßigen Vergleichstherapie). Solche Studien können und werden in manchen Dossiers auch von den pU für die Bewertung des Zusatznutzens vorgelegt. Die Nutzenbewertung auf Basis solcher Studien ist Gegenstand des vorliegenden Artikels. Dabei werden im Einzelnen folgende Punkte adressiert:

- Konzeptionelle und allgemeine methodische Überlegungen zur Eignung einarmiger Studien für die Bewertung des Zusatznutzens
- Empirische Untersuchung zur Bedeutung einarmiger Studien bei bislang abgeschlossenen frühen Nutzenbewertungen

Nicht Gegenstand des vorliegenden Artikels ist das Spezialthema „Extrapolation bei Kindern und Jugendlichen“, für das ggf. in Dossiers ebenfalls einarmige Studien vorgelegt

werden. Ebenso sind Nutzenbewertungen zu Arzneimitteln für seltene Erkrankungen (Orphan Drugs) nicht Gegenstand des vorliegenden Artikels, da für diese der Zusatznutzen bei Marktzugang unabhängig von der vorliegenden Evidenz als belegt gilt.¹

Fragestellung einer Zusatznutzenbewertung – das PICO-Schema

Die Fragestellung einer Nutzenbewertung nach AMNOG ist durch Paragraph 35a des Sozialgesetzbuches (SGB) V und die zugehörige AM-NutzenV vorgegeben.^{1, 3} Sie lautet: Hat der neue Wirkstoff im Vergleich zur zweckmäßigen Vergleichstherapie bei Patienten, für die der neue Wirkstoff zugelassen wurde, einen Zusatznutzen hinsichtlich patientenrelevanter Endpunkte? Diese Fragestellung kann durch das in der evidenzbasierten Medizin bekannte PICO-Schema abgebildet werden: Population, Intervention, Komparator, Endpunkte. Der Vergleich ergibt sich dann dadurch, dass man innerhalb einer definierten Population die Ergebnisse von Intervention und Komparator bezüglich definierter Endpunkte vergleicht, woraus sich dann der Therapieeffekt der Intervention im Vergleich zur Kontrolle ergibt.

Es ist bekannt, dass direkt vergleichende RCTs insbesondere deshalb für einen Therapievergleich die am besten geeignete Evidenz sind, weil durch die Randomisierung faire Ausgangsbedingungen für die Intervention und die Kontrolle geschaffen werden: Die der Intervention zugeordnete Gruppe ähnelt im Mittel der der Kontrolle zugeordneten Gruppe, sodass sich etwaige Unterschiede zwischen den Therapien am Studienende nicht allein durch unterschiedliche Ausgangsbedingungen erklären lassen (z. B. schwerer erkrankte Patientinnen und Patienten in der Vergleichsgruppe). Dies ist bei Vergleichen, die auf einzelnen Armen aus unterschiedlichen Studien basieren, nicht gegeben. Bei einer Bewertung auf Basis einarmiger Stu-

dien muss daher geprüft werden, ob die untersuchten Populationen ausreichend ähnlich sind.

Zusätzlich ist offen, ob es überhaupt einarmige Studien zur Vergleichstherapie gibt, in denen die tatsächlich interessierende Population untersucht wurde. Denn während die einarmigen Studien zum neuen Wirkstoff zumeist dessen Zulassung begründen und die Einschlusskriterien dieser Studien daher mit dem zugelassenen Anwendungsgebiet kongruent sind, ist dies für die Studien der Vergleichstherapie nicht per se gegeben. Bei einem Vergleich auf Basis einzelner Arme unterschiedlicher Studien ist der pU des neuen Wirkstoffs daher darauf angewiesen, dass zur Vergleichstherapie Studien anderer Hersteller oder Studiengruppen vorliegen, die die Zulassungspopulation seines neuen Wirkstoffs untersucht haben.

Weniger im Fokus der Diskussion um die Eignung einarmiger Studien steht zumeist die Ähnlichkeit der untersuch-



Dr. Thomas Kaiser ist Arzt und Systementwickler. Nach seiner Tätigkeit als Programmierer hat er in Köln Medizin studiert und hat einige Jahre im Bereich der Inneren Medizin gearbeitet. Im Jahr 2002 hat er das Institut für evidenzbasierte Medizin in Köln mitbegründet. Seit 2004, dem Gründungsjahr des IQWiG, ist er dort als Ressortleitung Arzneimittelbewertung tätig, seit 2011 als Doppelspitze gemeinsam mit Dr. Beate Wieseler.

ten Endpunkte zwischen den eingeschlossenen Studien. Dies bezieht sich nicht allein darauf, welche Endpunkte in den zu vergleichenden Studien untersucht wurden, sondern auch, ob diese Endpunkte unterschiedlich operationalisiert wurden, und ob die Beobachtungsdauer ausreichend ähnlich ist. Auch hierfür bietet die direkt vergleichende RCT auf der Hand liegende Vorteile, weil Menge und Operationalisierung der Endpunkte für den Interventions- und Kontrollarm gleichermaßen gelten (unterschiedliche Beobachtungsdauer zwischen den Behandlungsarmen kann hingegen auch bei RCTs problematisch sein, siehe z. B.⁴). Dem gegenüber ist nicht gewährleistet, dass die Studien zur Vergleichstherapie dieselben Endpunkte untersucht haben wie die Studien zum neuen Wirkstoff, da ggf. ein anderes Studienziel verfolgt wurde. Selbst wenn gleiche Endpunkte (z. B. Schmerz) untersucht wurden, können die Endpunkte so unterschiedlich operationalisiert worden sein (z. B. deutlich unterschiedliche Schwellenwerte für „Therapieerfolg“), dass eine vergleichende Aussage mit diesen Daten nicht möglich ist.

Erschwerend kommt bei Vergleichen einzelner Arme unterschiedlicher Studien hinzu, dass die konkrete Studienlage zur Vergleichstherapie durch mangelnde Publikationsqualität ggf. unklar bleibt, da wichtige Informationen zu Population und Endpunkten fehlen (siehe z. B. die Nutzenbewertungen zu Ceritinib bei Bronchialkarzinom⁵ oder Nivolumab bei Hodgkin-Lymphom⁶).

Aus all dem wird klar, dass eine qualitativ hochwertige einarmige Studie des pU allein für die Zusatznutzenbewertung nicht ausreichend ist: Die Frage ist nicht, ob die einarmige Studie geeignet ist, sondern ob der Vergleich auf Basis einarmiger Studien zur Intervention und zur Vergleichstherapie geeignet ist. Dabei hängt die Eignung einarmiger Studien des pU ganz wesentlich von der konkreten Studienlage zur Vergleichstherapie ab.

Nicht-adjustierte Vergleiche und der „dramatische Effekt“

Wie oben beschrieben, werden bei einer RCT durch Randomisierung ähnliche Ausgangsbedingungen für die Intervention und Kontrolle geschaffen, in dem die jeweils untersuchten Gruppen im Mittel ähnliche Risiken für das Versterben, die Morbidität und Nebenwirkungen haben. Dies wird durch ähnliche Verteilung bekannter, insbesondere aber auch unbekannter Risikomerkmale gewährleistet. Eine solche Gleichverteilung der Risikomerkmale ist beim Vergleich einzelner Studienarme nicht gewährleistet. Zwar gibt es Methoden zur Adjustierung der bekannten und in den Studien untersuchten Risikomerkmale (z. B.⁷).

Aber zum einen sind diese Methoden davon abhängig, dass die relevanten Merkmale in den Studien überhaupt erhoben wurden und die patientenindividuellen Daten für eine Adjustierung vorliegen (beides ist für die Studien zur

Kasten A: Dossier zur frühen Nutzenbewertung – Struktur der Ergebnisdarstellung in Modul 4 des Dossiers

4.3.1 Ergebnisse randomisierter kontrollierter Studien

4.3.2 Weitere Unterlagen

4.3.2.1 Indirekte Vergleiche auf Basis von RCT

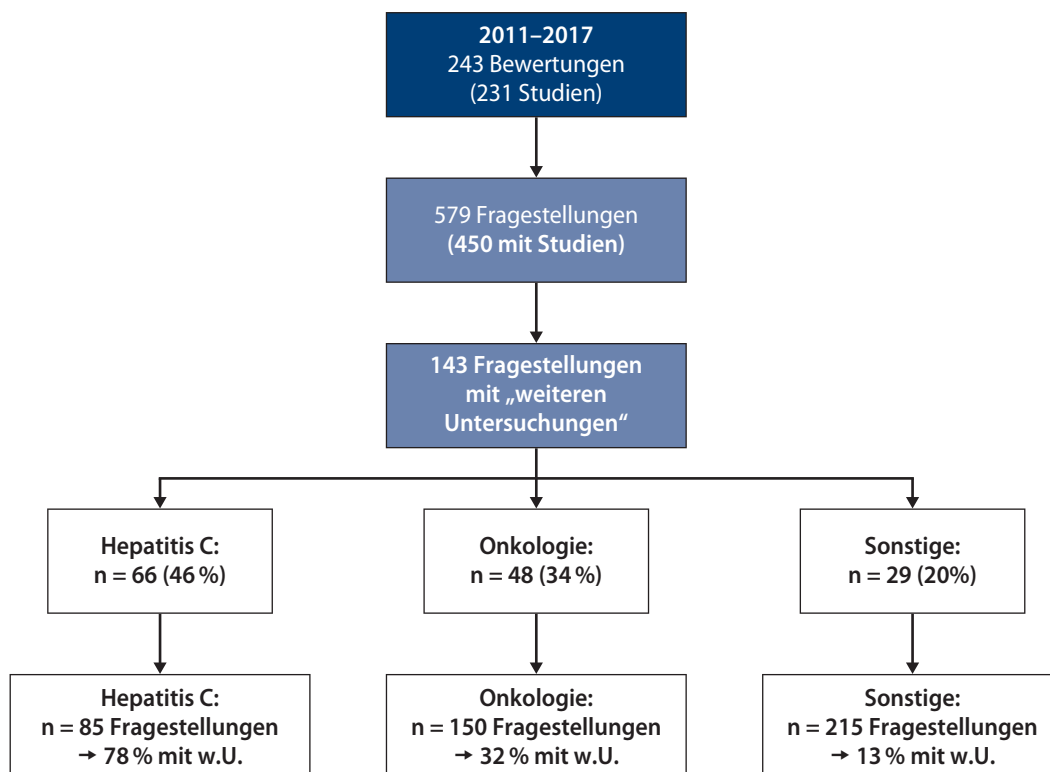
4.3.2.2 Nicht randomisierte vergleichende Studien

4.3.2.3 Weitere Untersuchungen

Quelle: Verfahrensordnung des G-BA¹⁰

Einarmige Studien sind den „Weiteren Untersuchungen“ zuzuordnen, sie werden daher im Kapitel 4.3.2.3 in Modul 4 des Dossiers dargestellt.

Häufigkeit „Weiterer Untersuchungen“ in Dossiers zur frühen Nutzenbewertung



Quelle: Eigene Analyse

Abbildung 1: Beim Vergleich der Häufigkeit, in der „Weitere Untersuchungen“ vorgelegt wurden, wird die Dominanz des Anwendungsgebiets Hepatitis C deutlich.

Vergleichstherapie oft nicht der Fall). Zum anderen lösen Adjustierungen das Grundproblem der potenziellen Ungleichverteilung unbekannter Risikomerkmale nicht. Die (hier beispielhaft zitierte) Annahme aus einer Publikation zu Adjustierungsmethoden „...fundamental assumption of no unmeasured confounders“⁷ ist in der Regel fundamental falsch.

Dennoch gibt es Situationen, in denen Vergleiche auf Basis einarmiger Studien trotz des sehr hohen Verzer-

rungsrisikos Aussagen zu Vor- oder Nachteilen einer Therapie erlauben, und zwar dann, wenn die beobachteten Unterschiede so groß sind, dass sie sich nicht mehr allein durch die Verzerrung erklären lassen. Glasziou et al. haben hierzu eine Empfehlung erarbeitet und diesen Ergebnisse von Simulationsstudien zugrunde gelegt, wonach ein beobachtetes relatives Risiko von 5 bis 10 nicht mehr plausibel allein durch Störgrößeneinflüsse erklärt werden könne.⁸ Die Grenze ist auch von den Begleitumständen abhän-

gig, z. B. der Qualität der betrachteten Studien, Unterschieden im Studiendesign [z. B. Vergleich der Ergebnisse einer unverblindeten mit denen einer verblindeten Studie] und der Konsistenz der Ergebnisse in einer Endpunktkategorie.⁹

Einarmige Studien in bislang abgeschlossenen Nutzenbewertungen: Struktur der Dossiers zur frühen Nutzenbewertung

Direkt vergleichende RCTs sind für die Bewertung des Zusatznutzens der am besten geeignete Studientyp.² Sofern solche Studien vorliegen, sind sie auch zwingend vom pU im Dossier zur frühen Nutzenbewertung vorzulegen.³ Dem pU steht es darüber hinaus frei, auch andere Studien für den Nachweis des Zusatznutzens eines neuen Wirkstoffs in dem Dossier zur frühen Nutzenbewertung einzureichen. Dies spiegelt sich entsprechend in der Struktur eines Dossiers zur frühen Nutzenbewertung wider. Diese Struktur des vom pU einzureichenden Dossiers ist in der Verfahrensordnung des Gemeinsamen Bundesausschusses (G-BA) beschrieben.¹⁰ Modul 4 des Dossiers enthält dabei die vom pU aufbereiteten Daten zur Nutzenbewertung.

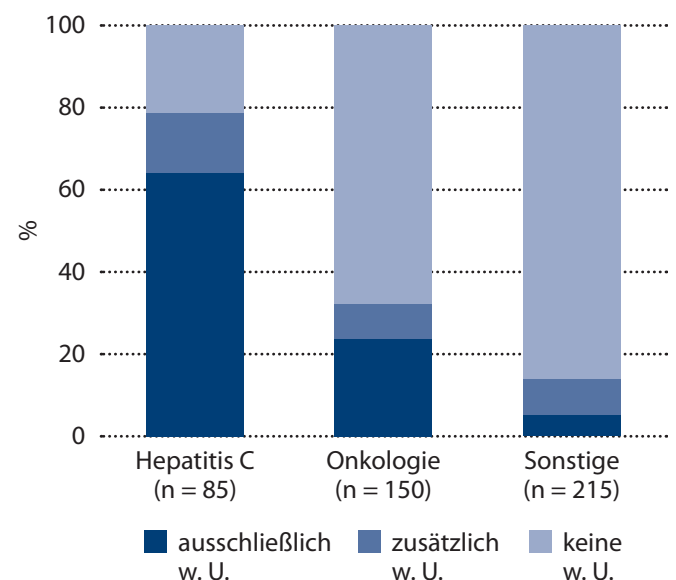
In Kasten A ist der Aufbau des Ergebniskapitels in Modul 4 des Dossiers gemäß den Vorgaben des G-BA dargestellt. Die im vorliegenden Artikel betrachteten einarmigen Studien sind dabei den „Weiteren Untersuchungen“ zuzuordnen, sie werden daher in der Regel von den pU in Kapitel 4.3.2.3 in Modul 4 des Dossiers dargestellt. Die nachfolgend beschriebenen Auswertungen beziehen sich daher auf die Vorlage von Nachweisen zum Zusatznutzen in diesem Kapitel des Dossiers. Zwar kann es zum einen vorkommen, dass solche Studien auch in anderen Kapiteln des Dossiers dargestellt werden. Zum anderen können im Kapitel „Weitere Untersuchungen“ vereinzelt auch lediglich Expertenmeinungen ohne zugrundeliegende Studien dar-

gestellt werden. Da es sich nach den bisherigen Erfahrungen jedoch jeweils um Einzelfälle handelt, wird diese Unschärfe bei den nachfolgenden Ausführungen akzeptiert, da davon auszugehen ist, dass die grundsätzliche Aussage hierdurch nicht relevant beeinflusst wird.

Vorlage „Weiterer Untersuchungen“ in bisherigen Dossiers

In Abbildung 1 ist dargestellt, zu wie vielen Nutzenbewertungen „Weitere Untersuchungen“ von den pU in den Jah-

Indikationsspezifische Häufigkeit „Weiterer Untersuchungen“ in Dossiers zur frühen Nutzenbewertung



Quelle: Eigene Analyse

Abbildung 2: Bei der deutlichen Mehrheit der Fragestellungen wurden „Weitere Untersuchungen“ im Anwendungsgebiet Hepatitis C als einzige Evidenz vorgelegt.

ren 2011 bis 2017 vorgelegt wurden. Insgesamt handelt es sich um 243 Bewertungen mit 579 Fragestellungen (eine Bewertung kann mehrere Fragestellungen umfassen, z.B. Erst- und Zweitlinientherapie). Zu 12 der 243 Bewertungen (5 Prozent) bzw. 129 der 579 Fragestellungen (22 Prozent) hat der jeweilige pU gar keine Studien zum Zusatznutzen im Dossier vorgelegt, weder direkt vergleichende RCT noch „Weitere Untersuchungen“.

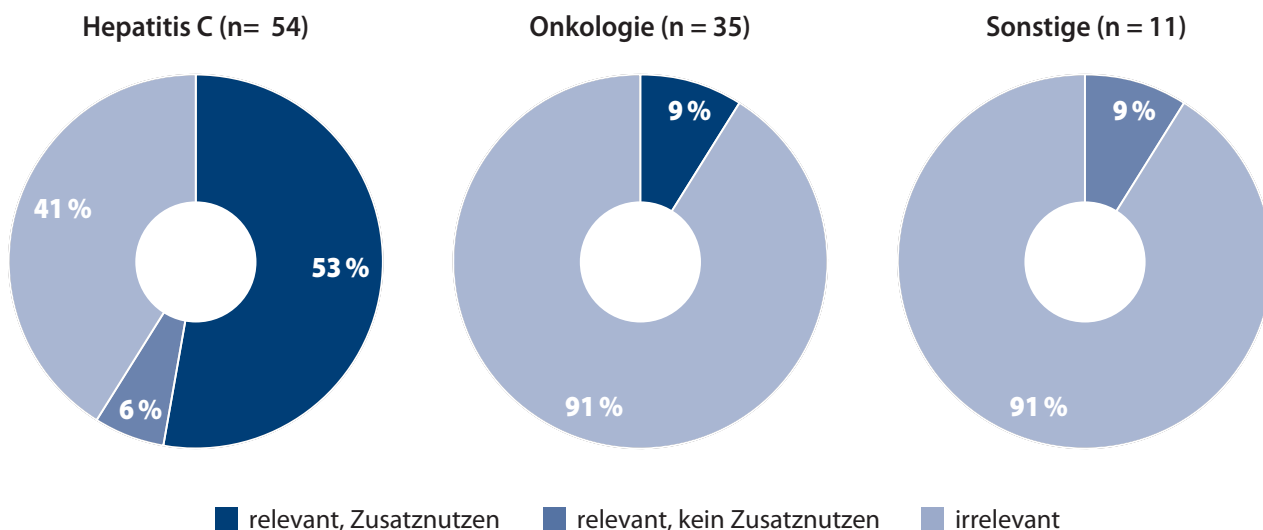
Zu 143 der 450 Fragestellungen mit Studien (32 Prozent) wurden „Weitere Untersuchungen“ vorgelegt, entweder als einzige oder als zusätzliche Evidenz. Dabei handelte es sich zumeist um Fragestellungen im Anwendungsgebiet Hepatitis C (66 Fragestellungen, 46 Prozent) oder im Bereich Onkologie (48 Fragestellungen, 34 Prozent). Auf alle übrigen

Indikationen entfielen nur 29 der Fragestellungen mit „Weiteren Untersuchungen“ (20 Prozent).

Die Präferenz der Vorlage „Weiterer Untersuchungen“ für das Anwendungsgebiet Hepatitis C wird noch deutlicher, wenn man die in den jeweiligen Indikationen insgesamt bewerteten Fragestellungen betrachtet. Dies ist in Abbildung 2 dargestellt.

Im betrachteten Zeitraum wurden im Anwendungsgebiet Hepatitis C insgesamt 85 Fragestellungen bewertet. Bei 66 und damit der deutlich überwiegenden Zahl dieser Fragestellungen (78 Prozent) wurden „Weitere Untersuchungen“ vorgelegt, zumeist als einzige Evidenz (54 Fragestellungen), teilweise auch als zusätzliche Evidenz (12 Fragestellungen). Dem gegenüber spielten „Weitere Untersu-

Indikationsspezifische Relevanz „Weiterer Untersuchungen“ in den Beschlüssen des G-BA



Quelle: Eigene Analyse

Abbildung 3: Im Anwendungsgebiet Hepatitis C sind die vorgelegten „Weiteren Untersuchungen“ aus Sicht des G-BA in rund der Hälfte der Fälle relevant gewesen. Bei anderen Indikationen ist diese Quote deutlich niedriger.

chungen“ in der Onkologie eine geringere Rolle, wobei immerhin für etwa ein Viertel der Fragestellungen „Weitere Untersuchungen“ die einzige vom pU vorgelegte Evidenz darstellten. Zu sonstigen Indikationen wurden weitere Untersuchungen nur im Ausnahmefall vorgelegt.

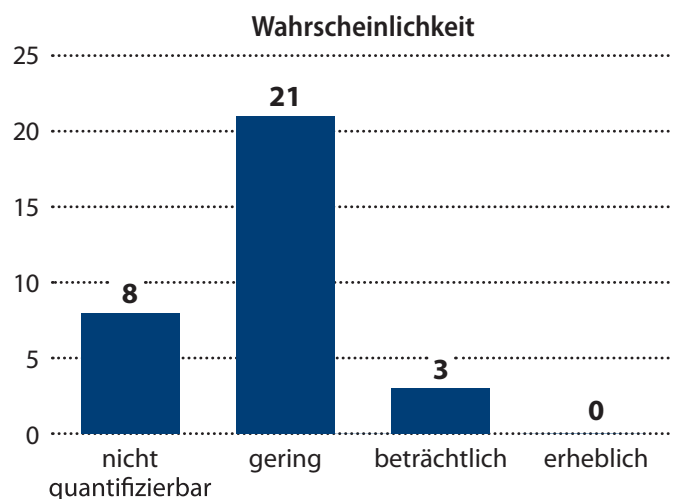
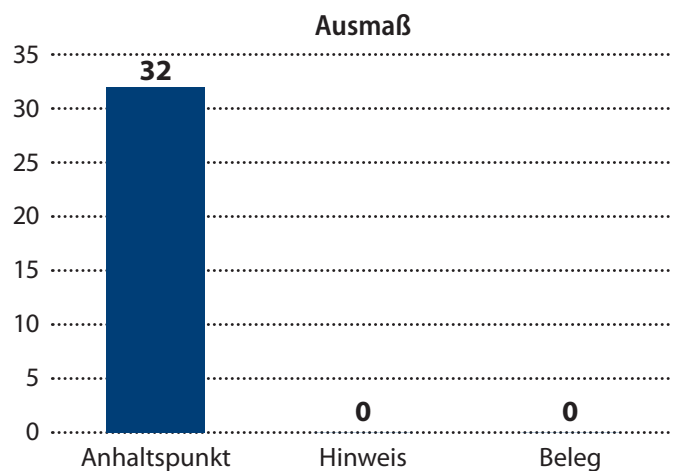
Ergebnisse aus „Weiteren Untersuchungen“

In Abbildung 3 ist dargestellt, ob der G-BA die vom pU vorgelegten „Weiteren Untersuchungen“ in seinen Beschlüssen als relevant angesehen hat, und wenn ja, ob daraus ein Zusatznutzen für den neuen Wirkstoff abgeleitet wurde. Relevant waren dabei diejenigen Studien, deren Ergebnisse der G-BA im Beschluss zur Nutzenbewertung dargestellt hat. Betrachtet wurden nur diejenigen Fragestellungen, bei denen „Weitere Untersuchungen“ die einzige vom pU vorgelegte Evidenz darstellten.

Die im Anwendungsgebiet Hepatitis C von den pU vorgelegten „Weiteren Untersuchungen“ waren in gut der Hälfte der Fälle auch aus Sicht des G-BA relevant, und aus den relevanten „Weiteren Untersuchungen“ wurde zumeist ein Zusatznutzen für den neuen Wirkstoff abgeleitet. Dem gegenüber waren sowohl die im Bereich Onkologie als auch die für sonstige Indikationen vorgelegten „Weiteren Untersuchungen“ zumeist irrelevant. Bei den drei Fragestellungen der Onkologie, bei denen nach Bewertung des G-BA die „Weiteren Untersuchungen“ relevant waren, wurde daraus jeweils ein Zusatznutzen für den neuen Wirkstoff abgeleitet.

Bei den sonstigen Indikationen führten „Weitere Untersuchungen“ in keinem Fall zu einem Zusatznutzen. In Abbildung 4 ist über alle Indikationen hinweg dargestellt, welches Ausmaß und welche Aussagewahrscheinlichkeit der G-BA in den Fällen, in denen aus „Weiteren Untersuchungen“ ein Zusatznutzen abgeleitet wurde, festgestellt hat (insgesamt 32 Fälle).

Ausmaß und Wahrscheinlichkeit des Zusatznutzens bei Bewertungen auf Basis „Weiterer Untersuchungen“



Quelle: Eigene Analyse

Abbildung 4: In 24 von 32 Fällen wurde der Zusatznutzen auf Basis „Weiterer Untersuchungen“ vom G-BA quantifiziert, fast immer war es ein „geringer“ Zusatznutzen.

In der Mehrheit der Fälle (24, 75 Prozent) wurde der Zusatznutzen auf Basis dieser Studien vom G-BA quantifiziert, zumeist lag dabei ein geringer Zusatznutzen vor. Ein nicht quantifizierbarer Zusatznutzen wurde bei 25 Prozent der Fälle festgestellt. In allen 32 Fällen war die Aussagewahrscheinlichkeit niedrig („Anhaltspunkt“).

Einordnung der Ergebnisse aus Beschlüssen zur Nutzenbewertung und Fazit

Die bisherigen Ergebnisse aus frühen Nutzenbewertungen zeigen, dass einarmige Studien (als „Weitere Untersuchungen“) je nach Indikation häufig oder nur im Ausnahmefall von den pU vorgelegt werden. Im Anwendungsgebiet Hepatitis C waren in den vergangenen Jahren einarmige Studien häufig die einzige vorgelegte Evidenz. Aufgrund der zuvor teilweise unzureichenden Therapieoptionen (z. B. bei der Behandlung von Kindern und Jugendlichen, Patienten mit dekompensierter Leberzirrhose oder mit HIV ko-infizierten Patienten, aber auch aufgrund der hohen Nebenwirkungsrate der vorhergehenden Standardtherapie) waren solche Studien in der Hälfte der Fälle auch ausreichend, um daraus einen Zusatznutzen abzuleiten, wenn auch durchweg mit geringer Aussagesicherheit.

Die im Bereich Onkologie vorgelegten einarmigen Studien waren hingegen nur selten geeignet, um daraus Aussagen zum Zusatznutzen abzuleiten. Dies stützt die Kritik an der zunehmenden Praxis beschleunigter Zulassungen ohne ausreichende Evidenzgenerierung.¹¹ Denn die hierfür durchgeführten Studien sind damit nachweislich häufig ungeeignet, die Therapieentscheidung zu informieren. Nur in Ausnahmefällen ist das Gegenteil der Fall, z. B. beim Wirkstoff Vismodegib zur Behandlung des Basalzellkarzinoms. Hier hat der G-BA aufgrund dessen, dass es für das Anwendungsgebiet von Vismodegib keine zufriedenstellenden Therapieoptionen gab und Spontanremissionen

nicht zu erwarten waren, einen Zusatznutzen auf Basis einer einarmigen Studie abgeleitet.¹²

Insgesamt lässt sich feststellen, dass einarmige Studien nicht per se für die Bewertung des Zusatznutzens ungeeignet sind. Ihre Eignung hängt jedoch stark vom Kontext ab, unter anderem davon, ob es ein ausreichend sicheres und publiziertes Wissen zur Vergleichstherapie gibt (z. B. ebenfalls aus einarmigen Studien oder zum natürlichen Verlauf der Erkrankung). Wenn aber keine sehr großen Effekte durch die neue Therapie zu erwarten sind, z. B. weil bereits mit der bisherigen Standardbehandlung gute Therapieergebnisse erzielt werden, dann wird sich aus einarmigen Studien auch dann kein Zusatznutzen ableiten lassen, wenn sie vom Grundsatz her für die Bewertung geeignet sind. Denn vergleichende Bewertungen auf Basis solcher Studien sind grundsätzlich potenziell so hoch verzerrt, dass sich kleine Therapieeffekte aus diesen Bewertungen nicht mit ausreichender Sicherheit ableiten lassen.

Zusammenfassend sollte die Durchführung einarmiger Studien aus methodischer und forschungsökonomischer Sicht, aber insbesondere auch mit Blick darauf, dass Wissensgenerierung primär die Patientenversorgung verbessern sollte, auf die wenigen Ausnahmefälle beschränkt werden, bei denen begründet (auch aufgrund ausreichend sicheren Wissens zur Vergleichstherapie) von sehr großen (dramatischen) Therapieeffekten des neuen Wirkstoffs ausgegangen werden kann.

Literatur:

¹ SGB V Handbuch (2017). Sozialgesetzbuch V Krankenversicherung. Altötting: KKF-Verlag; 2017.

² Sackett DL (1979). Bias in analytic research. *J Chronic Dis* 32(1-2): 51-63.

³ Bundesministerium für Gesundheit (2014). Verordnung über die Nutzenbewertung von Arzneimitteln nach § 35a Absatz 1 SGB V für Erstattungsverein-

barungen nach § 130b SGB V (Arzneimittel-Nutzenbewertungsverordnung – AM-NutzenV); Stand: 27.03.2014. URL: <http://www.gesetze-im-internet.de/bundesrecht/am-nutzenv/gesamt.pdf>.

⁴ Unkel S, Amiri M, Benda N, Beyersmann J, Knoerzer D, Kupas K, Langer F, Leverkus F, Loos A, Ose C, Proctor T, Schmoor C, Schwenke C, Skipka G, Unnebrink K, Voss F, Friede T (2019). On estimands and the analysis of adverse events in the presence of varying follow-up times within the benefit assessment of therapies. *Pharm Stat* 18(2): 166-183.

⁵ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2015). Ceritinib: Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung vom 29.09.2015; Auftrag A15-24 (IQWiG-Berichte; Band 329). URL: https://www.iqwig.de/download/A15-24_Ceritinib_Nutzenbewertung_35a_SGB-V.pdf.

⁶ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2017). Nivolumab Nutzenbewertung gemäß § 35a SGB V; Dossierbewertung vom 30.03.2017; Auftrag A16-76 (IQWiG-Berichte; Band 501). URL: https://www.iqwig.de/download/A16-76_Nivolumab_Nutzenbewertung-35a-SGB-V_V1-0.pdf.

⁷ Neugebauer R, Schmittziel JA, Zhu Z, Rassen JA, Seeger JD, Schneeweiss S (2015). High-dimensional propensity score algorithm in comparative effectiveness research with time-varying interventions. *Stat Med* 34(5): 753-781.

⁸ Glasziou PP, Chalmers I, Rawlins M, McCulloch P (2007). When are randomised trials unnecessary? Picking signal from noise. *BMJ* 334(7589): 349-351.

⁹ Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2017). Allgemeine Methoden Version 5.0. URL: https://www.iqwig.de/download/Allgemeine-Methoden_Version-5-0.pdf.

¹⁰ Gemeinsamer Bundesausschuss (2019). Anlage II zum 5. Kapitel der Verfahrensordnung des Gemeinsamen Bundesausschusses: Format und Gliederung des Dossiers, einzureichende Unterlagen, Vorgaben für technische Standards. URL: <https://www.g-ba.de/richtlinien/anlage/167/>

¹¹ Banzi R, Gerardi C, Bertele V, Garattini S (2015). Approvals of drugs with uncertain benefit–risk profiles in Europe. *Eur J Intern Med* 26(8): 572-584.

¹² Gemeinsamer Bundesausschuss (2016). Beschluss über eine Änderung der Arzneimittel-Richtlinie (AM-RL): Anlage XII - Beschlüsse über die Nutzenbewertung von Arzneimitteln mit neuen Wirkstoffen nach § 35a SGB V – Vismodegib. Beschluss vom 04.08.2016. URL: https://www.g-ba.de/downloads/39-261-2672/2016-08-04_AM-RL-XII_Vismodegib_D-213_BAnz.pdf.

Evidenzgewinnung und -beurteilung bei Volkskrankheiten am Beispiel Diabetologie

Prof. Dr. Dirk Müller-Wieland | Klinisches Studien-Zentrum, Medizinische Klinik I, Universitätsklinikum der RWTH Aachen

Das Arzneimittelmarktneuordnungsgesetz (AMNOG) von 2011 regelt durch die frühe Nutzenbewertung neuer Medikamente durch das Institut für Wirtschaftlichkeit und Qualität im Gesundheitswesen (IQWiG) und der abschließenden Entscheidung des Gemeinsamen Bundesausschuss (G-BA) den Marktzugang und die Finanzierung durch die Selbstverwaltung für unsere Solidargemeinschaft. Dies ist grundsätzlich begrüßenswert. Kriterien für die Nutzenbewertung sind laut SGB V Daten zur Mortalität, Morbidität, Symptomen und Krankheitslast im Vergleich zu einer vom G-BA festgelegten zweckmäßigen Vergleichstherapie (zVT). Die medizinischen Fachgesellschaften fordern unter anderem, prozedural bei der Festlegung der Fragestellung, der zVT und Operationalisierung von Endpunkten bzw. „Nutzen-Festlegung“ eingebunden zu werden. Problem bei häufigen und chronischen Krankheiten, wie zum Beispiel dem Typ-2-Diabetes, ist die optimale Studiengestaltung zur frühen und systematischen Erfassung der vom SGB V festgelegten Kriterien und ein direkter Vergleich mit der zVT.

A **MNOG und neue Diabetes-Medikamente**
Ziel der Bewertung eines Zusatznutzens im Rahmen des AMNOG durch das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) nach Beauftragung durch den Gemeinsamen Bundesausschuss (G-BA) betrifft nicht, um diesem Missverständnis vorzubeugen, eine Beurteilung von Effektivität und Sicherheit, sondern allein, ob eine neue Substanz im Rahmen der zugelassenen Anwendung, die in der Fachinformation niedergelegt ist, einen Zusatznutzen im Vergleich zu einer vom G-BA festgelegten Vergleichssubstanz hat.

Das AMNOG berücksichtigt nicht ausreichend und prozedural festgelegt den Sachverstand wissenschaftlicher Fachgesellschaften, zum Beispiel bei der Festlegung einer zweckmäßigen Vergleichstherapie, der Festlegung und operationalisierten Bewertung eines Zusatznutzens, der Einbindungen von externen Gutachtern und bei Abweichungen von Leitlinien.¹

In der Diabetologie sind zwischen 2011 und 2018 36 Verfahren mit 110 Subgruppen durchgeführt worden, bei den fast 90 Prozent eine Bewertung mit fehlendem Zusatznutzen erfolgte.² In aller Regel lag dies an der auf Grundlage des SGB V (Kriterien siehe oben) berechtigten Einschätzung, dass eine alleinige Absenkung des Surrogatparameters nicht ausreicht und Daten im Vergleich zur zVT fehlten. Weitere Diskussionspunkte sind unterschiedliche Beurteilungen und Bewertungen von Hypoglykämien sowie meist fehlenden Daten zur Patientenperspektive, mikrovaskulären Endpunkten und kardiovaskulärer Überlegenheit. In letzter Zeit hat sich letztere Situation etwas geändert, da seit 2008 die Food and Drug Administration (FDA) der USA sogenannte kardiovaskuläre Sicherheitsstudien zulasungsnah fordert und diese Entscheidung nach zehn Jahren auch nicht revidiert hat.^{3,4}

Kardiovaskuläre Sicherheitsstudien

Bei kardiovaskulären Endpunktstudien werden drei verschiedene Designs bzw. Arten von Studien unterschieden, d.h. Vergleich von Therapiestrategien, Überlegenheits- und Sicherheitsstudien (Übersicht:^{5,6}). Bei einem Vergleich von Therapiestrategien wird bei vergleichbarer Absenkung des Risikofaktors gefragt, ob zwischen zwei Substanzen ein Unterschied in Bezug auf kardiovaskuläre Ereignisse besteht. Bei der oralen Diabetestherapie läuft zurzeit eine Studie, die den Sulfonylharnstoff Glimperid mit dem DPP-4-Hemmer Linagliptin vergleicht. Bei Überlegenheitsstudien wird der Risikofaktor im Vergleich zur Placebogruppe gesenkt; bestes Beispiele sind Statin-Studien.

Sicherheitsstudien sind so angelegt, dass die zu untersuchende Substanz zwar mit Placebo verglichen wird, dass

aber auch im Placebo-Arm der Studie der zu beeinflussende Risikoparameter, in diesem Falle HbA_{1c} bzw. Blutzucker, Studienprotokoll-gemäß vergleichbar gesenkt wird. Daher testet ein solches Studiendesign allein die Frage, ob ein therapeutisches Molekül einen eigenen bzw. von der Blutzuckersenkung unabhängigen Effekt auf das kardiovaskuläre Risiko hat. Kardiovaskuläre Sicherheitsstudien zeigten Sicherheit für die DPP-4-Hemmer, Überlegenheit für einzelne GLP-1-Rezeptoragonisten, wie Liraglutid, Semaglutid, Albiglutid und laut Pressemitteilung (Veröffentlichung der Ergebnisse stehen noch aus) für Dulaglutid.

Die SGLT-2-Hemmer Dapagliflozin, Canagliflozin und Empagliflozin haben in den Sicherheitsstudien bereits einen konsistenten Hinweis für Renoprotektion gezeigt, die Hospitalisierung für Herzinsuffizienz gesenkt und bei Hochrisiko-Patienten eine Reduzierung von MACE und der kardiovaskulären Sterblichkeit. Letzterer Parameter war am meisten in der EMPA-REG-OUTCOME-Studie belegt, weshalb Empagliflozin in dieser Patientengruppe auch einen beträchtlichen Zusatznutzen vom G-BA beschieden wurde. Zudem wurden aufgrund dieser Studienlage Liraglutid und Empagliflozin als zVT bei Patienten mit Diabetes und vorbestehender kardiovaskulärer Erkrankung vom G-BA aufgenommen.

Translation dieser Studien in den „Medizinischer Standard“ bei der Behandlung des Typ-2-Diabetes aus nationaler und internationaler Sicht

Die medikamentöse Therapie des Typ-2-Diabetes ist eine Stufentherapie, die gemäß Leitlinien Patienten-orientiert und Evidenz-basiert umgesetzt werden soll. Die diesbezüglich jährlich aktualisierten Praxisempfehlungen der DDG sowie nun auch die gemeinsamen Empfehlungen der europäischen (EASD) und amerikanischen (ADA) Diabetes-Gesellschaften, die am 5. Oktober 2018 konsentiert und



Professor Dr. Dirk Müller-Wieland ist Präsident (2017-2019) der Deutschen Diabetes Gesellschaft (DDG) und seit 2011 Sprecher der DDG-Kommission Gesundheitspolitik. Nach dem Studium der Humanmedizin in Hamburg folgte ein zweijähriger Forschungsaufenthalt am Joslin Diabetes Center der Harvard Medical School. Seine Facharztausbildung zum Internisten mit den Schwerpunkten Endokrinologie, Diabetologie und Stoffwechsel erhielt er am UKE in Hamburg und in Köln. Seit 2016 ist er tätig in der I. Medizinischen Klinik des Universitätsklinikums der RWTH Aachen.

zeitgleich publiziert wurden, sind entsprechend in die Empfehlungen des „medizinischen Standards“ der ADA 2019 aufgenommen worden und werden im Folgenden kurz zusammengefasst.⁷⁻⁹ Metformin ist weiterhin Therapie der ersten Wahl. Die bevorzugte Empfehlung für eine zusätzliche Medikation über Metformin hinaus richtet sich nach patientenindividuellen Kriterien und der vorliegenden Evidenz für die jeweiligen Präparate hinsichtlich antihyperglykämischer Wirksamkeit, Hypoglykämierisiko, Beeinflussung des Körpergewichts, individuellem Nebenwirkungsprofil, und vor allem Beeinflussung kardiovaskulärer und renaler Endpunkte. Konkret heißt dies, dass auf diese Weise zeitnah die Ergebnisse neuer kardiovaskulärer Endpunktstudien (CVOT/Cardiovascular Outcome Trial) in die klinische Praxis translatiert und implementiert werden.

Die Patientenpopulationen, die in den CVOTs eingeschlossen wurden, werden bei der Empfehlung zum bevorzugten Einsatz einer spezifischen Substanz reflektiert. Dementsprechend werden bevorzugt die Natrium-Glukose-Transporter (SGLT) 2-Hemmer oder Glukagon-Like Peptide (GLP)-1 Rezeptoragonisten bei Patienten mit vorbestehender kardiovaskulärer Erkrankung (ASCVD/Atherosclerotic Cardiovascular Disease) empfohlen.

Es wird dann der spezifische Wirkstoff der jeweiligen Substanzklasse gewählt, der in einer CVOT positive bzw. kardioprotektive Ergebnisse aufzeigte und in dem jeweiligen Gesundheitssystem auch zugelassen ist; im Hinblick auf die GLP-1 Analoga gilt dies in Deutschland zur Zeit sowohl für Liraglutid aber auch Semaglutid und bei den SGLT-2-Hemmern für Empagliflozin und Dapagliflozin.

Stärken und Schwächen kardiovaskulärer Sicherheitsstudien

Stärken kardiovaskulärer Sicherheitsstudien (Cardiovascular Outcome Trials/ CVOTs) sind in Abbildung 1 zusammen-

gefasst (siehe auch^{5,10}). Das Design testet nicht, ob eine Blutzuckersenkung durch ein neues Antidiabetikum einen Effekt auf kardiovaskuläre Endpunkte hat, sondern ob das neue Medikament unabhängig von der Glukosesenkung sicher ist. Daher soll per Protokoll in der Placebogruppe die Diabetestherapie entsprechend der lokalen, meist leitlinienbasierten Gegebenheiten eskaliert werden. Deshalb geben diese Studien meist auch einen Bereich der Nüchtern-Glukose als Zielwert an, anhand dessen die Therapie modifiziert werden soll.

Da dies meist ein „Add-on“-Konzept ist, d.h. Hinzugabe der neuen Medikation zu der bereits bestehenden, erhält

Kardiovaskuläre Sicherheitsstudien

Ziel und Stärken

- Testen NICHT einen Effekt der BZ-Senkung
- zulassungsnaher Erhöhung der Zahl untersuchter Patienten
- Sicherheit, Verträglichkeit, Medikamente (zusätzlich)

Klinische Implikationen

- Placebo widerspiegelt Versorgungsrealität
- Patienten-„Gruppen“ orientiert – Implikationen für „LL“
- Morbidität der Patientenpopulationen und Effekte?

Schwächen auch für frühe Nutzenbewertung

- „Add-on“ – Konzept
- Kein Vergleich von zVT oder Therapie-Strategien
- Kurze Zeit der Studien (schnell Ereignis-Zahl und „Non-Inferior“)

Quelle: Prof. Müller-Wieland

Abbildung 1: Stärken und Schwächen kardiovaskulärer Sicherheitsstudien für die frühe Nutzenbewertung im Rahmen des AMNOG.

man durch diese großen Studien an meist mehreren Tausend Patienten einen großen Datenpool zur Sicherheit und Verträglichkeit der Medikation bei Patienten, die meist aufgrund ihrer Ko-Morbiditäten bereits mehrere Medikamente einnehmen. Für die frühe Nutzenbewertung und einer zweckmäßigen Vergleichstherapie ist relevant, dass der Placebo-Arm in aller Regel die Versorgungsrealität widerspiegelt. Da bei positiven Effekten die zugrundeliegenden glukose-unabhängigen Mechanismen nicht bekannt sind, werden diese Studien in „bevorzugte Indikationen“ bei Praxisempfehlungen translatiert, die sich dann aber auf die in der jeweiligen Studie untersuchte Patienten-Population bezieht.

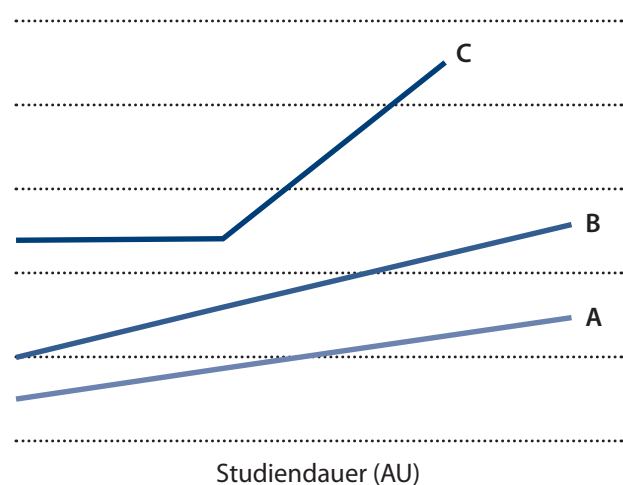
Daher bleibt offen, ob manch ein klinischer Effekt, der in der meist relativ kurzen Beobachtungszeit observiert wird, auch in einem direkten Zusammenhang mit der Grundmorbidität bzw. dem Krankheitsverlauf des Patienten besteht, welches die Generalisierbarkeit der Studien potenziell einschränkt. Dies gilt insbesondere für Endpunkte, bei denen Krankheitsstadien-bedingt kein linearer Zusammenhang zwischen Drug-target-Modulation und klinischem Endpunkt besteht (siehe als Beispiel Abbildung 2).

Schwächen dieses Studiendesigns ist zudem, dass die pharmazeutischen Unternehmen diese Sicherheitsstudien schnell durchführen wollen, d.h. man nimmt eine Patienten-Population mit einer hohen klinischen Ereignisrate und macht ein Ereignis-getriebenes Design. Somit wird die Studie bei einer statistisch relevanten Zahl von Ereignissen beendet. Daher sind die eingeschlossen Patienten in aller Regel bereits kardiovaskuläre Hochrisikopatienten, die Studiendauer ist kurz (zwei bis drei Jahre) und es wird als primärer Endpunkt meist die Zeit bis zum Auftreten des ersten Ereignisses gewählt.

Hierfür wird in aller Regel ein zusammengesetzter Endpunkt gewählt, d.h. MACE (Major Adverse Cardiovascular

Beziehungen zwischen Ereignisraten und Krankheitsdauer

Inzidenz des klinischen Endpunktes in der untersuchten Population (AU)



Quelle: Prof. Müller-Wieland

Abbildung 2: Bei A und B besteht ein linearer Zusammenhang, d.h. eine relative Risikoreduktion wird in den beiden unterschiedlichen Populationen zu sehen sein, allerdings mit einer unterschiedlichen absoluten Risikoreduktion. Im Fall C ist die Entwicklung von Ereignisraten abhängig sowohl von der Grundmorbidität bzw. -Risikos (senkrechte Achse) sowie dem Verlauf der Krankheit, die sich in diesem Fall nicht linear verhält. Damit könnte ein entsprechender Effekt – selbst auf die relative Risikoreduktion – in der Population A und B nicht beobachtet werden.

Event), der in aller Regel aus kardiovaskulärem Tod, nicht-tödlichem Myokardinfarkt und Schlaganfall besteht. Aufgrund der kurzen Studiendauer und der unterschiedlichen Wahrscheinlichkeit bzw. Ereignisraten der Einzelkompo-

nenen dieses zusammengesetzten Endpunktes, kommt es häufig zu Diskrepanzen, d.h. zum Beispiel einen Effekt auf den kardiovaskulären Tod, aber nicht auf Myokardinfarkt, oder einen Effekt auf Schlaganfall und nicht die anderen Komponenten. So kommt es auch zu Effekten auf Myokardinfarkt „ohne Effekt“ auf kardiovaskulären Tod; es wird eben häufig bei der Interpretation vergessen, dass diese Studien zeitlich limitiert sind. Deswegen gibt es berechnete Forderungen, dass diese Studien unabhängig von der Ereignisrate eine Mindestlaufzeit haben sollten und zudem auch kumulative Ereignisse in ein und demselben Individuum zusätzlich präspezifiziert ausgewertet werden sollten.^{10,11}

Kardiovaskuläre Endpunktstudien und Subgruppenanalysen zu einer zVT

Sicherheitsstudien, die global bzw. internationalisiert organisiert werden, machen eine dezidierte Subgruppen-Analyse zu einer für die frühe Nutzenbewertung meist nach Beginn der Studie durch den G-BA festgelegten zweckmäßigen Vergleichstherapie nicht möglich. Daher muss bei der Beurteilung dieser Studien im Rahmen der frühen Nutzenbewertung der Zeitpunkt des Entstehens des Studiendesigns und des zu diesem Zeitpunkt bestehenden „medizinischen Standards“ berücksichtigt werden.

Wünschenswert wäre, wenn auch die Zulassungsbehörden weitere per Protokoll festgelegte sekundäre Endpunkte dieser Studien für eine eventuelle Zulassungserweiterung berücksichtigen würden. Dies wäre auch aus Patientenperspektive effektiver und man müsste nicht auf Studien mit entsprechendem primären Endpunkt lange warten. Voraussetzung ist natürlich, dass die Ergebnisse eindeutig sind und ggf. so operationalisiert werden, als wenn sie bei der Erfassung einem primären Endpunkt entsprechen.

Operationalisierung von klinischen Ereignissen

Daher wäre es auch sinnvoll, dass nicht nur die zweckmäßige Vergleichstherapie frühzeitig und transparent festgelegt wird, sondern auch klinische Endpunkte für mikrovasculäre Komplikationen beim Diabetes, wie z.B. Nephropathie, Retinopathie und ggf. auch Neuropathie, operationalisiert und transparent seitens des IQWiG (ggf. in Zusammenarbeit mit der medizinischen Fachgesellschaft) festgelegt werden würde. Dies wäre gegenüber den in die Studien investierenden pharmazeutischen Unternehmern fair, sie hätten eine belastbare Grundlage nationaler Interessen beim Design globaler Studien durchzusetzen und würden dem methodischen Verfahren der frühen Nutzenbewertung eine höhere Legitimität und Plausibilität geben.

Dies gilt beim Diabetes auch für die Erfassung von Symptomen oder Nebenwirkungen; das häufigste Beispiel der Diskussion bei der frühen Nutzenbewertung ist die Operationalisierung hypoglykämischer Ereignisse. Das Problem bei diesem Symptom ist, dass die klinische Bedeutung nicht allein vom absolut erreichten niedrigen Wert abhängt, sondern auch von der Dauer, der Ausgangslage, der Geschwindigkeit einer Blutzucker-Senkung sowie der Grundmorbidität (z.B. mit oder ohne autonome Neuropathie, Alter etc.).

Zudem kann eine Hypoglykämie ein Indikator oder Mediator für Komplikationen sein und bei der Erfassung kann es im Rahmen von Studien zu Protokoll-bedingten Verzerrungen zwischen den Behandlungsrmen kommen. Ein neues Thema hierbei ist die durch eine kontinuierlich gemessene Glukosekonzentration in Geweben erfasste sogenannte „Time-in-Range“, bei der der klinische Stellenwert in Bezug auf Krankheitsverlauf sowie den Kriterien nach SGB V, Mortalität, Morbidität, Symptome und gesundheitsbezogene Krankheitslast, noch völlig unbekannt sind.

Patientenperspektive mit guter Qualität erfassen

Bei diesem Themenkomplex und bei Diabetes als Beispiel für eine chronische Krankheit ist die Patienten-Perspektive bzw. Patient-Reported-Outcome (PRO) wichtig. Hier zeigt sich, dass das in aller Regel verwendete Instrumentarium unzureichend standardisiert ist, um einen Vergleich zwischen Studien zu ermöglichen und die Erfassungen der verschiedenen Domänen (z.B. körperliche Funktionsfähigkeit, Angst, Müdigkeit, Schlafstörung, Einschränkung durch Schmerzen, Teilhabe an sozialen Rollen und Aktivitäten etc.) nicht auf die jeweilige spezifische Grunderkrankung ausgerichtet sind.

Hier ist dringender methodischer und klinischer Entwicklungsbedarf, da ansonsten die Sensitivität und Spezifi-

tät dieser Instrumente zu niedrig ist, das heißt, die Rate an falsch negativen Evaluationen ist unakzeptabel hoch und schadet ggf. der adäquaten Berücksichtigung patientenbezogener versorgungsrelevanter Endpunkte.

Perspektiven

Aus Patientenperspektive könnte nach Ansicht des Autors bei versorgungsbezogenen Endpunkte auch das vom Patienten definierte Ziel einer Therapie evaluiert werden. Dies würde bei chronischen Krankheiten die Patienten-Perspektive der Versorgung individueller berücksichtigen.

Zusätzlich sollten bei spezifischen Fragestellungen in der frühen Nutzenbewertung soweit möglich sogenannte „Real-World-Data“ und Registerdaten mit berücksichtigt werden. Die Priorisierung der Studienqualitäten für das Verfahren muss aber transparent festgelegt werden. In Zukunft werden ergänzende Daten aus sogenannten Cluster-Analysen oder mit neuen Methoden des „Deep Learnings“ Responder-Profilen oder Sub-Gruppen bei der großen Zahl von Patienten mit Volkskrankheiten neue Perspektiven bilden und eine Evaluierung eines Zusatznutzens patientenzentrierter gestalten.

Nach Ansicht des Autors sollten in Leitlinien auch „offene Fragen“ konsentiert werden, die damit einen Handlungs- bzw. Forschungsrahmen für die Solidargemeinschaft, den pharmazeutischen Unternehmer und die Forschungspolitik aufzeigen. Grundsatzfragen mit gesamtgesellschaftlicher Relevanz, die aufgrund von Interessenkonflikten nicht durch pharmazeutische Unternehmen angegangen werden können, sollten durch ein unabhängiges Gremium festgelegt und durch eine spezielle Finanzierung initiiert werden. Hierzu müssen kurzfristige und langfristige Ziele einer Diabetestherapie diskutiert, festgelegt und operationalisiert werden (siehe Abbildung 3).

Ziele einer Therapie des Diabetes

kurzfristig

- HbA_{1c} kontrollieren ohne Hypos und Gewichtszunahme
- zulassungsnahe Erhöhung der Zahl untersuchter Patienten
- Sicherheit, Verträglichkeit, Medikamente (zusätzlich)

mittelfristig

- Vermeidung mikro- und makrovaskulärer Komplikationen
- Reduktion der Morbidität und gute Lebensqualität

künftig

- Remission – Heilung – Prävention
- Verlauf (Progression) aufhalten können
- neurodegeneratives und malignes Risiko reduzieren

Quelle: modifiziert nach¹²

Abbildung 3: Ziele einer Diabetes-Therapie und Kriterien für die Entwicklung neuer therapeutischer Interventionen.

Schlussfolgerung

Ein Zusatznutzen gemäß AMNOG ist ein patientenrelevanter therapeutischer Effekt hinsichtlich des Gesundheitszustandes, Krankheitsdauer, Überleben, Nebenwirkungen und Lebensqualität, der qualitativ oder quantitativ höher ist im Vergleich zu einer zVT. Alle Beteiligten müssen die AMNOG-Kriterien für einen Zusatznutzen akzeptieren. Die wissenschaftlichen Fachgesellschaften sollten prozedural festgelegt in den Prozess der frühen Nutzenbewertung eingebunden werden.

Die Erfassung und Beurteilung der entsprechenden klinischen Endpunkte sollten seitens des IQWiG bzw. G-BA ggf. in Zusammenarbeit mit den jeweiligen wissenschaftlichen Fachgesellschaften operationalisiert werden, damit frühzeitig die notwendige Transparenz für eine auch später im Verfahren belastbare Studienplanung möglich ist. Krankheitsspezifisch erfasste gesundheitsbezogene Lebensqualität ist ein wichtiges Kriterium für die Patientenperspektive, ist verfahrensrelevant bei einer Nutzenbewertung und bisher sowohl methodisch und klinisch unzureichend erarbeitet.

Literatur:

¹ Gallwitz B, Kellerer M, Müller-Wieland D. Diabetes mellitus in Deutschland – politische Handlungsfelder. In: Deutscher Gesundheitsbericht Diabetes 2019. Hrsg. J. Kröger und D. Müller-Wieland. Verlag Kirchheim + Co GmbH, Mainz; S 276-282.

² Sackett DL (1979). Bias in analytic research. *J Chronic Dis* 32(1-2): 51-63.
Wörmann B und Vertreter verschiedener Fachgesellschaften für die Arbeitsgemeinschaft wissenschaftlicher medizinischer Fachgesellschaften (AWMF), Ad-Hoc-Kommission Nutzenbewertung und Deutsche Gesellschaft für Hämatologie und Medizinische Onkologie (DGHO). Frühe Nutzenbewertung neuer Arzneimittel in Deutschland 2011 – 2018. Februar 2019.

³ US Department of Health and Human Services Food and drug Administration. Guidance for Industry. Diabetes mellitus – Evaluating Cardiovascular Risk in New Antidiabetic Therapies to treat Type 2 Diabetes. 2008. www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm071627.pdf

⁴ Phend C. FDA advisors: tweak, don't dump CV outcomes trials. 2018. www.medpagetoday.com/cardiology/prevention/75938.

⁵ Cefau WT, Kaul S, Gerstein HC et al. Cardiovascular outcome trials in type 2 dia-

betes: Where do we go from here? Reflections from a Diabetes Care editors' expert forum. *Diabetes care* 2018; 41: 14-31

⁶ Home P. Cardiovascular outcome trials of glucose-lowering medications: an update. *Diabetologia* 2019; 62: 357-69

⁷ Davies MJ, D'Álessio DA, Fradkin J, Kernan WN, Mathieu C, Mingrone G, Rossing P, Tsapa A, Wexler DJ, Buse JB. Management of hyperglycemia in type 2 diabetes, 2018. A consensus report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes EASD. *Diabetologia* 2018; 61: 2461-2498 und *Diabetes Care* 2018; 41: 2669-2701.

⁸ American Diabetes Association. Standards in Medical Care in Diabetes – 2019. Pharmacological Approaches to glycaemic treatment. *Diabetes Care* 2019; 42 (Suppl 1): S90-S102.

⁹ Landgraf R, Kellerer M, Aberle J et al. Therapie des Typ-2-Diabetes. *Diabetologie und Stoffwechsel* 2018; 13 (Supplement): S83-290

¹⁰ McGuire DK, Marx N, Johansen OE, Inzucchi SE, Rosenstock J, George JT. FDA guidance on antihyperglycemic therapies for type 2 diabetes: One decade later. *Diabetes Obes Metab* 2019; 21: 1073-1078.

¹¹ Claggett B, Pocock S, Wie LJ et al. Comparison of time-to-first event and recurrent-event methods in randomized clinical trials. *Circulation* 2018; 138: 570-577.

¹² Müller-Wieland D. Orale Antidiabetika ermöglichen Individualisierung der Behandlung. *MMW Fortschr Med* 2015; 157 (Suppl1): 69-74

EUnetHTAs Perspektive im Hinblick auf Evidenzanforderungen

Dr. Giovanni Tafuri, Wissenschaftlicher Direktor | EUnetHTA

Das Europäische Netzwerk für Health Technology Assessment (engl. European Network of Health Technology Assessment, EUnetHTA) möchte dazu beitragen, dass ein nachhaltiges Modell für die wissenschaftliche und technische Zusammenarbeit im Bereich HTA auf EU-Ebene geschaffen wird. Seine wesentlichen Aktivitäten umfassen hierbei die Erstellung von gemeinsamen Bewertungsberichten sowie Verfahren im Zusammenhang mit dem frühen Dialog mit Herstellern; beides erfolgt in enger Zusammenarbeit mit der EMA. Obwohl die Angleichung von HTA- und regulatorischen Evidenzanforderungen sowie die Festlegung gemeinsamer Methoden im Bereich HTA und gemeinsame Bewertungen der EU-Mitgliedstaaten eine große Rolle spielen, um den Zugang für Patienten zu neuen Behandlungsoptionen zu verbessern, bleibt die größte Hürde in Sachen Zugang zu neuen Arzneimitteln der Preis.

Das EUnetHTA besteht aus 83 nationalen bzw. regionalen Institutionen, die sich mit HTA beschäftigen. Mit Hilfe dieses Projektes, das nun in die dritte Runde geht (Gemeinsame Aktion 3, 2016-2020), soll in enger Zusammenarbeit mit den beteiligten Akteuren und der Europäischen Kommission ein nachhaltiges Modell für die wissenschaftliche und technische Zusammenarbeit im Bereich HTA auf EU-Ebene etabliert werden. Einerseits werden im Rahmen einer freiwilligen Zusammenarbeit gemeinsame Themen im Bereich HTA erarbeitet, es soll aber auch eine stärkere Berücksichtigung und Implementierung dieser gemeinsamen Arbeit auf nationaler, regionaler und lokaler Ebene erzielt werden.

Eine der Hauptaktivitäten des EUnetHTA ist die Erstellung von gemeinsamen HTA-Berichten, die Bewertungen der relativen Wirksamkeit (engl. Relative Effectiveness Assessments, REAs) von Arzneimitteln oder Medizinprodukten. Bei Arzneimitteln werden diese Berichte üblicherweise von einer nationalen HTA-Institution mit Unterstützung einer anderen HTA-Institution aus einem anderen Mitgliedsstaat erstellt, die als Co-Autor fungiert. Bei der Bewertung fungieren eine Reihe von HTA-Institutionen als Rezensenten des Berichts.

So sind aus der aktuellen Gemeinsamen Aktion für Arzneimittel bis jetzt vier Berichte hervorgegangen, fünf Beurteilungen laufen derzeit und weitere befinden sich in der Vorbereitungsphase (<https://www.eunetha.eu/rapid-reas/>). Um diese Berichte leichter auf nationaler Ebene verwenden zu können, werden sie zeitnah nach dem Europäischen Öffentlichen Beurteilungsbericht (engl. European Public Assessment Report, EPAR) veröffentlicht, der kurz nach der Marktzulassung durch die Europäische Kommission erscheint. Gemeinsame Merkmale von REAs sind die Definition einer sogenannten PICO (engl. Population, Inter-

vention, Comparators and Outcomes) als Ausgangspunkt für die Beurteilung der Technologie, die Verwendung von GRADE (Grading of Recommendations Assessment, Development and Evaluation), welche das Risiko einer Bias-Beurteilung misst durch das Qualitäts-Rating-Tool der Cochrane Collaboration, das Vorhandensein von systematischen Übersichtsberichten zur Erfassung der gesamten verfügbaren Evidenz, die indirekten Vergleiche und Netzwerk-Metaanalysen sowie die Berücksichtigung von Rückmeldungen von Patienten und Experten.

Bemerkenswerterweise wurden in einem aktuellen Dokument von EUnetHTA-Partnern gemeinsame Methoden zum Umgang mit Patienten und zur Berücksichtigung ihrer



Dr. Giovanni Tafuri ist Wissenschaftlicher Direktor des Europäischen Netzwerks für Health Technology Assessment (EUnetHTA) beim EUnetHTA-Sekretariat in den Niederlanden. Vor seiner Berufung arbeitete er bei der HTA und in der Zulassungsabteilung der italienischen Arzneimittelagentur (AIFA) (2006 bis 2018) sowie als entsandter nationaler Experte für wissenschaftliche Beratung bei der EMA (2015 bis 2017). Er ist Apotheker, hat einen Magister der Naturwissenschaften in „International Health Care Management, Economics and Policy“, einen Doktorgrad in „Pharmaceutical Policy and Regulation“ und war als wissenschaftlicher Mitarbeiter bei verschiedenen nationalen und internationalen Institutionen tätig.

Rückmeldungen bei der Erstellung von REAs festgestellt.¹ Ein Vergleich zwischen den veröffentlichten REAs und den entsprechenden Europäischen Öffentlichen Beurteilungsberichten (engl. European Public Assessment Reports, EPARs), die von der EMA veröffentlicht wurden, zeigt gemeinsame Unsicherheiten bei HTA-Institutionen und Verordnungsgebern bei Arzneimitteln zum Zeitpunkt der behördlichen Zulassung; es zeigt jedoch auch, dass man diesen Unsicherheiten unterschiedlich begegnet.

Die unterschiedlichen Aufgabenbereiche der EMA- und HTA-Institutionen spiegeln sich auch in der Evidenz wider, die für die Beurteilung verwendet wird. In der Tat müssen HTA-Institutionen zusätzliche Studien berücksichtigen, um indirekte Vergleiche zu ziehen, und Netzwerk-Metaanalysen durchführen, um die relative Wirksamkeit im Vergleich zu therapeutischen Alternativen zu beurteilen, die nicht in den Zulassungsstudien verwendet wurden (Abbildung 1).

Als weitere zentrale EUnetHTA-Aktivität wurden Verfahren zum frühen Dialog mit Herstellern während der klinischen Entwicklung in Zusammenarbeit mit der EMA entwickelt. Ziel ist es, dass die Hersteller spezielle HTA- und regulatorische Anforderungen in ihren Entwicklungsplan integrieren und so die Anforderungen hinsichtlich der Evidenz von Behörden und HTA-Institutionen erfüllen.

In einer retrospektiven Analyse basierend auf verschiedenen Verfahren im Zusammenhang mit dem frühen Dialog zwischen 2010 und 2015 fand man heraus, dass es Gemeinsamkeiten hinsichtlich der Evidenzanforderungen zwischen Verordnungsgebern und HTA-Institutionen gibt.² Obgleich es weniger Gemeinsamkeiten hinsichtlich der Empfehlung von Komparatoren gab, stellten die Untersucher insgesamt ein hohes Maß an Übereinstimmungen zwischen EMA- und HTA-Institutionen fest.

In einer anderen Analyse desselben Autorentams wurden die tatsächlichen Auswirkungen von parallelen wis-

senschaftlichen Beratungen zu klinischen Entwicklungen untersucht, indem die Berücksichtigung von Empfehlungen von regulatorischen und Gesundheitstechnologiebewertungen beurteilt wurden.³ Eine der wichtigsten Erkenntnisse war, dass Hersteller ihre Entwicklungsprogramme eher auf der Grundlage von regulatorischen und HTA-Beratungen über die Auswahl von primären Endpunkten und Komparatoren ändern. Die Analyse zeigte jedoch auch, dass die Auswahl des Studienkomparators eine echte Herausforderung darstellt, bei der die Hersteller tendenzi-

ell eher regulatorischen Beratungen folgen.

In der wissenschaftlichen Fachwelt wurde das Problem der unterschiedlichen Evidenzanforderungen zwischen Verordnungsgebern und HTA-Institutionen breit diskutiert. Ein einziges Genehmigungsverfahren in Europa mit einer einzigen Gesetzgebung und klar definierten Bewertungskriterien, wird in der Tat mit dem Umfeld von 28 unterschiedlichen Mitgliedsstaaten verglichen, die alle unterschiedlichen Gesetzgebungen, HTA-Methoden, Kriterien und Erstattungssystemen unterliegen.

Eichler und Kollegen stellten Überlegungen an, wie man diese beiden ganz unterschiedlichen Systeme anpassen kann und veröffentlichten 2010 eine Arbeit, in der sie die aktuellen und mögliche zukünftige Paradigmen für die Interaktion zwischen Verordnungsgebern und Experten bei der Bewertung von Gesundheitstechnologien aufzeigten.⁴ Seither haben viele Autoren die Unterschiede und Konsequenzen dieser Ungleichheiten zwischen den beiden Interessengruppen untersucht.

Bei diesen Analysen wurden unterschiedliche Zielsetzungen verfolgt. So zeigte eine systematische Beurteilung von Onkologika-Zulassungen durch die EMA von 2009 bis 2013 beispielsweise, dass die meisten Medikamente ohne Evidenz für einen Nutzen hinsichtlich Überleben oder Lebensqualität zugelassen wurden; diese Punkte sind für HTA und Entscheidungsträger im Bereich Preisgestaltung und Erstattung entscheidend.⁵

Grundsätzlich stellen Unsicherheiten hinsichtlich Nutzen & Risiko von Produkten zum Zeitpunkt der Zulassung eine große Herausforderung für die Definition des zusätzlichen Nutzens neuer Technologien durch HTA-Institutionen sowie deren Gesamtwert dar, was für Entscheidungsträger in der Preisgestaltung und Erstattung unerlässlich ist.

Obwohl Zulassungsbehörden den Herstellern häufig die Verpflichtung auferlegen, nach der Zulassung Studien

Regorafenib für Patienten mit hepatozellulärem Karzinom (HCC)

EMA/CHMP EPAR ¹	EUnetHTA REA ²
<ul style="list-style-type: none"> ■ RESORCE-Studie, OS-Steigerung (2,8 Monate) gilt als klinischer Nutzen ■ Unsicherheiten: Sorafenib-intolerante Patienten; Patienten mit ECOG PS> 1 und/oder Child Pugh B → geregelt durch Änderungen in der Fachinformation (SmPC) 	<ul style="list-style-type: none"> ■ RESORCE-Studie, OS-Steigerung (2,8 Monate) gilt als moderate Verbesserung ■ Unzureichende Evidenz über die Auswirkungen auf die HRQoL („bedauerlich“ für Patienten im Endstadium) ■ Evidenzlücken: Sorafenib-intolerante Patienten; Patienten mit ECOG PS> 1 und/oder Child Pugh B → es werden weitere Forschungsdaten benötigt

Legende:

1. EMA/CHMP EPAR EMEA/H/C/002573/II/0020.

2. EUnetHTA REA Projekt-D: PTJA02

Quelle: Dr. Tafuri

Abbildung 1: Die unterschiedlichen Aufgabenbereiche der EMA- und HTA-Institutionen spiegeln sich auch in der Evidenz wider, die für die Beurteilung verwendet wird.

durchführen zu müssen, um die bestehende Evidenzlücke zum Zeitpunkt der Zulassung zu füllen, zeigte sich in der Analysen solcher Studien, dass mehr als die Hälfte entweder mit deutlicher Verzögerung oder gar nicht abgeschlossen wurden.⁶ Es gilt als hinreichend belegt, dass das aktuelle Szenario durch die unterschiedlichen Entscheidungen im Rahmen der Preisgestaltung und Erstattung und die Zeit bis zur Erstattung in den einzelnen EU-Ländern noch komplexer wird.⁷

Obwohl die weitere Angleichung von HTA- und regulatorischen Evidenzanforderungen sowie die Festlegung gemeinsamer Methoden im Bereich HTA sowie gemeinsame Bewertungen eine große Rolle spielen, um den Zugang für Patienten zu neuen Behandlungen zu verbessern, bleibt die größte Hürde der Preis.

Es ist mittlerweile international anerkannt, dass steigende Medikamentenpreise ein ernsthaftes Problem für die Nachhaltigkeit von Gesundheitssystemen darstellen. Eine häufige Begründung für hohe Medikamentenpreise seitens der Hersteller sind die beträchtlichen Forschungs- und Entwicklungsausgaben, die erforderlich sind, um ein Medikament auf den Markt zu bringen. Die Ausgaben für Forschung und Entwicklung werden auf 650 Millionen bis 2,7 Milliarden (2017) US-Dollar geschätzt.⁸

Deshalb wird mehr Transparenz bei den Preisdeterminanten der Hersteller benötigt. Hier wurden schon viele Vorschläge gemacht. So soll mit einer neuen von der WHO verabschiedeten italienischen Verordnung der öffentliche Informationsaustausch über aktuelle Preise, die Regierungen und andere Beschaffer von Gesundheitsprodukten bezahlen, verbessert werden und gleichzeitig mehr Transparenz bei Arzneimittelpatenten, Ergebnissen von klinischen Studien und anderen Determinanten für die Preisgestaltung entlang der Wertschöpfungskette vom Labor bis zum Patienten geschaffen werden.⁹

Große Profite können gerechtfertigt sein, wenn neue Produkte signifikante Vorteile für Patienten bieten oder einen bedeutenden pharmakologischen Fortschritt darstellen. In einer Analyse von Mailankody und Prasad von Onkologika, die zwischen 2009 und 2013 von der US-amerikanischen Zulassungsbehörde (Food and Drugs Administration, FDA) zugelassen wurden, stellten die Autoren keinen Zusammenhang zwischen der Neuheit oder dem relativen Vorteil dieser Medikamente und ihrem Preis fest.¹⁰ Tatsächlich wurde nur ein geringer Preisunterschied zwischen Medikamenten, die auf Grundlage der Time-to-Event-Endpunkte zugelassen wurden, und solchen festgestellt, bei denen die Zulassung auf Grundlage der Ansprechrate erfolgte. Die Ergebnisse lassen vermuten, dass die aktuellen Preismodelle nicht rational sind, sondern lediglich widerspiegeln, was der Markt bereit ist, für die Präparate zu bezahlen.

Villa et al. analysierten den Verhandlungsprozess der italienischen Arzneimittelagentur (AIFA) zwischen 2013 und 2017 für 133 neue Präparate (44 Orphan-Drugs und 89 anderen neuen Wirkstoffen). In ihrer Arbeit wiesen sie auf ein Missverhältnis zwischen dem von den Herstellern empfundenen Wert und dem Wert hin, den Kostenträger beziffern.¹¹ Nach dem Verhandlungsprozess wurden die Preise um durchschnittlich 25,1 bzw. 28,6 Prozent für Orphan-Drugs bzw. andere Moleküle gesenkt.

Bei innovativen Medikamenten gab es eine größere Preissenkung (-32,2 Prozent). Zudem fanden die Autoren heraus, dass sich durch die Implementierung von sogenannten Managed Entry Agreements (insbesondere Vereinbarungen über die Kostenteilung) bei den Verhandlungen größere Preissenkungen erzielen ließen.

Vor dem Hintergrund dieses komplexen Szenarios wird derzeit über eine neue EU-Rechtsvorschrift für die Zusammenarbeit im Bereich HTA diskutiert.¹² Eines ihrer Ziele ist

die verstärkte Zusammenarbeit bei Beurteilungen im Bereich HTA, die gemeinsam von Vertretern der nationalen HTA-Behörden durchgeführt werden sollen.

Die Idee, die dahinter steckt, ist, den nationalen EU-Behörden Berichte zur Verfügung zu stellen, in denen die verfügbare Evidenz mit Hilfe von Methoden zusammengeführt ist, die im Rahmen von Gemeinsamen EUnetHTA-Aktionen erarbeitet wurden, und so den Entscheidungsprozess auf nationaler Ebene zu erleichtern, Doppelungen zu reduzieren und gleichzeitig die Entscheidungen zur Preisgestaltung und Erstattung bei den einzelnen Mitgliedsstaaten zu belassen. Das Gesetzgebungsverfahren für diese Verordnung läuft, und ein Ergebnis wird im Laufe des Jahres 2020 erwartet.

Haftungsausschluss: Die Inhalte dieses Artikels entsprechen meiner persönlichen Meinung und liegen ausschließlich in meiner Verantwortung. Mit diesem Artikel soll nicht der Anschein erweckt werden, als würde er die Ansichten der Europäischen Kommission und/oder der Exekutivagentur für Verbraucher, Gesundheit, Landwirtschaft und Lebensmittel (engl. Consumers, Health, Agriculture and Food Executive Agency, CHAFEA) sowie anderer Institutionen der Europäischen Union widerspiegeln. Die Europäische Kommission und die Agentur übernehmen keine Verantwortung für die eventuelle Verwendung der Inhalte dieses Artikels.

Literatur:

¹ Patient Input in Relative Effectiveness Assessments, EUnetHTA. Online erhältlich: https://www.eunetha.eu/wp-content/uploads/2019/06/Final_290519_Patient-Input-in-REAs.pdf.

² Tafuri, G., Pagnini, M., Moseley, J. et al. *Br J Clin Pharmacol.* 2016;82:965-973.

³ Tafuri, G., Lucas, I., Estevão, S. et al. *Br J Clin Pharmacol.* 2018;84:1013-1019.

⁴ Eichler, H.G., Bloechl-Daum, B., Abadie, E. et al. *Nat Rev Drug Discov.* 2010, 4:277-91.

⁵ Davis, C., Naci, H., Gurpinar, E. et al. *BMJ.* 2017;359:j4530.

⁶ Pease, A.M., 1 Krumholz, M.H., Downing, N.S. et al. *BMJ* 2017;357:j168.

⁷ Variations Between European Countries in Time to Reimbursement Decisions for New Cancer Treatments Approved by the EMA. ESMO 2018. Online erhältlich: <https://www.ascopost.com/News/59385>.

⁸ Vinay, P., Mailankody, S. *JAMA Intern Med.* 2017;177(11):1569-1575.

⁹ 72. Weltgesundheitsversammlung. Improving the transparency of markets for medicines, vaccines, and other health products. Online erhältlich: https://apps.who.int/gb/ebwha/pdf_files/WHA72/A72_ACONF2Rev1-en.pdf.

¹⁰ Mailankody, S., Prasad, V. *JAMA Oncol.* 2015; 4:539-40.

¹¹ Villa, F., Tutone, M., Altamura, G. *Health Policy* 123 (2019) 595–600.

¹² Regulation of the European Parliament and of the Council on health technology assessment and amending Directive 2011/24/EU. Online erhältlich: https://ec.europa.eu/health/sites/health/files/technology_assessment/docs/com2018_51final_en.pdf.

Wie umgehen mit Real-World-Data und Registern? Eine politische Betrachtung

Michael Henrich | Mitglied des Deutschen Bundestages

Die steigende Zahl von Orphan Drugs verlangt eine Weiterentwicklung ihres Sonderstatus' im Rahmen des AMNOG-Verfahrens. Ziel der Gesundheitspolitik muss dabei sein, dass auch weiterhin diese innovativen Arzneimittel schnellstmöglich bei den Patienten ankommen, gleichzeitig jedoch auch ihr Nutzen belegt ist. Das Gesetz für mehr Sicherheit in der Arzneimittelversorgung (GSAV) stärkt hierfür die Rolle von Registern und Real-World-Data in der Nutzenbewertung¹. Der Beitrag rekapituliert in einem ersten Schritt mit Blick auf die bestehende Gesetzeslage sowie Beispiele für Register den bisherigen Sachstand. In einem zweiten Schritt wird die Neureglung im GSAV und damit einhergehende Probleme und Herausforderungen dargelegt.

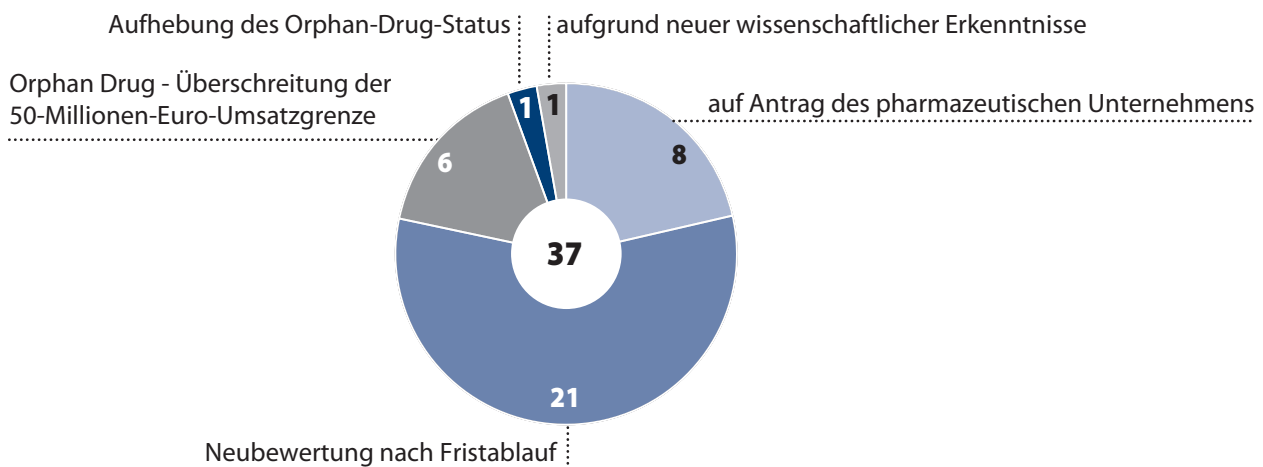
N **eu**e Wege in der Nutzenbewertung: **Die Sicht der Politik auf Register und Real-World-Data**

Eine Konstante der Politik in einer modernen Welt ist die Herausforderung, einen schnellen technologischen Fortschritt zu erfassen, den Nutzen zu definieren und dafür Sorge zu tragen, dass er im Alltag der Menschen ankommt.

Im Bereich der Arzneimittel finden wir ein Politikfeld vor, in dem diese Maxime in besonderem Maße gilt: einerseits weil wir hier einen schnellen Fortschritt erleben, andererseits weil es hier schlicht um das hohe Gut Gesundheit geht. Mit dem AMNOG-Verfahren haben wir grundsätzlich einen Modus vivendi, der Patienten einen schnellen Zugang zu den neuesten heilversprechenden Innovationen ermöglicht und dabei gleichzeitig die Gesamtausgaben des Systems im Blick hat. Gleichzeitig stehen wir vor der Herausforderung einer immer feineren Ausdifferenzierung innovativer Arzneimittel, mit der für viele Patienten verbesserte Heilungschancen einhergehen. Als Gesetzgeber wollen wir weiterhin, dass diese Innovationen schnell bei den Patienten ankommen.

Die Innovationen machen deutlich, dass die existierenden Regelungen an Grenzen stoßen und wir eine Weiterentwicklung der Ausnahmeregelung für Orphan-Arzneimittel benötigen. Ausnahmen bestimmen die Regel – und so soll es auch bleiben. Gleichzeitig macht die zunehmende Ausdifferenzierung innovativer Arzneimittel eine Überprüfung der geltenden Anforderungen an die erforderlichen Evidenzgrundlagen sowohl für die Zulassung als auch für den G-BA-Prozess notwendig. Für die Zulassung hat die Europäische Union mit der Verordnung (EG) Nr. 141/2000 sowie der Verordnung 726/2004 einen vernünftigen Rahmen geschaffen. Um die Wirkungsweise und den Nutzen von Orphan-Arzneimittel besser zu verstehen und

Anzahl der erneuten Nutzenbewertungen nach Gründen, 2011–2017



Quelle: BPI-MARIS 2018

Abbildung 1: Orphan Drugs spielen bei den Fällen einer erneuten Nutzenbewertung eine wichtige Rolle.



Michael Henrich hat Rechtswissenschaften in Passau und Bonn studiert. Von 1991-1995 arbeitete er als Wissenschaftlicher Mitarbeiter des Bundestagsabgeordneten Elmar Müller in Bonn. Seit 1995 ist er selbständiger Rechtsanwalt. Von 1998 bis 2003 war er Landesgeschäftsführer des Wirtschaftsrats der CDU Baden-Württemberg. Seit 2002 ist er Abgeordneter für den Wahlkreis Nürtingen, seit 2013 Berichterstatter für den Bereich Arzneimittelversorgung und Apotheken. Seit 2015 ist er Obmann im Gesundheitsausschuss.

zu dokumentieren, sollen deshalb im Bereich der frühen Nutzenbewertung zukünftig Register eine ergänzende und zunehmend wichtige Rolle spielen.

Als Instrument ermöglichen sie es, zwei Ziele miteinander zu verbinden: Weiterhin eine schnelle Verfügbarkeit innovativer Orphan-Arzneimittel für Patienten und gleichzeitig ein langfristiger Nachweis ihres Zusatznutzens. Im Rahmen des Gesetzes für mehr Sicherheit in der Arzneimittelversorgung (GSAV) werden wir diese Überlegungen auf gesetzlicher Ebene konkretisieren.

Blick auf das Bestehende: Rechtliche Grundlagen der Nutzenbewertung

Ausgangspunkt ist der Paragraph 35a SGB V. Er besagt, dass der G-BA den Nutzen von erstattungsfähigen Arzneimitteln mit neuen Wirkstoffen zu bewerten hat. Maßstab dieser Bewertung ist der Zusatznutzen gegenüber der zweck-

mäßigen Vergleichstherapie. Für die Umsetzung verweist Paragraf 35a in Absatz 1 Satz 9 auf eine durch das BMG (ohne Zustimmung der Länder) zu erstellende Rechtsverordnung. Bei den Anforderungen an diese ist insbesondere festzulegen:

„Grundsätze für die Bestimmung der zweckmäßigen Vergleichstherapie und des Zusatznutzens, und dabei auch die Fälle, in denen zusätzliche Nachweise erforderlich sind, und die Voraussetzungen, unter denen Studien bestimmter Evidenzstufen zu verlangen sind; Grundlage sind die internationalen Standards der evidenzbasierten Medizin und der Gesundheitsökonomie“.

Die zugehörige Arzneimittel-Nutzenbewertungsverordnung (AMNV) fasst den Zusatznutzen als therapeutische Verbesserung, dessen Nachweis im Wege von Fachinfor-

mationen und klinischen Studien zu erfolgen hat. Paragraf 5 Absatz 2 Satz 3 liefert uns auch gleich hierfür das Instrument der Wahl: „Vorrangig sind klinische Studien, insbesondere direkte Vergleichsstudien mit anderen Arzneimitteln dieser Festbetragsgruppe mit patientenrelevanten Endpunkten, insbesondere Mortalität, Morbidität und Lebensqualität, zu berücksichtigen.“

Der darauf folgende Absatz 3 liefert dann die Ausnahme, die sich im Verordnungstext zuvor bereits angekündigt hat. Insofern Studien mit höchster Evidenz nicht durchführbar sind oder nicht gefordert werden können, kommen die darunter liegenden Evidenzstufen ins Spiel. Im Paragrafen 5 Abs. 6 der AMV werden die Evidenzstufen gegliedert aufgezählt. Im Übrigen haben wir in der rechtlichen Systematik, hier im Europäischen Recht in Art. 14 Abs.

Klassen zur Beurteilung der Evidenz

Klasse		Anforderungen an die Studien
I	Ia	Evidenz aufgrund einer systematischen Übersichtsarbeit randomisierter, kontrollierter Studien (evtl. mit Metaanalyse)
	Ib	Evidenz aufgrund mindestens einer hoch qualitativen randomisierten, kontrollierten Studie
II	IIa	Evidenz aufgrund mindestens einer gut angelegten, kontrollierten Studie ohne Randomisierung
	IIb	Evidenz aufgrund einer gut angelegten, quasi-experimentellen Studie
III		Evidenz aufgrund gut angelegter, nicht experimenteller deskriptiver Studien
IV		Evidenz aufgrund von Berichten/Meinungen von Expertenkreisen, Konsensuskonferenzen und/oder klinischer Erfahrungen anerkannter Autoritäten

Quelle: in Anlehnung an Kleespies, C.; Kaise, T.; Sawicki, P.T.; für die Arbeitsgruppe Praktische evidenzbasierte Medizin, St. Franziskus Hospital, Köln; IeM - Institut für evidenzbasierte Medizin, Köln: Begriffe und Methode der evidenzbasierten Medizin - Ein Glossar.

Abbildung 2: Die randomisierte, kontrollierte Studie ist der Goldstandard bei der frühen Nutzenbewertung. Zugleich kennt der einschlägige Paragraf 35a Ausnahmetatbestände, wenn höchste Evidenzklassen nicht erreicht werden können.

8 der Verordnung 726/2004, eine Parallele zum nationalen deutschen Recht.²

Als Zwischenfazit können wir daher mit Blick auf die gegenwärtig bereits bestehende grundsätzliche Regelung festhalten: Die randomisierte kontrollierte Studie (RCT) ist der Goldstandard unseres Zulassungssystems sowie für die frühe Nutzenbewertung nach Paragraph 35a SGB V. Gleichzeitig kennt es Ausnahmebestände, wenn höchste Evidenzklassen nicht erreicht werden können. Und, auch hierauf sei hingewiesen, haben wir nicht zuletzt spezialrechtliche Regelungen im Paragraphen 5 Absatz 5a AMNV zu den europäischen Zulassungen.

Er ermöglicht es dem G-BA, einen Zusatznutzen für andere Patientengruppen oder Teilindikationen anzuerkennen, „sofern die Übertragung der Evidenz nach dem Stand der wissenschaftlichen Erkenntnis auch im Hinblick auf die Nutzenbewertung zulässig und begründet ist“.³

An dieser Stelle möchte ich eine erste Zwischenbewertung mit dem Blick auf die bestehenden Regelungen ziehen. Aus meiner Sicht bedarf es allenfalls Änderungen in der Arzneimittelverordnung. Einen weitergehenden Änderungsbedarf ließe sich auf der Ebene des SGB V nur schwer realisieren und stünde in Widerspruch zu der bisherigen Systematik. Damit verbinde ich aber auch die Aufforderung: Wer weitere notwendige Anpassungen sieht, diese klar und transparent zu formulieren und konkret aufzuzeigen, was erforderlich ist. Die Politik steht manchmal ratlos vor, mitunter sehr weitgehenden, aber unpräzisen und nicht ausformulierten Forderungen, die an uns herangetragen werden.

Wo stehen wir heute beim Thema Register?

Zur nötigen Klarheit trägt dabei auch der Blick auf bestehende Beispiele von Register bei. Ich möchte hier drei Bereiche anführen.

1. Klinische Krebsregister

Wenn wir über zusätzliche Evidenz aus Registern und aus Real-World-Data sprechen, dann fällt natürlich ein erster Blick auf die klinischen Krebsregister und die ihnen zugrundeliegenden Regelungen. Hier sei auf das Krebsregistergesetz vom 3. April 2013 verwiesen und den damals neu hinzugekommen Paragraphen 65c im SGB V. Die zentralen Passagen lauten:

1. „Zur Verbesserung der Qualität der onkologischen Versorgung richten die Länder klinische Krebsregister ein“ (Absatz 1, S. 1).

In Satz 2 folgt dann gegliedert in acht Unterpunkten die Spezifizierung der Aufgaben und die zu erfüllenden Parameter: In welchem Gebiet sind Daten zu erfassen, Auswertung und Rückmeldung der Ergebnisse an Leistungserbringer, Fragen der Zusammenarbeit etc.

2. „Die klinische Krebsregistrierung erfolgt auf der Grundlage des bundesweit einheitlichen Datensatzes [...]“ (Absatz 1, Satz 3). Sowie, auf diesen Knackpunkt sei verwiesen:

3. „Die für die Einrichtung und den Betrieb der klinischen Krebsregister nach Satz 2 notwendigen Bestimmungen einschließlich datenschutzrechtlicher Regelungen bleiben dem Landesrecht vorbehalten.“

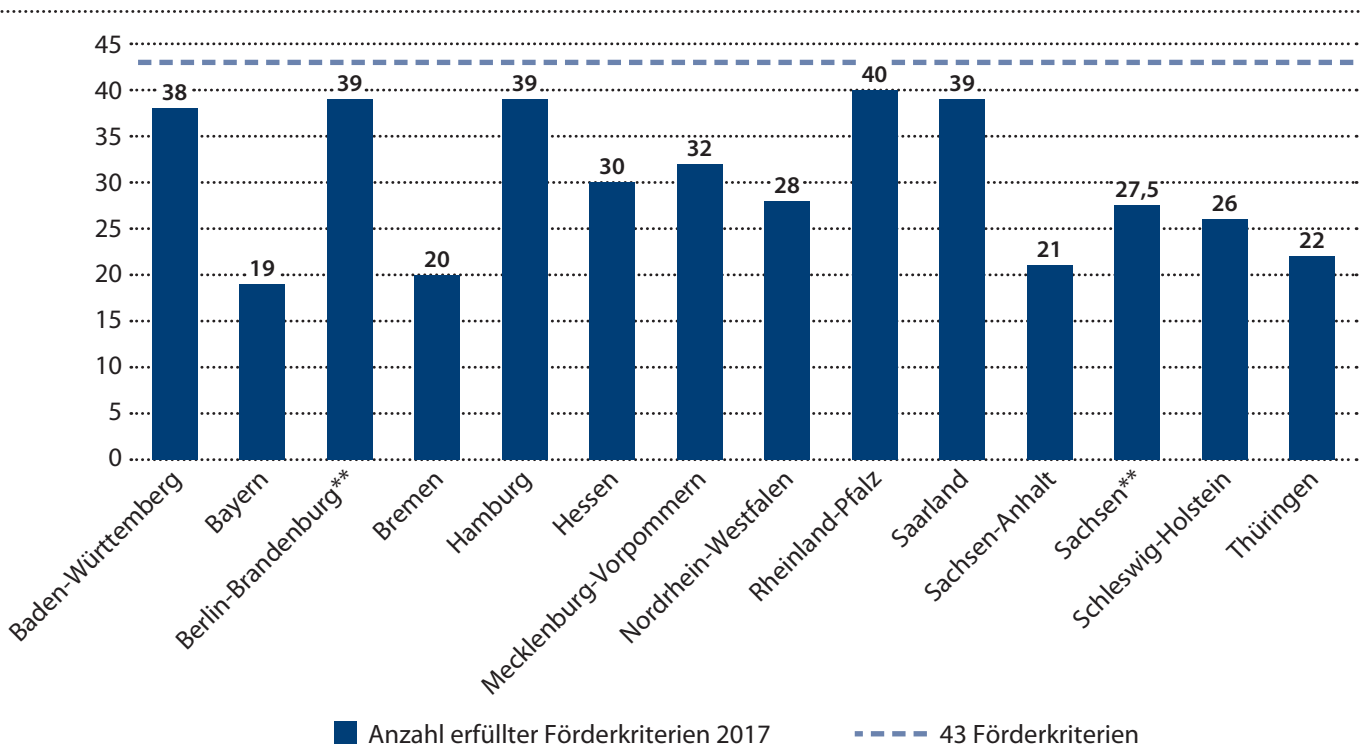
Eine Prognos-Studie aus dem Jahr 2017 zeigt, wie heterogen sich die Krebsregister-Landschaft entwickelt hat. Sie verweist auf einen unterschiedlichen Verfahrensstand sowie darauf, dass jedes Bundesland eigene Konzepte entwickelt hat und oftmals nur Basissätze vorliegen.⁴ Die Studie zeigt auf, dass der unterschiedliche Stand auf Faktoren wie der Weiterentwicklung der klinischen aus epidemiologischen Register zurückzuführen ist. Die Studie hält fest:

„Zwischen den klinischen Krebsregistern der einzelnen Bundesländer variiert der Erfüllungsgrad der Förderkriterien zum 31. Dezember 2017. So haben die klinischen

Krebsregister Baden-Württemberg, Berlin-Brandenburg, Hamburg, Rheinland-Pfalz und Saarland die Förderkriterien bis zum Stichtag weitestgehend erfüllt, während die übrigen klinischen Krebsregister teils deutlich hinter der Erfüllung der 43 Förderkriterien zurücklagen. Für manche klinische Krebsregister weisen die Erfüllungsberichte weniger als die Hälfte aller Förderkriterien als erfüllt aus.⁴⁵

Der Stand der Krebsregister macht deutlich, dass hier zwar einiges bewegt wurde, aber dennoch weiterhin Arbeit auf die verantwortlichen Akteure wartet. Darüber hinaus bestehen grundlegende Probleme. Aus meiner Sicht benötigen wir eine noch bessere Vernetzung zwischen den Akteuren und auf Ebene der Daten. Das bedeutet auch, dass wir die Kompatibilität der Daten sicherstel-

Anzahl erfüllter Förderkriterien der klinischen Krebsregister nach Bundesland* zum 31. Dezember 2017



Angaben basieren auf: Erfüllungsberichte 2017; *ohne Niedersachsen; **gemeinsames KKR Berlin-Brandenburg; ***Durchschnitt der dezentralen KKR Chemnitz (27 erfüllt), Dresden (28 erfüllt), Leipzig (28 erfüllt), Zwickau (27 erfüllt)

Quelle: Prognos AG, 2018

Abbildung 3: Nur fünf Bundesländer hatten Ende 2017 Krebsregister aufgelegt, die annähernd alle Förderkriterien erfüllt haben. Andere Bundesländer sind teilweise weit hinter diesen Vorgaben zurückgeblieben.

len und Doppelerfassungen vermeiden müssen. Grundsätzlich stehen hierfür zwei Wege offen: Wir können die bestehenden Systeme optimieren oder wir könnten eine neue Erfassungssystematik schaffen. Zwingend benötigen wir aber eine Zusammenführung der Ergebnisse.

2. Fallbeispiel RABBIT – Datenerfassung in der Rheumatologie: Mit seinem Logo eines flinken Hasens steht RABBIT auch als (nahezu) Akronym für den vollständigen Titel des Registers: Rheumatoide Arthritis: Beobachtung der Biologika-Therapie. RABBIT wurde 2001 vom Deutschen Rheumaforschungszentrum (DRFZ) in Abstimmung mit der Deutschen Gesellschaft für Rheumatologie, dem Berufsverband Deutscher Rheumatologen und dem Kompetenznetz Rheuma ins Leben gerufen. Das Register erfasst Krankheits- und Therapieverläufe von mehr als 17 000 Patienten, die von rheumatoider Arthritis betroffen sind. Ihnen stehen auf Seiten der Ärzteschaft deutschlandweit rund 400 Rheumatologinnen und Rheumatologen als Behandler gegenüber.

RABBIT untersucht die langfristige Sicherheit von Biologika, Biosimilars und JAK-Inhibitoren. Es handelt sich um eine reine Beobachtungsstudie mit einer Mindest-Teilnahmezeit von fünf Jahren. Finanziert wird das Register durch alle Unternehmen, die Biologika zur Therapie rheumatischer Arthritis herstellen. In seinem Design und der Umsetzung steht RABBIT für eine Möglichkeit, Evidenz mit Registern zu generieren. Für die mit dem GSAV beabsichtigten Änderungen Evidenz verstärkt auf anderen Wegen zu schaffen, kann es daher in vielen Aspekten als Beispiel dienen.

3. Fallbeispiel CRISP (Clinical Research platform into molecular testing, treatment and outcome of non-Small cell lung carcinoma Patients): Test- und Therapierealität bei Lungenkrebs:

Verantwortlich für diese große Studie ist die Arbeitsgemeinschaft internistische Onkologie in der Deutschen Krebsgesellschaft. Bei CRISP handelt es sich um eine große, offene, ebenfalls nicht interventionelle, prospektive Registerstudie an der 117 Zentren aktiv beteiligt sind. Erfasst werden die Daten von 8000 Patienten mit metastasiertem nicht-kleinzelligem Lungenkarzinom, wie Therapien bei ihnen wirken und welchen Einfluss sie auf die Lebensqualität haben. Patienten werden bis zu ihrem Ableben oder maximal drei Jahre begleitet.

Anhand dieser drei Beispiele werden sowohl Chancen als auch die Limitierungen einer weiterführenden Verwendung von Registerdaten im Rahmen der AMNOG-Prozesse sehr deutlich. So ist fraglos anzustreben, dass insbesondere bei extrem hochpreisigen Therapien, die zudem noch mit sehr begrenzten, oft einarmigen Datensätzen in den Markt eintreten, eine weitere systematische Datenerhebung im Versorgungsalltag die Einführung dieser Produkte flankiert. Andererseits zeigen z.B. die Krebsregister, wie herausfordernd es ist, bundesweit einheitliche Datensätze zu generieren. Aber auch das vorbildhaft eingeführte RABBIT-Register hat seinen Ausgangspunkt und Schwerpunkt in der Beurteilung des Risikoprofils der Biologika in der Rheumatologie. Aussagen zur vergleichenden Wirksamkeit wurden auch auf Basis des RABBIT-Registers kaum gemacht.

Die methodische Stringenz prospektiv angelegter randomisierter Datensätze aus klinischen Studien lässt sich unter Versorgungsbedingungen kaum nachbilden. Die im GSAV avisierte Erhebung von Registerdaten zum vergleichenden Nachweis eines Zusatznutzens bedarf aus diesem Grund selbstverständlich einer weiterführenden methodischen Diskussion zur Angemessenheit und Verwendung solcher Versorgungsdaten für vergleichende Wirksamkeitsbeurteilungen gemäß AMNOG.

Neureglung des Paragraphen 35a SGB V –

Inhalt und Ziele

Paragraf 35a Absatz 3b sieht in seiner Neufassung vor, dass der G-BA bei bestimmten Arzneimitteln eine anwendungsbegleitende Datenerhebung für eine Nutzenerhebung verlangen kann. Diese Arzneimittel werden in zwei Unterpunkten genauer gefasst, wobei Punkt 2 jene für seltene Leiden erfasst.

In dem Absatz wird weiter ausgeführt:

„Der Gemeinsame Bundesausschuss kann die Befugnis zur Verordnung des Arzneimittels zu Lasten der gesetzlichen Krankenversicherung auf solche Vertragsärzte oder zugelassene Krankenhäuser beschränken, die an der anwendungsbegleitenden Datenerhebung mitwirken.“ Die Zielrichtung ist eindeutig, betonen möchte ich, dass wir ein „kann“ in der Regelung vorfinden. Für das Nähere – Dauer, Art und Umfang, Auswertung und Formate – verweist der Paragraf im Übrigen auf den G-BA.

Ebenfalls neu: Die Beratungspflicht in Paragraf 35a Abs. 7 wird künftig weiter gefasst und die anwendungsbegleitende Datenerhebung wird ausdrücklich mit aufgenommen. Die gewonnenen Daten und die Erhebung hat der G-BA regelmäßig, mindestens einmal jährlich zu überprüfen. Dies in Abstimmung mit dem Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) sowie dem Paul Ehrlich-Institut. Adressat der Regelung sind die pharmazeutischen Unternehmen, die wiederum die anfallenden Beratungskosten bei den beiden Institutionen zu tragen haben (Absatz 7).

In der Begründung finden sich nähere Angaben. So ist mit dem Gesetz keine generelle Beschränkung auf Einrichtungen oder Zentren verbunden. Der Gesetzgeber macht zudem auch keine Vorgaben zum Studiendesign, etwa der Randomisierung. Welche Datenerhebung gefordert ist, bemisst sich nach der Zweck- und Verhältnismäßigkeit. Wich-

tig sind in diesem Kontext zudem die europäischen Regelungen, auf die sich Paragraf 35a Absatz 3 b bezieht.

Hier sei Art. 14 Abs. 7 und 8 der Verordnung (EG) 726/2004 angeführt:⁶

„(7) Nach Konsultation des Antragstellers kann eine Genehmigung vorbehaltlich besonderer Bedingungen erteilt werden, die jährlich von der Agentur neu beurteilt werden. Die Liste dieser Bedingungen wird öffentlich zugänglich gemacht.“

„(8) In Ausnahmefällen und nach Konsultation des Antragstellers kann eine Genehmigung vorbehaltlich der Verpflichtung des Antragstellers erteilt werden, besondere Verfahren zu schaffen, die insbesondere die Sicherheit des Arzneimittels, die Information der zuständigen Behörden über alle Zwischenfälle im Zusammenhang mit seiner Verwendung und die zu ergreifenden Maßnahmen betreffen. Diese Genehmigung kann nur aus objektiven und nachprüfbaren Gründen erteilt werden und muss auf einem der in Anhang I der Richtlinie 2001/83/EG genannten Motive beruhen. Die Aufrechterhaltung der Genehmigung ist von der jährlichen Neubeurteilung dieser Bedingungen abhängig.“

Mit Blick auf die Zulassung von Medikamenten gegen seltene Erkrankungen ist zudem die Verordnung (EG) 141/2000 maßgeblich.⁷ Sie hält die Auflagen und Bedingungen zur Zulassung von Medikamenten gegen seltene Erkrankungen durch die Europäische Arzneimittel-Agentur (EMA) fest. Spezifischer – hinsichtlich der Umsetzung – fasst es die Verordnung (EG) 507/2006. Hier möchte ich insbesondere die Art. 3-5⁸ hervorheben. Die zentralen Auszüge in Stichpunkten lauten:

- Auch für eine lediglich bedingte Zulassung wird ein positives Nutzen-Risiko-Verhältnis angestrebt.⁹
- Bedingte Zulassungen sollten auf Fälle beschränkt

bleiben, „in denen nur der klinische Teil der Antragsunterlagen weniger umfassend ist als üblich“ (Abs. 4).

- Ziel ist das Schließen medizinischer Versorgungslücken.
- Zulassungen sind mit Auflagen zu verbinden, wie die Erstellung von Studien, um ein positives Nutzen-Risiko-Verhältnis darzulegen.

Der Blick auf EU-Ebene zeigt, dass wir hier bereits eine umfängliche rechtliche Systematik vorliegen haben, um einen gangbaren Mittelweg zwischen schneller Verfügbarkeit von Arzneimitteln und dem Nachreichen von medizinischer Evidenz zu bestreiten. Ein anzuführender Kritikpunkt am GSAV ist sicherlich, dass diese existierenden Regelungen zu stiefmütterlich behandelt wurden. Mich erinnert das fatal an die Debatte um Orphan Drugs, als der damalige Vorsitzende des G-BA die Regulierung und Voraussetzungen für eine Erteilung des Orphan-Drug-Status im Gesetzgebungsverfahren nicht kannte.

Kritik an der Neuregelung

Patienten in Deutschland haben einen sehr guten und schnellen Zugang zu Orphan Drugs.¹⁰ Die Herausforderung der Novelle besteht also nicht zuletzt darin, einen bereits sehr guten Status quo weiter zu verbessern. Als Kritikpunkte wurde dabei insbesondere angeführt:

1. Die Ausweitung der Kalkulationsbasis der 50 Millionen Euro, die im Paragraphen 35a Absatz 1, Satz 12 als Umsatzgrenze für Orphan Drugs zu Grunde gelegt wird. Bislang bezieht sich diese Umsatzgrenze auf die vertragsärztliche Versorgung, zukünftig soll auch die stationäre Versorgung hier mit einbezogen werden.

Darüber hinaus sollen die Hersteller zur Auskunft über die Abgabe von Medikamenten verpflichtet werden. Hiergegen werden nun verschiedene Argumente angebracht.

I. Kritisiert wird etwa, dass diese Selbstverpflichtung den etablierten Interessenausgleich in Frage stellen würde.

II. Unternehmen können dies gar nicht leisten, da sie selbst die Medikamente und wo sie angewendet werden, nicht nachverfolgen können.

III. Eine Umsatzgrenze müsste zumindest an die Inflationsrate angepasst werden.

2. Ein weiterer Punkt betrifft die Auflagen zur Datenerhebung und ihre schlussendliche Anerkennung durch den G-BA. Hier zeigt sich, dass der G-BA in der Vergangenheit aus einarmigen Studien selten einem Wirkstoff einen anerkannten Zusatznutzen zugesprochen hat. Registerstudien können randomisierte kontrollierte Studien ergänzen und zusätzliche Evidenz kreieren, aber ob sie diese zukünftig ersetzen können und ob sie als Nachweis für einen Zusatznutzen Anerkennung finden, muss mit Blick auf Bisheriges zumindest kritisch betrachtet werden.

In der bisherigen Praxis jedenfalls setzt die Anerkennung eines Zusatznutzens vergleichende Evidenz voraus. Für einen Nachweis über Registerstudien bedeutet das eine qualitätsgesicherte Datenerhebung vor dem geplanten Marktzugang sowie indikationsspezifische Register.

3. Für die Bewertung des Nutzens von Arzneimitteln zur Behandlung seltener Leiden, bei bedingten Zulassungen und Zulassungen unter außergewöhnlichen Umständen erhält der G-BA künftig nach Paragraph 35 Absatz 3b SGB V die Befugnis, vom pharmazeutischen Unternehmer die Durchführung begleitender Datenerhebungen zu verlangen. Kommt der pharmazeutische Unternehmer dem nicht nach oder zeigt sich kein quantifizierbarer Zusatznutzen, können angemessene Abschläge auf den Erstattungsbeitrag vereinbart werden (Paragraph 130b SGB V). Hier stellt sich die Frage, was passiert, wenn Ärzte oder Patienten sich einer Datenerhebung verweigern. Muss sich ein Patient im Zweifelsfall dann an ein weit entferntes Zentrum wenden?

4. Eine fachspezifische Frage betrifft die Anforderungen

bezüglich der Evidenz bei den jährlichen Überprüfungen, die in Paragraph 35a Absatz 3b S.6 festgelegt sind. Hierbei ist sicherlich einschränkend darauf hinzuweisen, dass für valide Datenerhebungen ausreichend Zeit eingeräumt werden muss und gerade bei seltenen Erkrankungen die Verfügbarkeit einer ausreichenden Patientenzahl erforderlich ist.

Über diese Aspekte, die konkrete Regelungsinhalte betreffen, steht deshalb die grundsätzliche Frage, wie die Datenerhebung mit der europäischen Ebene verzahnt werden sollte und wie der G-BA mit den Daten umgeht. Im Rahmen der europäischen Zulassung findet bereits eine anwendungsbegleitende Datenerhebung statt. Daher ist die Frage berechtigt, ob dies am Ende zu Doppelstrukturen führt. Hierbei sei ausdrücklich auf die aktuelle Initiative der European Network for Health Technology Assessment (EUnetHTA) zu Etablierung methodischer Eckpfeiler für die Erstellung von Registern verwiesen.¹¹

Preisbildung¹²

Nach Ablauf der Frist zur Durchführung einer anwendungsbegleitenden Datenerhebung und nach erneutem Beschluss über die Nutzenbewertung wird der Erstattungsbetrag neu verhandelt (Paragraph 130b Abs. 3 Satz 7 SGB V). Wenn sich keine Quantifizierung des Zusatznutzens belegen lässt, dann sind geringere Jahrestherapiekosten zu vereinbaren, die unter dem zuvor vereinbarten Erstattungsbetrag liegen.

Hier bedarf es allerdings aus meiner Sicht einer weiteren gesetzlichen Einschränkung. Der Gesetzesentwurf sah keine Begründung für den Fall vor, wenn sich trotz Erhebung von Daten kein Zusatznutzen begründen lässt. Hier sind aber Konstellationen denkbar, bei denen der Hersteller dies durchaus nicht zu verantworten hat. Etwa wenn Daten durch behandelnde Ärzte nicht sachgemäß eingetragen

werden oder von Seiten der Technik Probleme bei der digitalen Verwaltung der Register bestehen. Zudem ist hier zu berücksichtigen, dass gemäß dem bereits genannten Entwurf der EUnetHTA der Hersteller an der ‚Governance des Registers‘ gar nicht beteiligt sein sollte, was seine Einflussnahme zwangsläufig limitiert.

Die Lösung könnte darin liegen, die Anforderung einzuschränken und dem G-BA einen größeren Spielraum einzuräumen. Statt einer Ist-Regelung wäre dies bei einer Soll-Regelung der Fall. Die Regelung könnte dann in einer Neufassung lauten: „Sofern [...] anhand der gewonnenen Daten keine Quantifizierung des Zusatznutzens belegen lässt, soll ein Erstattungsbetrag vereinbart werden, der in angemessenem Umfang zu geringeren Jahrestherapiekosten führt als der zuvor vereinbarte Erstattungsbetrag.“

Arzneimittel für neuartige Therapien im Arzneimittelgesetz (AMG, Paragraph 4 Absatz 9)

Ausgangspunkt ist hier zunächst die Entscheidung des G-BA. Verdeutlichen kann man dies anhand einer schematischen Übersicht: Das Bundessozialgericht hat in diesem Zusammenhang darauf verwiesen, dass es entscheidend ist, welches Gewicht der ärztliche Behandlungsanteil hat.¹³

Wenn es sich um ein ATMP handelt, dann müssen Mindestanforderungen an Struktur-, Prozess- und Ergebnisqualität festgelegt werden. Hierzu zählen insbesondere die Qualifikation der Leistungserbringer, strukturelle Anforderungen sowie sonstige Anforderungen an die Qualitätssicherung. Der G-BA erlässt hierfür die Durchführungsbestimmungen nach Paragraph 136 Abs. 2 und 3 SGB V. Eine Einbindung des Paul-Ehrlich-Instituts ist bei diesen Prozessen sicherlich wünschenswert und zielführend. Bei der letztgültigen Ausgestaltung des Gesetzes wird dieser Aspekt sicherlich noch eine Rolle spielen.

Arzneimittel oder ATMP

Einordnung als Arzneimittel (ambulant oder stationär) Dossierpflicht <ul style="list-style-type: none"> • Pharmazeutischer Charakter • Einfache Verabreichung 	Einordnung als Behandlungsmethode Keine Dossierpflicht <ul style="list-style-type: none"> • Komplexere Verabreichung • Verabreichung mit ärztlicher Intervention verbunden
→ Frühe Nutzenbewertung nach §35a SGB V	→ Bewertung nach §135 bzw. 137c SGB V

Quelle: eigene Darstellung

Abbildung 4: Entscheidend für die Frage, ob im konkreten Fall ein ATMP vorliegt, ist das Gewicht des ärztlichen Behandlungsanteils im Rahmen der Therapie.

Welche Rolle spielen Register und Real-World-Data?

Register werden heute bereits im Rahmen des Paragraphen 35 SGB V genutzt, wenn es um Prävalenz, Inzidenz und den Behandlungsverlauf geht. Bei seltenen Erkrankungen und bei Langzeitverläufen haben sie bereits ihre Bedeutung. Wie können Sie darüber hinausgehend zusätzliche Evidenz schaffen? Und wie können Sie möglicherweise RCTs ergänzen?

Wenn wir vor diesen Fragen stehen, dann sollte man sich vergegenwärtigen, dass Register in zwei Richtungen wirken können: Sie können Evidenz in Richtung Zusatznutzen untermauern, sie können aber auch positive Befunde relativieren. Für diese Erkenntnis reicht im Zweifelsfall ein Blick in das „Deutsche Ärzteblatt“: „Orale Antikoagulation: Wenn Studienergebnisse und die reale Welt divergieren“¹⁴. Der Artikel zeigt anhand von Daten dreier Krankenkassen, dass die direkten oralen Antikoagulanzen (DOAKs) unter Alltagsbedingungen hinsichtlich Morbidität und Mortalität

schlechter abschneiden als die herkömmlichen Vitamin-K-Antagonisten.

Als naheliegende Ursache steht die Vermutung im Raum, dass die Patienten der RCT uniform gehalten werden, gerade um die Studienergebnisse vergleichbar zu halten. Demgegenüber waren die Patienten, aus denen sich die RWDs zusammensetzen, älter und insgesamt kränker mit verschiedensten Vorerkrankungen. Wenn Real-World-Data wie in diesem Fall zeigen, dass weniger Patienten sterben oder einen Schlaganfall erleiden, dann müssen wir hierauf eine Antwort finden, wie wir damit umgehen.

Die detaillierten Antworten auf diese Fragen müssen wir im Rahmen der Gesetzgebung und der Praxis finden. Rahmenbedingungen bleiben dabei aber gesetzt:

- Die Organisation der Register soll durch unabhängige Institutionen erfolgen.
- Wir benötigen eine verbindliche Finanzierung der Register mit Beteiligung der Industrie.
- Die gewonnenen Daten müssen vollständig und geeignet sein für eine Verwendung im System der Nutzenbewertung.

Es ist absehbar, dass sich aus den Änderungen weitere rechtliche und systematische Fragen ergeben: Muss die Arzneimittelnutzenverordnung angepasst werden? Benötigen wir Änderungen an der Verfahrensordnung des G-BA? Muss das zugrundeliegende Methodenpapier des IQWiG angepasst werden?

Die erfreuliche Nachricht ist, dass der medizinische Fortschritt unaufhaltsam weitergeht und neue Heilungschancen schafft. Die Herausforderung bleibt es, einen Weg zu finden, damit er schnell und nutzenbringend bei den Patienten ankommt, ohne unser solidarisches Gesundheitssystem falsch zu belasten. Für die Politik und Selbstverwaltung bedeutet das: Nach der Reform ist vor der Reform.

Literatur:

¹ Der Beitrag basiert auf einer Rede, die ich während des laufenden Gesetzgebungsverfahrens im März 2019 auf der Tagung INTERDISZIPLINÄRE PLATTFORM ZUR NUTZENBEWERTUNG¹ gehalten habe. Das GSAV wurde vor Drucklegung des vorliegenden Beitrags am 5. Juni vom Bundestag verabschiedet.

² „In Ausnahmefällen und nach Konsultation des Antragstellers kann die Genehmigung für das Inverkehrbringen unter bestimmten Bedingungen, die insbesondere die Sicherheit des Arzneimittels, die Information der zuständigen Behörden über alle Zwischenfälle im Zusammenhang mit seiner Anwendung und die zu ergreifenden Maßnahmen betreffen, erteilt werden. Die Genehmigung für das Inverkehrbringen kann nur erteilt werden, wenn der Antragsteller nachweisen kann, dass er aus objektiven und nachprüfbaren Gründen keine vollständigen Daten über die Wirksamkeit und Sicherheit des Arzneimittels bei bestimmungsgemäßem Gebrauch zur Verfügung stellen kann, und muss auf einem der in Anhang I der Richtlinie 2001/83/EG genannten Gründe beruhen. Die Aufrechterhaltung der Genehmigung für das Inverkehrbringen ist von der jährlichen Neubeurteilung dieser Bedingungen abhängig.“, Verordnung (EG) Nr. 726/2004 des Europäischen Parlaments und des Rates vom 31. März 2004 zur Festlegung von Gemeinschaftsverfahren für die Genehmigung und Überwachung von Human- und Tierarzneimitteln und zur Errichtung einer Europäischen Arzneimittel-Agentur.

³ Bei der Bewertung von Arzneimitteln mit einer Genehmigung für die pädiatrische Verwendung im Sinne des Artikels 2 Absatz 4 der Verordnung (EG) Nr. 1901/2006 des Europäischen Parlaments und des Rates vom 12. Dezember 2006 über Kinderarzneimittel und zur Änderung der Verordnung (EWG) Nr. 1768/92, der Richtlinien 2001/20/EG und 2001/83/EG sowie der Verordnung (EG) Nr. 726/2004 (ABl. L 378 vom 27.11.2006, S. 1), die zuletzt durch die Verordnung (EG) Nr. 1902/2006 (ABl. L 378 vom 27.12.2006, S. 20) geändert worden ist, prüft der Gemeinsame Bundesausschuss, ob für Patientengruppen oder Teilindikationen, die von der Zulassung umfasst sind, die jedoch in der Studienpopulation nicht oder nicht hinreichend vertreten sind und für die die Zulassung aufgrund der Übertragung von Evidenz ausgesprochen wurde, ein Zusatznutzen anerkannt werden kann. Er kann in diesen Fällen einen Zusatznutzen anerkennen, sofern die Übertragung der Evidenz nach dem Stand der wissenschaftlichen Erkenntnis auch im Hinblick auf die Nutzenbewertung zulässig und begründet ist. Der Gemeinsame Bundesausschuss regelt das Nähere in seiner Verfahrensordnung.

⁴ Stand der klinischen Krebsregistrierung. Ergebnisse der Überprüfung der Förderkriterien zum 31.12.2017, Gutachten der Prognos-AG, S. 5 ff. Mit der Studie kommt der GKV-SV den gesetzlichen Anforderungen des § 65c Absatz 10 SGB V nach, ab dem Jahr 2018 alle fünf Jahre einen Bericht über die bundesweiten Ergebnisse der klinischen Krebsregistrierung zu veröffentlichen.

⁵ S. XV.

⁶ Verordnung (EG) Nr. 726/2004 des Europäischen Parlaments und des Rates vom 31. März 2004 zur Festlegung von Gemeinschaftsverfahren für die Genehmigung und Überwachung von Human- und Tierarzneimitteln und zur Errichtung einer Europäischen Arzneimittel-Agentur (Text von Bedeutung für den EWR), Amtsblatt Nr. L 136 vom 30/04/2004 S. 0001 - 0033.

⁷ Verordnung (EG) Nr. 141/2000 des Europäischen Parlaments und des Rates vom 16. Dezember 1999 über Arzneimittel für seltene Leiden, Amtsblatt Nr. L 018 vom 22/01/2000 S. 0001 - 0005.

⁸ Artikel 1 Nummer 28a der Richtlinie 2001/83/EG.

⁹ Die Definition des Nutzen-Risiko-Verhältnisses enthält Artikel 1 Nummer 28a der Richtlinie 2001/83/EG.

¹⁰ Vgl. bspw., Deticek, A., Locatelli, I. & Kos, M. (2018). Patient access to medicines for rare diseases in european countries. *Value Health*, 21, 553-560. Die Studie kommt zu dem Ergebnis, dass Deutschland sowohl hinsichtlich der Schnelligkeit in der Verfügbarkeit, als auch in den Ausgaben (berechnet auf die Einwohneranzahl) die beste Versorgung bietet.

¹¹ EUnetHTA. (2018). NetworkREQueST[®] tool and its Vision paper are now available for Public Consultation. Abgerufen am 22. Mai, 2019, von <https://www.eunetha.eu/request-tool-and-its-vision-paper-are-now-available-for-public-consultation/>.

¹² In den Verhandlungen zu dem Gesetz wurde der hier vorliegende Vorschlag zugunsten einer Verlängerung der Übergangsfrist verworfen.

¹³ BSG, Urteil vom 19. Oktober 2004 – B 1 KR 27/02 R –, BSGE 93, 236-252, SozR 4-2500 § 27 Nr 1.

¹⁴ Eckert, N. (2019). Orale Antikoagulation: Wenn Studienergebnisse und die reale Welt divergieren. *Dtsch Arztebl* 2019, 116(9): A-416 / B-344 / C-340. Abgerufen am 22. Mai, 2019, von <https://www.aerzteblatt.de/archiv/205793/Orale-Antikoagulation-Wenn-Studienergebnisse-und-die-reale-Weltdivergieren>.

Generierung zusätzlicher Evidenz – Nötig sind konsentierete Spielregeln

Von Dr. Florian Staeck

Sind bei der frühen Nutzenbewertung von Arzneimitteln randomisierte, klinische Studien (RCT) nicht verfügbar oder soll deren Validität besser abgeschätzt werden, so stellt sich die Frage nach den Methoden zur Gewinnung kontextueller Evidenz sowie deren Stellenwert im Gesamtgefüge der frühen Nutzenbewertung.

Anlässlich der 9. Tagung der Interdisziplinären Plattform zur Nutzenbewertung am 15./16. März 2019 in Fulda erörterten die Teilnehmer, mittels welcher Instrumente Daten jenseits von RCT aufbereitet und für die Festlegung eines Zusatznutzens gangbar gemacht werden können. Ein Ziel war dabei, auch jenseits des deutschen AMNOG-Kontexts von Verfahrensweisen anderer HTA-Behörden zu lernen, um deren Potenzial für eine mögliche Anwendung in Deutschland abschätzen zu können. Deutlich wurde im Verlauf der Tagung, dass sich diese Diskussion in Deutschland noch in einem frühen Stadium befindet.

Als ein mögliches Werkzeug für die Beurteilung des Vertrauens in Effektschätzer für Studienendpunkte diskutierten die Teilnehmer die im Ausland bereits seit vielen Jahren konzeptualisierte GRADE-Methodik (Grading of Recommendations, Assessment, Development and Evaluation). Ziel ist es hierbei, die Qualität von Evidenz nicht nur als Bias-Risiko einzelner Studien wiederzugeben, sondern in einem systematischen Ansatz die Präzision und mögliche Inkonsistenz von Studienergebnissen zu beleuchten. Dazu werden bei dem Konzept eindeutige Kriterien zur Einschätzung der Qualität der Evidenz wie Studiendesign, Bias-Risiko, Inkonsistenz oder Stärke des Effekts verwendet. GRADE könne dazu beitragen, die Arbeitsschritte der Evaluation zu strukturieren und damit eine Basis für eine transparente Entscheidungsfindung zu produzieren.

Dabei gehe es um eine bessere Operationalisierung der evidenzbasierten Medizin, nicht aber um eine Senkung

von Standards, betonten Teilnehmer. Hinzu komme, dass in den vergangenen 10 bis 15 Jahren im GRADE-Konzept keine neuen Bewertungsdomänen hinzugekommen sind – insoweit habe sich ein Expertenkonsens herauskristallisiert, welche Domänen bei der Bewertung der Qualität von Evidenz entscheidend sind.

Bei der Diskussion, ob GRADE im deutschen Bewertungskontext fruchtbar gemacht werden kann, wurden verschiedene Positionen artikuliert. Einerseits betonten Teilnehmer, das GRADE-Konzept werde in Deutschland aufmerksam beobachtet und sei in Form von Einzelelementen auch bereits adaptiert worden. Andererseits wurde hervorgehoben, Fragen wie beispielsweise die nach der Verteilungsgerechtigkeit oder der Gleichheit des Zugangs zu neuen Arzneimitteln oder Behandlungsmethoden würden im deutschen Bewertungskontext bisher selten oder aber gar nicht thematisiert. Insoweit könne der GRADE-Ansatz eine methodische Bereicherung darstellen.

Nutzen-Risiko-Analyse bei der Zulassung

Nicht nur bei der Nutzenbewertung, sondern auch bei der Zulassung von Arzneimitteln oder Impfstoffen geht es um den Umgang mit Unsicherheit in der Beurteilung von Evidenz. Gefordert wird dabei eine Nutzen-Risiko-Analyse, und zwar unabhängig davon, dass im Arzneimittelgesetz der Begriff „Evidenz“ gar nicht verwendet wird. Bei einer Zulassung in Deutschland wird großen Wert auf einen vollständigen Datensatz über ein neues Produkt gelegt. Die Behörden verfügten über ein Arsenal regulatorischer Maßnahmen, um im Fall noch nicht ausreichender Evidenz reagieren zu können – beispielsweise durch eine bedingte Zulassung unter Auflagen. Angemerkt wurde von Teilnehmern, dass die Zulassungsinstanz über die ausdrückliche Möglichkeit der Durchsetzung von Datennachlieferungen verfüge – ein Punkt, an dem der Gemeinsame Bundesaus-

schuss in der Vergangenheit nicht selten gescheitert sei. Zwar gebe es bei Zulassungsverfahren eine klare Präferenz für RCT, doch würden in Ausnahmefällen auch einarmige Studien oder historische Vergleiche als Basis für weitere Evidenz herangezogen, hieß es. Da viele Produkte das Verfahren der Zulassung durchlaufen, verfügten die Experten der Behörde über eine breite Erfahrung, die ihnen Vergleiche hinsichtlich der Einschätzung von Evidenz erlaubten.

Unterschiedliche Auffassung wurden darüber laut, inwieweit die Bindung eines Assessors der Zulassungsbehörde an ein Produkt über dessen Lebenszyklus hinweg mit Interessenkonflikten einhergehen kann. Diese Situation könne insbesondere dann eintreten, wenn ein neues Produkt vom Assessor bereits ab einem frühen Stadium der Entwicklung begleitet wird. Dem wurde entgegengehalten: Entscheidungen fielen bei der Zulassungsbehörde immer in der Gruppe. Außer dem Co-Rapporteur gebe es als weiteres Sicherungssystem immer auch ein Peer-Review-Verfahren, so dass Interessenkollisionen im Zulassungsprozess weitestgehend eliminiert werden.

Die konkreten Herausforderungen beim Umgang mit abweichenden Evidenzlagen diskutierten die Teilnehmer am Beispiel einarmiger Studien. Es gebe einen breiten wissenschaftlichen Konsens darüber, dass die Anwendung nicht adjustierter indirekter Vergleiche – zum Beispiel durch Berücksichtigung einzelner Studienarme unterschiedlicher Studien – nicht angemessen ist, wurde ausgeführt. Doch auch durch Adjustierung lasse sich beispielsweise mit Blick auf die Studienpopulation niemals der gleiche Effekt wie in einem RCT durch die Verblindung erreichen. Dennoch seien Konstellationen denkbar, in denen Daten aus derartigen Studien nicht völlig verworfen werden können, um kontextuelle Evidenz zu generieren, hieß es vorsichtig. Allerdings sei der klinisch relevante Nutzen im Kontext der frühen Nutzenbewertung vom Einzelfall

abhängig – etwa einer hohen Nebenwirkungsrate der zweckmäßigen Vergleichstherapie (zVT) oder der Eignung dieser Daten im Verhältnis zur zVT.

Nach den bisherigen Erfahrungen im Verfahren der frühen Nutzenbewertung sind solche Studien vergleichsweise häufig im Zusammenhang mit Wirkstoffen gegen Hepatitis C eingereicht worden – hier hat der G-BA stark überproportional häufig aus den zusätzlichen Daten einen Zusatznutzen abgeleitet. In der Onkologie seien derartige Datensätze dagegen aus Sicht des Bundesausschusses nur sehr selten geeignet gewesen. Doch auch im günstigsten Fall wurde aus Daten beispielsweise einarmiger Studien nie mehr als ein Anhaltspunkt für einen Zusatznutzen abgeleitet. Das Ausmaß wurde dabei als nicht quantifizierbar oder als gering eingeschätzt, wurde berichtet.

Sinnvolles Instrument oder Ressourcenvergeudung?

In der Diskussion wurden solche Ergebnisse besprochen, doch überwog Skepsis, ob es – außer in Fällen dramatischer Effekte – nicht „programmierte Ressourcenverschwendung“ aus Sicht der Dossierhersteller sei, Geld in derartige Studien zu investieren. Dem wurde entgegengehalten, durch einarmige Studien könne gegebenenfalls eine Beobachtungsgleichheit dahingehend erreicht werden, dass Endpunkte in verschiedenen Studien identisch erhoben worden sind. Oder es könne durch eine qualitativ hochwertige prospektive Datenerhebung die Möglichkeit der Einschätzung von Daten insgesamt verbessert werden. Allerdings sei es von den Umständen des Einzelfalls abhängig, ob dadurch die Gesamteinschätzung des Zusatznutzens eines Produkts positiv verändert wird.

Die aktuell fehlende Erwartungssicherheit vieler beteiligter Stakeholder prägte auch die Diskussion über die Wertigkeit kontextueller Evidenz, die anhand der Onkologie und der Behandlung von Patienten mit Diabetes melli-

tus Typ 2 exemplarisch beleuchtet wurde. In der Diabetologie ist im Oktober 2018 mit den neuen Leitlinien der American Diabetes Association (ADA) und der European Association for the Study of Diabetes (EASD) zur Therapie des Typ-2-Diabetes ein Paradigmenwechsel eingeläutet worden, demzufolge die Behandlung nicht mehr primär in Abhängigkeit von der Glukosestoffwechseleinstellung, unter anderem beurteilt am HbA1c-Wert erfolgt, sondern mehr patientenzentriert.

Demnach sollen die Herz-Kreislaufsituation, aber auch andere klinische Besonderheiten, stärker als bisher bei der Behandlung im Vordergrund stehen. Entsprechend verschiebe sich die Perspektive weg von einem „One size fits all“-Ansatz hin zur Beachtung und Behandlung mikro- und makrovaskulärer Komplikationen. Somit stünden die Reduktion von Morbidität und die Erhaltung guter Lebensqualität im Vordergrund der Behandlung. Diese Akzentverlagerung führe dazu, dass der Fokus viel stärker als bisher auf einem gesamttherapeutischen Ansatz liege.

Das habe auch Auswirkungen auf das Design klinischer Studien. Das Plädoyer lautete, die Kriterien des SGB V gleichrangig ernst zu nehmen, also neben der Mortalität auch Morbidität und gesundheitsbezogene Lebensqualität zu beachten. Diese müssten dann bezogen auf Endpunkte operationalisiert und evaluiert werden. Viel zu selten stehe bisher in Studien die Frage im Vordergrund, ob die gesundheitsbezogenen Probleme der Patienten tatsächlich gelöst werden. Hier stelle sich das Problem, dass der höhere Aufwand der Berücksichtigung beispielsweise von Retinopathien oder Neuropathien in Studien sich nur dann lohne, wenn absehbar ist, wie damit in der frühen Nutzenbewertung umgegangen wird.

Der Fokus verschiebe sich als Folge des Paradigmenwechsels von der Einzelerkrankung hin zu einer Risikointervention. Hierfür müssten die Instrumente in der Diabe-

tologie aber noch spezifischer und sensitiver gemacht werden, lautete eine Forderung. Mit einer gewissen Skepsis wiesen Teilnehmer auf die stark unterschiedlichen Ausgangssituationen in der Diabetologie und der Onkologie hin. Prinzipiell sei es möglich, Daten zur Morbidität und Lebensqualität beim Patienten abzufragen. Doch dadurch werde der Grad der Komplexität von Studien nochmals zunehmen, wurde zu Bedenken gegeben.

Lebensqualität wird immer häufiger erfasst

Die Ausgangsvoraussetzungen in der Onkologie stellen sich deutlich anders dar. So ist etwa der Anteil der Studien, bei denen die Lebensqualität erfasst wurde, in den vergangenen Jahren stetig von 60 auf zuletzt rund 80 Prozent gestiegen. Zudem spielt bei den Evidenzgewinnungskonzepten in der Onkologie die molekulare Diagnostik eine herausragende Rolle. Die Companion-Diagnostik bewege sich in Deutschland auf einem hohen Niveau, wurde berichtet. Die Förderung der Deutschen Krebsgesellschaft sei beim Aufbau regionaler Diagnostikzentren zusammen mit den jeweiligen Zuweisernetzen sehr hilfreich.

Randomisierte klinische Studien seien in der Mehrzahl der Indikationen möglich. Besondere Herausforderungen stellten sich immer dann, wenn einzelne Krebs-Entitäten biologisch sehr heterogen sind, so dass es mitunter schwierig wird, ein eindeutig definiertes Patientenkollektiv zu benennen. Ähnlich wie in der Diabetologie könne es auch in diesem Fall keinen „One size fits all“-Ansatz geben. Dies gelte zumal dann, wenn Therapiestandards und Biomarker sich während der Studienlaufzeit ändern. Vor dem Hintergrund dieser Herausforderungen werde auch künftig die frühe Nutzenbewertung neuer Arzneimittel mit einer hohen Datenunsicherheit belastet bleiben.

Aktuelle Fragen der konkreten politischen Ausgestaltung von Datenerhebungen, die zusätzliche Evidenz gene-

rieren sollen, diskutierten die Teilnehmer am Beispiel des geplanten Gesetzes für mehr Sicherheit in der Arzneimittelversorgung (GSAV). Vorgesehen war im Entwurf, der zum Zeitpunkt der Tagung im März 2019 vorlag, dass der G-BA im Falle von Orphan Drugs die Möglichkeit erhalten soll, anwendungsbegleitende Datenerhebungen oder Auswertungen zum Zweck der Nutzenbewertung zu fordern. Die so gewonnenen Daten sowie die Durchführung der Datenerhebung soll der G-BA laut Gesetzentwurf mindestens jährlich überprüfen.

Diese geplante Bestimmung rief unter den Teilnehmern eine kontroverse Diskussion hervor. Einerseits wurde es begrüßt, dem G-BA die Möglichkeit zu geben, fehlende oder unzureichende Evidenz einzufordern. Andererseits wurde moniert, dass der Gesetzgeber de facto nicht-interventionelle Studien vorgebe, mit denen die Akteure der Zusatznutzenbewertung zumindest bisher nichts hätten anfangen können, hieß es. So sei es auch nach acht Jahren AMNOG nicht einmal gelungen, beispielsweise aus Registerdaten einen quantifizierbaren Zusatznutzen abzuleiten.

Die geplante Bestimmung im GSAV zur Möglichkeit zu Abschlüssen sei ein dem AMNOG-System bislang systemfremdes Element, wurde kritisiert. Derartige Sanktionen habe es bislang nur in ausdrücklichen Fällen von Non-Compliance gegeben. Besonders skeptisch wurde bewertet, dass der G-BA bereits ein Jahr nach Start der Datenerhebung die Ergebnisse bewerten soll. Diese Frist sei etwa mit Blick auf Daten aus Registern viel zu kurz bemessen.

Ein anderslautendes Plädoyer von Teilnehmern ging dahin, das Signal des Gesetzgebers zu sehen, dass auch andere Modi von Evidenz Eingang in das AMNOG-Verfahren finden können. Insoweit solle die anwendungsbegleitende Datenerhebung von pharmazeutischen Unternehmen eher als Chance gesehen werden, nicht als Bedrohung. Das wurde unterstrichen mit der Bemerkung, es werde für sich

genommen einen Mehrwert darstellen, wenn im Zuge der anwendungsbegleitenden Erhebungen Daten im deutschen Versorgungskontext gewonnen werden können.

Hohe Anforderungen an Datenerhebung

Im konkreten AMNOG-Verfahren werde es für den G-BA nur die Pflicht geben, nachträglich erhobene Daten zu berücksichtigen. Hingegen existiere kein Recht des dossierreichenden Herstellers, dass aus den Daten auch eine positive Zusatznutzenbewertung abgeleitet wird. Hinzu komme, dass es sich im konkreten Fall immer um kleine Patientenpopulationen handeln werde. Vor diesem Hintergrund könnten auch die Daten nicht durch Einschreibemodelle für Patienten gewonnen werden, die lückenhaft bleiben würden. Abschließend plädierten Teilnehmer dafür, die neuen Verfahren der Evidenzgewinnung nicht zu schnell scharf zu stellen, um so das AMNOG – wie bisher – als lernendes System fortzuschreiben.

Mit Interesse verfolgten die Teilnehmer die ersten Impulse, die von der Arbeit auf Ebene des European Network for Health Technology Assessment (EunetHTA) ausgehen. Ziel der Partner auf EU-Ebene ist es seit 2004, die Zusammenarbeit bei der Bewertung von Gesundheitstechnologien zu verstetigen. Es habe in der Vergangenheit relativ wenig Interesse von Herstellern gegeben, mit EunetHTA-Wissenschaftlern ins Gespräch zu kommen, wurde berichtet. Dies ändere sich langsam. Dazu beitragen würden auch sechs weitere Reports, die bei dem Expertenkonsortium in Erarbeitung sind. Das EunetHTA-Projekt sei als Lernprozess angelegt, der die Chance biete, das Vorhaben von einem akademischen Projekt zu einer Art Referenz weiterzuentwickeln, an der sich nationale HTA-Behörden orientieren können. Ziel sei es, dass sich die Mitgliedsländer durch die strukturierte Datengewinnung bei EunetHTA aus einem breiten Fundus an Daten zur Evidenz bedienen können.

BEIRAT DER INTERDISZIPLINÄREN PLATTFORM

Dr. Jürgen Bausch
Wolfgang van den Bergh
Dr. Markus Frick
Dr. Antje Haas

Dr. Harald Herholz
Prof. Dr. Wolfgang Greiner
Dr. Ulf Maywald
Dr. Heinz Riederer

Prof. Dr. Jörg Ruof
Dr. Sibylle Steiner
Dr. Florian Staeck
Prof. Dr. Bernhard Wörmann

DISKUSSIONSTEILNEHMER

Dr. Jürgen Bausch
Sabine Beckmann
Dr. Antje Behring
Dr. Johannes Bruns
Claus Burgardt
Prof. Dr. Klaus Cichutek
Prof. Dr. Eva-Susanne Dietrich
Dr. Mathias Flume
Dr. Markus Frick
Prof. Dr. Wolfgang Greiner
Dr. Antje Haas
Michael Hennrich
Dr. Harald Herholz

Philipp Huwe
Dr. Thomas Kaiser
Dr. Dietrich Knoerzer
Lothar Kuntz
Dr. Wolfgang Lang-Heinrich
Martin Lack
Dr. Eberhard Lüdtke
Dr. Ulf Maywald
Dr. Thomas Mittendorf
Prof. Dr. Dirk Müller-Wieland
Dr. Heinz Riederer
Prof. Dr. Jörg Ruof
Prof. Dr. Dr. Holger J. Schünemann

Dr. Florian Staeck
Dr. Giovanni Tafuri
Dr. Katharina Thiele
Wolfgang van den Bergh
Dr. Julia Wagle
Frank Wallbrecht
Prof. Dr. Jürgen Wasem
Jan Weißflog
Dr. Elena Wolff-Holz
Prof. Dr. Bernhard Wörmann

IMPRESSUM**HERAUSGEBER**

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

VERLAG

Springer Medizin Verlag GmbH
Wolfgang van den Bergh,
Chefredakteur
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 Staeck
Wolfgang van den Bergh

AUTOREN

Prof. Dr. Klaus Chichutek
Michael Hennrich
Dr. Thomas Kaiser
Prof. Dr. Dirk Müller-Wieland
Prof. Dr. Dr. Holger J. Schünemann
Dr. Florian Staeck
Dr. Giovanni Tafuri
Prof. Dr. Bernhard Wörmann
Dr. Elena Wolff-Holz

BILDNACHWEIS

Titelbild: sdecoret / stock.adobe.com

LAYOUT / GRAFIK

Sandra Bahr
Oliver Hippmann

DRUCK

F&W Druck- und Mediacenter GmbH
Holzhauser Feld 2, 83361 Kienberg

 Springer Medizin

© Springer Medizin Verlag GmbH,
Berlin, Oktober 2019
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 GmbH

INTERDISZIPLINÄRE PLATTFORM ZUR NUTZENBEWERTUNG

Kontextuelle Evidenz – Wege zur gezielten Therapie

Ausgabe 9
Oktober 2019
ISBN 2364-916X