

Bundesgesundheitsbl 2015 · 58:298–307  
 DOI 10.1007/s00103-014-2106-1  
 Online publiziert: 8. Januar 2015  
 © Die Autor(en) 2014. Dieser Artikel ist auf  
 Springerlink.com mit Open Access verfügbar

Karl Wegscheider<sup>1</sup> · A. Drabik<sup>1</sup> · C. Bleich<sup>2</sup> · H. Schulz<sup>2</sup>

<sup>1</sup> Institut für Medizinische Biometrie und Epidemiologie, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Deutschland

<sup>2</sup> Institut für Medizinische Psychologie, Universitätsklinikum Hamburg-Eppendorf, Hamburg, Deutschland

## Nutzenbewertung aus Sicht der Versorgungsforschung und der Epidemiologie

### Die Nutzenbewertung von Arzneimitteln und Medizinprodukten

Der Begriff der Nutzenbewertung findet vielfältige Verwendung im Kontext mit dem Arzneimittelrecht (AMG), im Recht der gesetzlichen Krankenversicherung (GKV), d. h. im SGB V, in der Verfahrensordnung des Gemeinsamen Bundesausschusses (GB-A) [1] sowie im Methodenpapier des IQWiG [2] und schließlich auch in der Literatur zu Evidence Based Medicine oder Health Technology Assessments<sup>1</sup>. Will man eine Nutzenbewertung vornehmen, die eine kollektive Verbindlichkeit beanspruchen kann und sich als Grundlage für politische Entscheidungen eignet, so kommt man nicht ohne die Festlegung eines Verfahrens aus, wie man Nutzen und Schaden zu quantifizieren gedenkt und welche Konsequenzen mit welchen Ergebnissen verbunden sein sollen. Hierzu müssen sowohl die Qualitätsanforderungen an Daten als auch die Messmethoden für den Nutzen, die Auswertemethoden und die Regeln für die nach dem jeweiligen Ergebnis zu ziehenden Schlussfolgerungen festgelegt werden. An dieser Stelle haben Politik und Wissenschaft, die im Wesentlichen unterschiedlichen Aufgaben nachgehen, einen gemeinsamen Gestaltungsauftrag. Im Bereich der Arzneimittelzulassung wurde dieser Auf-

trag im 20. Jahrhundert überwiegend von den Arzneimittelbehörden mit Unterstützung externer Experten wahrgenommen. Mit einem verschärften Bewusstsein für die besonderen Herausforderungen, die mit der Datenbeschaffung und methodischen Bewältigung der Nutzenbewertung verbunden sind, wurden in einigen europäischen Ländern in diesem Jahrhundert zusätzliche Behörden oder wissenschaftliche Institute gegründet, die diese Aufgabe heute wahrnehmen, in Deutschland das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Das IQWiG bewertet bei Beauftragung die Evidenzlage für den Nutzen bzw. Zusatznutzen eines Produktes und quantifiziert das Ausmaß des Zusatznutzens für Kosten-Nutzen-Analysen. Die zugeordneten politischen Entscheidungen trifft in Deutschland der G-BA als höchstes Organ der Selbstverwaltung. Die Grundlagen dafür sind gesetzlich verankert, vor allem im SGB V sowie den darauf bezogenen untergesetzlichen Regelungen wie insbesondere der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV; siehe u. a. den Beitrag von Kaiser et al. in diesem Heft).

Mit der Nutzenbewertung greifen IQWiG und G-BA selbst in das Versorgungsangebot ein und verändern durch den Fortfall oder das Hinzukommen von Angeboten den Zusatznutzen von Konkurrenzangeboten. Die resultierende Versorgungslandschaft ist Gegenstand der Versorgungsforschung, die sich in einem sehr umfassenden Sinne mit der Kranken- und Gesundheitsversorgung beschäftigt. Versorgungsforschung untersucht die in der Routine erbrachten Versorgungsleis-

tungen als Ergebnis aus Gesundheits- (z. B. Behandlung) und Kontextleistung (Gestaltung des Versorgungsumfeldes). Ein bedeutender Fokus der Versorgungsforschung ist dabei die Untersuchung des unter Alltagsbedingungen erreichten Nutzens präventiver, kurativer, rehabilitativer und palliativer Versorgungsleistungen für den Patienten und die Gesellschaft [3]. Im Unterschied zu G-BA und IQWiG, die ihre Bewertungen nach Produkt und Indikation gliedern und dabei möglichst allgemeingültige Aussagen anstreben, ist der Ausgangspunkt der Versorgungsforschung dabei eine konkrete Versorgungssituation, in der eine oder auch mehrere Versorgungsangebote in einem beschreibbaren Umfeld zur Verfügung stehen und von Patienten und Gesunden mit unterschiedlichen Präferenzen angenommen oder nicht angenommen werden. Dabei spielen die Beweggründe, die selbst zum Gegenstand der Forschung werden, ebenso eine große Rolle wie die Bildung von Patientenclustern mit unterschiedlichen Präferenzen. Im Idealfall nimmt die Versorgungsforschung eine in Zeit und Ort verankerte Bewertung der Versorgungslage und des Nutzens einer Vielfalt von Angeboten aus Sicht verschiedener Patientengruppen vor. Ziel ist nicht die allgemeine auf möglichst alle Situationen passende Nutzenbewertung eines Produktes, sondern die Darstellung der Bedingungsfaktoren, unter denen ein Versorgungsangebot einen Nutzen bewirken kann, der den möglichen Schaden überwiegt. Zunehmend geraten dabei auch komplexe Interventionen (z. B. Vorsorge- oder Früherkennungsprogramme) in den Fokus, die in der Regel aus vielen inein-

<sup>1</sup> Zu einer Definition von Nutzenbewertung und Einordnung des Begriffes z. B. bei Medizinprodukten, Arzneimitteln, medizinischen Leistungen oder im Rahmen von diagnostischen Maßnahmen siehe die Beiträge in diesem Heft von Zens et al., Kaiser et al., Windeler und Lange oder Vach et al.

andergreifenden einzelnen Maßnahmen bestehen. Die Komplexität besteht darin, dass diese Maßnahmen nur gemeinsam in ihrer Gesamtheit und nur simultan im Hinblick auf ihren spezifischen Beitrag bewertet werden können. Die Bewertung des Einzelbeitrages ist dabei wichtig zur weiteren Verbesserung des Programms. Zur Bewältigung der damit verbundenen Aufgaben bedient sich die Versorgungsforschung der Methoden der Epidemiologie, die die Entwicklung von Gesundheit und Krankheit im Allgemeinen zum Gegenstand hat und dafür eine Reihe von Studientypen und Auswertungsmethoden entwickelt hat. Die aus der gesteigerten Komplexität erwachsenden Anforderungen haben dabei zur Weiterentwicklung der klassischen Methoden und zur Entwicklung neuer Methoden im Bereich der Studientypen und der mathematischen Modellierung geführt, die es uns heute erlauben, Zusammenhänge und Effekte quantitativ und grafisch sichtbar zu machen, die früher nicht dargestellt werden konnten. Epidemiologie und Versorgungsforschung ergänzen auf diese Weise die Nutzenbewertung um wesentliche Perspektiven, die sich in den nach der Gesetzeslage erfolgten formalen Nutzenbewertungen nicht ausreichend abbilden. Ziel dieses Artikels ist die beispielhafte Darstellung des besonderen Beitrages zur Nutzenbewertung, den Versorgungsforschung und Epidemiologie leisten können, und der dafür nutzbaren modernen Methoden.

### Die frühe Nutzenbewertung von Arzneimitteln und Medizinprodukten durch das IQWiG

Die Vorgehensweise des IQWiG bei der Umsetzung der Gesetzesregelungen zur frühen Nutzenbewertung von Arzneimittel- und Medizinprodukten wird in diesem Heft ausführlich in anderen Beiträgen dargestellt [Zens et al., Kaiser et al., Windeler und Lange] und kann im Detail im Methodenpapier des IQWiG [2] nachgelesen werden. Kurz zusammengefasst folgt das IQWiG dabei zunächst den bereits erwähnten, aus dem Gesetz und den zugehörigen Verordnungen abgeleiteten Grundsätzen:

- Bewertungen sind nur auf der Basis der evidenzbasierten Medizin unter Berücksichtigung der Evidenzstufen vorzunehmen,
- es ist eine zweckmäßige Vergleichstherapie heranzuziehen und
- sich an patientenrelevanten Endpunkten zu orientieren.

Dazu kommen aber weitere Grundsätze, die im Rahmen der Operationalisierung der Gesetzesvorlagen vom IQWiG im geregelten Diskurs mit Wissenschaftlern erarbeitet wurden:

- Die Ursächlichkeit des zu bewertenden Produktes für den ermittelten Nutzen muss klar bewiesen sein. Dafür ist eine (am besten metaanalytische) Zusammenfassung von randomisierten kontrollierten Efficacy-Studie(n) erforderlich. Es werden also nur Studien der Evidenzklasse 1 berücksichtigt.
- Die Bewertung eines Nutzens erfolgt indikationsspezifisch und wird innerhalb einer Indikation nur für genügend homogene Patientengruppen vorgenommen, bei denen ein gleicher Nutzen erwartet werden kann. Dazu müssen Studienpopulationen gegebenenfalls des Öfteren in Unterkollektive verminderten Stichprobenumfangs aufgeteilt werden.
- Die Sicherheit eines positiven Befundes wird in 3 Stufen berichtet (Beleg, Hinweis, Anhaltspunkt).

### Nutzenbewertung aus der Sicht von Versorgungsforschung und Epidemiologie

Vergleichende Nutzenbewertungen folgen auch in der Versorgungsforschung und Epidemiologie den im Gesetz genannten Grundsätzen. Die Evidenzklassen werden bei der Beurteilung der Aussagekraft von Studien berücksichtigt. Es wird die jeweils bestmögliche Evidenz angestrebt. Grundsätzlich werden patientenrelevante Endpunkte betrachtet. Mortalität, Morbidität und Lebensqualität stehen im Vordergrund, aber nicht ausschließlich. Kontrollgruppen werden verwendet, allerdings nicht notwendig nur eine Vergleichstherapie, sondern mehrere Vergleichstherapien, wenn mehrere Optionen zur Verfü-

gung stehen. Größere Unterschiede gibt es hingegen in Bezug auf die vom IQWiG vorgenommenen Konkretisierungen. Hier gehen Versorgungsforschung und Epidemiologie teilweise deutlich andere Wege. Diese Wege ergeben sich aus der Aufgabenstellung und Perspektive von Versorgungsforschung und Epidemiologie und dem erweiterten Gegenstand. Bevor diese Punkte im Detail erläutert werden, werden zunächst einige Beispiele für vergleichende Nutzenbewertungen in der Versorgungsforschung vorgestellt, die die Besonderheiten erkennen lassen.

### Beispiele für die vergleichende Nutzenbewertung in der Versorgungsforschung

Gegenstand der Versorgungsforschung ist die Kranken- und Gesundheitsversorgung. Typische Fragestellungen sind etwa die tatsächliche Verwendung von Medikamenten und medizinischen Hilfsmitteln in der Praxis und die damit verbundenen Risiken, der Nutzen neuer diagnostischer Maßnahmen, ungerechtfertigte Unterschiede in der Versorgung durch Arztpraxen oder Krankenhäuser [4] oder die regionale Verteilung und der Nutzen unterschiedlicher Versorgungsangebote. Die Versorgungsforschung fokussiert damit primär auf die Mikro- und Mesoebene des Gesundheitssystems. Die Epidemiologie hat demgegenüber die gesamte Bevölkerung im Fokus und beschäftigt sich mit der lebenslangen Entwicklung von Risikofaktoren für Krankheiten sowie deren Auftreten, Ausbreitung und Letalität. Typische Fragestellungen sind etwa die regionale Verteilung von spezifischen Krebserkrankungen, die Krebsmortalität und der Zusammenhang zwischen Lebensstil und Krebshäufigkeit. Zu der Bewältigung dieser Aufgabe bedient sich die Epidemiologie eines umfangreichen, stark statistisch geprägten Instrumentariums. Die folgenden Beispiele haben in den letzten Jahren das Interesse einer breiteren Öffentlichkeit gefunden und sind deshalb gut geeignet, die Besonderheiten der Nutzenbewertung in der Versorgungsforschung und der Epidemiologie zu begründen und zu veranschaulichen:

1. Screeninguntersuchungen zur Früherkennung von Tumorerkrankungen

sind spezielle diagnostische Verfahren und werden von großen Hoffnungen im Hinblick auf die Verringerung von Krebsodesfällen und die Erhöhung der Gesamtlebensdauer begleitet. Zur Umsetzung dieser Programme werden erhebliche Investitionen benötigt, ein Finanzprogramm muss aufgestellt werden und Regularien und Verwaltungsabläufe müssen verändert werden. Vor Einführung werden daher im Idealfall große randomisierte Studien mit langer Nachverfolgungszeit begonnen. Politische Entscheidungen werden dann allerdings wegen des großen Interesses an einer Besserung der Heilungsrate häufig bereits auf der Basis noch unvollständiger Daten getroffen. Bei verbreiteten Programmen wie dem Mammografiescreening zur Früherkennung von Brustkrebskrankungen oder dem Screening nach dem prostataspezifischen Antigen (PSA) zur Früherkennung von Prostatakarzinomen ergaben sich bis zum Zeitpunkt der flächendeckenden Einführung unklare Befunde, obwohl bereits mittlerweile Millionen von Teilnehmern in die Studien eingeschlossen waren. Selbst bei Analyse derselben Datenbasis sahen einige Forscher Nutzenbelege, andere keine Belege [5, 6]. Auch Jahrzehnte später führt die Analyse der Langzeitergebnisse zu keinem eindeutigen Bild. Im Ergebnis gibt es in den USA und Europa gegensätzliche Guidelines zum PSA-Test [7, 8]. Das Mammografiescreening ist aktuell wieder in der Diskussion [9, 10]. Wie kann es sein, dass trotz einer überwältigenden Menge von Daten aus randomisierten Studien, epidemiologischen Studien und Registerdaten dennoch keine Einigkeit über den Nutzen der beiden genannten Screeningverfahren zu erzielen ist? Ein wesentlicher Grund liegt in der unterschiedlichen Bewertung von zahlreichen nachgewiesenen oder auch nur plausiblen Biasquellen, denen diese Studien unterliegen. Auch randomisierte Studien sind davon betroffen, verteilt die Randomisierung doch nur die initiale Zuteilung zum Screening oder zur Kontrolle zufällig. Alle anderen

Bundesgesundheitsbl 2015 · 58:298–307 DOI 10.1007/s00103-014-2106-1  
© Die Autor(en) 2014. Dieser Artikel ist auf Springerlink.com mit Open Access verfügbar

K. Wegscheider · A. Drabik · C. Bleich · H. Schulz

## Nutzenbewertung aus Sicht der Versorgungsforschung und der Epidemiologie

### Zusammenfassung

Der vom Gesetzgeber vorgegebene Begriff des (Zusatz-)Nutzens von medizinischen Produkten und Dienstleistungen und das damit zusammenhängende Prüfungsverfahren dienen in erster Linie der Regulierung des Marktzugangs durch Beschränkung auf Medikamente und Produkte, deren Nutzen mit hoher Evidenz belegt ist. Der Begriff beinhaltet implizit die Fiktion eines einheitlichen Nutzens in einer Patientengruppe gleicher Indikation. Aus der Perspektive von Versorgungsforschung und Epidemiologie ist dieser Nutzenbegriff viel zu eng, um die vielfältigen Formen und Abstufungen abzubilden, in denen Patienten bzw. unterschiedlichen Risiken ausgesetzte Bürger aus Versorgungsangeboten wie z. B. Krankenhausbehandlungen, Impfprogrammen oder Screeningverfahren Nutzen ziehen oder Schaden erleiden können. Die Bewertung des Nutzens von Versorgungsangeboten erfordert den Vergleich aller lokal und aktuell erreichbaren Alternativen unter Berücksichtigung der Nutzerpräferenzen und der individuellen Bedingungen. Die Nutzenbewertung in der Versor-

gungsforschung ist dementsprechend nicht eindimensional und erfordert Studien neuer Typs, die sich erheblich von den klassischen Phase-III-Zulassungsstudien unterscheiden. Neue Ansätze entwickeln sich unter den Begriffen „Comparative Effectiveness Research“ (CER) und „Patient Centered Outcome Research (PCOR)“. Auch in der CER und PCOR sind zufällige Zuteilungen von Versorgungsangeboten fester Bestandteil der Studienmethodik, da nur sie den Nachweis der Ursächlichkeit erlauben. Diese Zuteilungen werden jedoch in einem größeren Rahmen betrachtet, der Erkenntnisse aus Registern und Datenbanken, aus epidemiologischen Studien, Machbarkeitsstudien und Post-hoc-Analysen einbezieht. Hierfür bieten CER und PCOR geeignete Designs und moderne statistische Auswertungsverfahren an.

### Schlüsselwörter

Nutzenbewertung · Versorgungsforschung · Epidemiologie · Effectiveness · Comparative Effectiveness Research

## Benefit assessment in health services research and epidemiology

### Abstract

The legal terms “benefit” and “added benefit” and the procedures related to their assessment, given the application of a new medical intervention for approval, should help to restrict the market to those products for which there is much evidence of their benefit or added benefit. The term implies the fiction of an overall benefit for all patients with the same disease. However, from the perspective of health services research and that of epidemiology the term inevitably has to be extended to cover the benefit to a variety of groups of patients or users of a broad spectrum of medical interventions in the real world, ranging from inpatient treatment to vaccination or screening programs. Thus, the assessment of benefit requires a comparison of the new product with all the alternatives currently available for routine care, explicitly taking into account user preferences. Hence, the assessment of benefit in health services research is not one-dimensional and

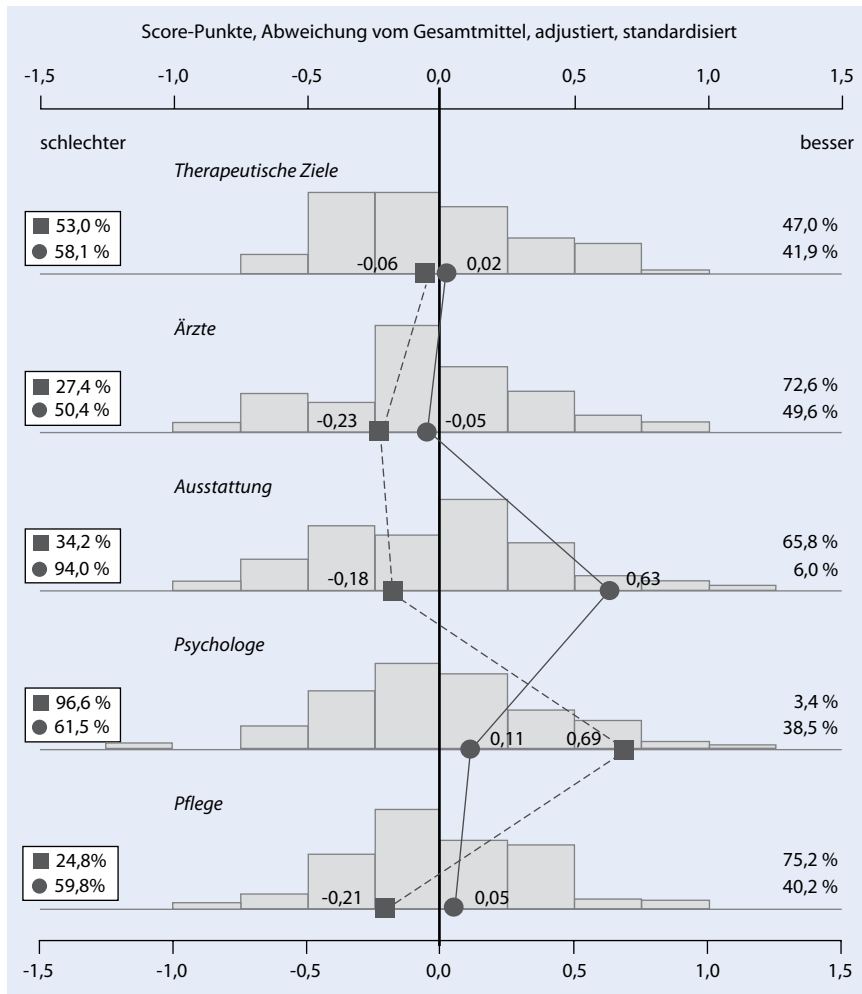
requires new types of studies that go beyond the traditional phase III trials (efficacy trials). New approaches are mainly developed by comparative effectiveness research (CER). CER studies also use randomized designs, because they are currently the best available method for causal inferences. However, randomization in CER is extended to a much broader framework, including health-related databases, registers, epidemiological studies, feasibility studies, and post hoc analyses. CER has developed the necessary and appropriate designs and statistical methods. In addition, some of these methods allow an adaptive assessment of benefit, which can be used to monitor the implementation of new health care interventions or programs.

### Keywords

Benefit assessment · Health services research · Epidemiology · Effectiveness · Comparative effectiveness research

Biasquellen in einem langen Follow-up können prinzipiell auch in randomisierten Studien auftreten. Diese po-

tenziellen Fehlerquellen werden heftig diskutiert. Die Diskussion enthüllt aber darüber hinaus eine weitgehende



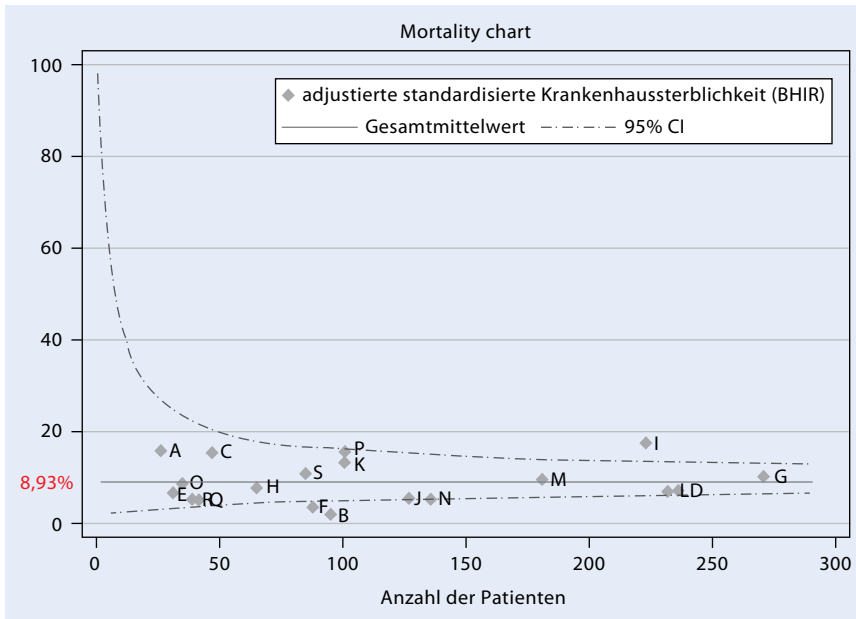
**Abb. 1** ▲ Profilvergleich von zwei für einen Patienten zur Auswahl stehende Rehabilitationskliniken. Gezeigt werden Scorepunkte, berechnet als standardisierte adjustierte Empirical Bayes-Schätzer aus einer von früheren Patienten abgegebenen Zufriedenheitsbewertung in 5 patientenrelevanten Bereichen. Man erkennt, dass eine Klinik (*Kreis* + durchgezogene Linie) in allen Endpunkten durchschnittliche Bewertungen erhält, aber offensichtlich eine herausragende Ausstattung aufweist, während die andere Klinik (*Karos* + gestrichelte Linie) eher unterdurchschnittlich bewertet wird, aber eine als hervorragend wahrgenommene psychologische Betreuung bietet. Im Hintergrund die Verteilung der Scorewerte aller in der Studie berücksichtigten Kliniken, links und rechts der Anteil der Kliniken, die jeweils schlechter oder besser sind. Solche Darstellungen erleichtern Patient und Arzt die Klinikwahl.

Uneinigkeit über die Frage, wie unterschiedliche Nutzenaspekte zu gewichten seien. Soll die verminderte Brustkrebssterblichkeit der Gescreenten oder ihre kaum veränderte Gesamtsterblichkeit, die statistisch unsicher ist, weil die Studien darauf nicht gepowert waren, der Maßstab für einen Nutzen sein? Sollen gut dokumentierte unnötige seelische Belastungen und unnötige diagnostische Eingriffe in die Rechnung einbezogen werden, auch wenn sie sich in der langfristigen Lebensqualitätsmessung nicht nachweisen lassen?

Das Beispiel verdeutlicht, dass trotz Beschränkung auf Mortalität, Morbidität und Lebensqualität die Frage der richtigen Messskala zur Nutzenbewertung offenbleiben kann. Im vorliegenden Fall wird sie je nach persönlicher Präferenz und persönlichem Hintergrund verschieden beantwortet werden. Während wir klare allgemein akzeptierte Regeln zur Bestimmung der Wirksamkeit haben, sind wir beim Nutzen häufig uneins. Dem Begriff des Nutzens haftet ein individuelles Verständnis an, das von den persönlichen Präferen-

zen und der eigenen Positionierung bestimmt ist. Versorgungsforschung mittelt diese unterschiedlichen Sichtweisen nicht weg, sondern berücksichtigt sie bei der Nutzenbewertung und macht sie selbst zum Forschungsgegenstand (Präferenzmessung u. Ä.). Hier ist nicht eine zusammenfassende Bewertung einer Maßnahme gegenüber einer zum Vergleichsstandard erhobenen Kontrolle gefragt, wie sie der Gesetzgeber für die frühe Nutzenbewertung vorsieht. Die Entscheidung für oder gegen die Teilnahme an einem Screening, die eine Frau am besten gemeinsam mit ihrem Arzt trifft, profitiert am meisten von einer Nutzenbewertung, bei der für alle zur Auswahl stehenden Optionen (hier z. B. Zuwarten, regelmäßige Brustabtastung, regelmäßige Echografie, Mammografiescreening alle zwei Jahre bzw. jedes Jahr) der Nutzen in Hinblick auf alle möglicherweise relevanten Endpunkte in einem Endpunktprofil analog zu **Abb. 1** dargestellt wird. Die Frau kann dann die für sie besonders relevanten Endpunkte simultan betrachten und eine ihren Präferenzen entsprechende Wahl treffen.

2. Disease-Management-Programme (DMP) wurden in Deutschland vor mehr als 10 Jahren politisch motiviert für ausgewählte Indikationen flächendeckend eingeführt, ohne die Ergebnisse von Modellversuchen oder die Resultate einer geplanten randomisierten kontrollierten Studie abzuwarten [11]. Sie werden von zahlreichen Kassen auf regionaler Ebene angeboten. Eine Bewertung des Nutzens der DMP unter den Bedingungen der Gesundheitsversorgung in Deutschland ist mangels Kontrollgruppe heute nicht mehr möglich. Betrachtet man Ergebnisse aus dem Ausland, so erhebt sich die Frage, ob DMP tatsächlich im Sinne einer besseren Versorgung wirken oder nur die ohnehin gut motivierten und versorgten Patienten selektieren. Was jedoch weiterhin untersucht werden kann, ist die Frage, inwieweit die zahlreichen DMP der unterschiedlichen Krankenkassen zu gleicher Indikation denselben Nut-



**Abb. 2** ▲ Control Chart zu einem pseudonymisierten Krankenhausvergleich zur Krankenhaussterblichkeit nach akutem Myokardinfarkt, Interimanalyse aus einem laufenden Register. Auf der x-Achse die Zahl der Patienten, die bisher ins Register eingebracht wurden. Auf der y-Achse die für zahlreiche Kovariate adjustierte standardisierte Krankenhaussterblichkeit. Innerhalb des markierten Bereichs können Abweichungen zufällig sein. Krankenhaus B, F und N haben eine niedrigere Sterblichkeit, Krankenhaus I eine höhere als der Durchschnitt. Die Darstellung ist auch zum fortlaufenden Monitoring geeignet.

zen aufweisen. Patienten können zwischen diesen Programmen nicht randomisiert werden. Wegen der Unterschiede in der Zusammensetzung der Teilnahme muss ein statistischer Ausgleich für die unterschiedlichen Anforderungen an die Programme gefunden werden. Dies geschieht durch Adjustierung für zahlreiche Kovariaten, die den Case-Mix beschreiben. Adjustierung allein reicht jedoch nicht aus, um zu unverfälschten Bewertungen zu kommen. Die Nutzenzuschreibungen würden aufgrund des zufälligen Fehlers in der Bewertung der einzelnen DMPs eine zu starke Varianz aufweisen und somit zahlreiche falsch-positive oder falsch-negative Bewertungen produzieren. Besonders häufig sind davon DMPs betroffen, die weniger Patienten in die Evaluation einbringen als die meisten anderen Programme. Einzelne DMP würden im Ergebnis zu Unrecht gelobt oder getadelt werden. Mit dem Übergang zu einem sogenannten hierarchischen Modell, d. h. der Verwendung von bayesianischen Schätzern, kann dieses Problem auf ein Mi-

nimum reduziert werden [12]. Das Bundesversicherungsamt (BVA) hat ein solches mathematisches Modell in die Praxis umgesetzt [13] und ist damit auch ohne Randomisierung zu akzeptierten validen Nutzenbewertungen gekommen, die sich teilweise von den optimistischen Selbstdarstellungen der Krankenkassen erheblich unterscheiden haben [11]. Die Nutzenbewertungen weisen zwar nicht die höchste Evidenzklasse auf. Sie sind dennoch wichtige Marksteine in einem sich ständig verändernden Umfeld, die Hinweise auf Schwachstellen und Verbesserungsmöglichkeiten geben. Und sie waren ausreichend zur Ausübung der Aufsichtsfunktion durch das BVA.

Das Beispiel zeigt, dass vergleichende Nutzenbewertungen in der Versorgungsforschung nicht immer auf randomisiert kontrollierten Studien basieren können. Trotzdem sind sinnvolle Nutzenbewertungen möglich, die z. B. zur Qualitätssicherung eingesetzt werden können.

3. Ein ähnlich gelagertes Problem findet sich im Bereich der Krankenhausver-

gleiche, die in Deutschland seit Jahren intern von Trägern und Management, extern und flächendeckend, für ausgewählte Indikationen vom BQS-Institut für Qualität und Patientensicherheit und dem Institut für Angewandte Qualitätsförderung und Forschung im Gesundheitswesen (ACQUA), öffentlich lokal von zahlreichen Zeitungen sowie auf freiwilliger Basis von Fachgesellschaften, Netzwerken und Registern vorgenommen werden. Diese Vergleiche werden an Brisanz und Bedeutung gewinnen, wenn die Vergütung mit der Ergebnisqualität in Verbindung gebracht werden sollte, was politisch zurzeit intensiv diskutiert wird [14]. Ein solches System nimmt, ohne dass das Wort dabei bisher zur Anwendung kommen muss, eine ständige vergleichende Nutzenbewertung der Krankenhäuser für den Durchschnittspatienten vor. Das ideale statistische Auswertungsmodell eines solchen Vergleiches ist analog zum oben im Zusammenhang mit den DMP beschriebenen Modell (Anwendungsbeispiele zeigen **Abb. 1 und 2**). Mit Hilfe dieses Modells und einer Zugriffsmöglichkeit zur Datenbank ließen sich auch relativ problemlos vergleichende Nutzenbewertungen für spezielle Patientengruppen durchführen. In Deutschland scheitern solche Bemühungen bis auf Weiteres am Datenschutz und an der weitverbreiteten Furcht von Entscheidungsträgern, Politikern und anderen in der Öffentlichkeit stehenden Personen vor Transparenz. Faire und valide Vergleiche zwischen Einrichtungen findet man deshalb am ehesten in Registern mit freiwilliger Teilnahme und in Qualitätszirkeln. Das Beispiel zeigt, wie Routinedaten für Nutzenbewertungen verwendet werden können, wenn die Datensammlung von vornherein so angelegt ist, dass die erforderlichen Daten an der wissenschaftlichen Zielsetzung orientiert in hinreichender Qualität gesammelt werden. Das setzt allerdings die allgemeine Akzeptanz der Evaluation durch die beteiligten Institutionen voraus.

## Erweiterung des Nutzenbegriffs in der Versorgungsforschung

### Übertragbarkeit von Studienergebnissen

Ein häufig gegen die Begrenzung auf randomisierte Studien ins Feld geführtes Argument ist das der mangelnden **Übertragbarkeit** auf die Behandlung im klinischen Alltag. Von den alltäglich mit einer bestimmten Therapie behandelten Patienten erfüllt zumeist nur ein sehr kleiner Teil die Einschlusskriterien der Zulassungsstudien und weiterer randomisierter Studien. Viele spätere Nutzergruppen wie z. B. komorbide bzw. multimorbide, alte oder gebrechliche Patienten, Schwangere oder stillende Frauen, Kinder, Behinderte oder Menschen mit Migrationshintergrund werden aufgrund der restriktiven Ein- und Ausschlusskriterien nicht berücksichtigt. Windeler und Lange setzen sich in ihrem Artikel in diesem Heft ausführlich mit diesem Argument auseinander. Zutreffend stellen sie fest, dass die mangelnde Übertragbarkeit nur dann gegeben wäre, wenn in diesen Nutzergruppen tatsächlich eine Effektmodifikation zu beobachten wäre. Die Beweislast sehen sie bei den Kritikern.

Für Wirksamkeitsstudien mag die Argumentation von Windeler und Lange zutreffen. Bei der Nutzenbewertung in der Versorgungsforschung liegen die Dinge anders. Hier können Einflussgrößen auch ohne Effektmodifikation eine dominante Bedeutung gewinnen. Eine wirksame Dauertherapie für eine chronische Krankheit, die in der Stadt mit einem dichten Versorgungsnetz angeboten einen hohen Nutzen entfaltet, kann auf dem Land mit einem Fahrweg von mehreren Stunden von geringerem Nutzen sein als eine direkt verfügbare, aber weniger wirksame Therapie. Der Nutzen von Therapien hängt auch dann nicht unwesentlich vom lokalen Versorgungsangebot, von den Lebensumständen des Kranken, seinen Komorbiditäten und weiteren Einflussfaktoren ab, wenn die Wirksamkeit überall gleich ist. Eine bedeutsame Einflussgröße ist dabei häufig das Alter. Selbst wenn Sensitivität und Spezifität von Screeningverfahren altersunabhängig sind, wird die Nutzenbewertung es nicht sein, wie das Bei-

spiel des Mammografiescreenings zeigt. Bei weitgehend gleicher Wahrscheinlichkeit von Fehldiagnosen ist der mögliche Gewinn junger Frauen durch das Screening gering, da ihr Brustkrebsrisiko kleiner ist, sodass das Screening für sie nutzlos oder sogar schädlich ist. Ebenso ist der Nutzen für ältere Frauen gering, weil bei langsamer wachsenden Karzinomen und weniger zu gewinnenden Lebensjahren die möglichen Schädigungen durch die Diagnostik überwiegen. Bei der Nutzenbewertung in der Versorgungsforschung kann man sich deshalb nicht auf die randomisiert kontrollierten Vergleiche in den großen Studien beschränken, sondern muss zumeist mehrere Einflussgrößen simultan ins Kalkül ziehen. Hierfür verfügt die Epidemiologie über ein umfangreiches statistisches Instrumentarium. In einem solchen Modell würde zum Beispiel die Abhängigkeit des Nutzens vom Alter mithilfe einer Funktion beschrieben. Bei richtiger Wahl des Funktionstyps und der Parameterschätzung aus einem genügend großen Kollektiv, das auch aus einer randomisierten Studie stammen könnte, lässt sich die Abhängigkeit des Nutzens vom Alter sehr genau und belastbar quantifizieren. Hier zeigt sich der Unterschied zu der Methode des IQWiG besonders deutlich: Das IQWiG würde in einem solchen Fall das Kollektiv so lange in kleine Unterkollektive einteilen, bis die Inhomogenität zwischen Alten und Jungen innerhalb des Unterkollektivs nicht mehr sichtbar ist und damit das Unterkollektiv als genügend homogen klassifiziert werden kann. Dann wären aber die Unterkollektive so klein geworden, dass Anhaltspunkte, Hinweise und Belege höchstens zufällig zustande kämen. Die Altersabhängigkeit des Nutzens wäre so nicht sichtbar geworden. Eine adäquate statistische Modellierung ist ein mächtiges Instrument, das Zusammenhänge sichtbar machen kann, die mit Ja/Nein-Entscheidungen nicht adäquat adressiert werden können.

### Randomisation und epidemiologische Modellbildung

Gegen die Verwendung epidemiologischer Modelle in der Nutzenbewertung wird häufig vorgebracht, dass sie anders als randomisierte Studien zum Beweis der

Ursächlichkeit nicht in der Lage sind, weil die Zuordnung von Patienten zu Therapien zur ungleichmäßigen Verteilung anderer relevanter Einflussgrößen geführt haben kann (sog. Allocation Bias). Dieses Argument ist richtig und wird auch von der Epidemiologie sehr ernst genommen. Die Epidemiologie verfügt zwar über spezielle Methoden (z. B. Adjustierung, Stratifizierung, Propensity Score Matching), die dazu geeignet sind, den Allocation Bias zu reduzieren. Sicher ausschließen lässt er sich auch dadurch nicht, da der Bias von unbeobachtbaren und somit nicht adjustierbaren Einflüssen verursacht werden kann. Am Ende jeder epidemiologischen Studie muss deshalb sehr sorgfältig die Frage beantwortet werden, welche Größenordnung dieser Bias annehmen kann und ob das gegenüber anderen Einflüssen, die das Studienergebnis in randomisierten wie in epidemiologischen Studien verfälschen können (wie z. B. fehlende Werte), überhaupt ins Gewicht fällt.

Nichtsdestotrotz würden auch Versorgungsforscher bei Nutzenbewertungen die Randomisierung als Studienelement wählen, wenn die Möglichkeit dazu gegeben ist. Wie das im Rahmen von Nutzenstudien erfolgen kann, wird im nächsten Kapitel erörtert. Generell ersetzt die statistische Modellbildung nicht die Randomisation, sondern ergänzt sie. Manchmal ist sie wichtiger für eine Nutzenbewertung als die Randomisation, wie die oben vorgestellten DMP-Bewertungen und Krankenhausvergleiche zeigen.

### Nutzen ist individuell

Die Wirksamkeit einer Therapie mag kollektiv sein, der Nutzen ist primär individuell. Aufgabe des behandelnden Arztes ist es, unter Berücksichtigung der Evidenzlage, meist vermittelt über Leitlinien, zusammen mit seinem Patienten nicht die Therapie maximaler Wirksamkeit, sondern die Therapie mit dem größten individuellen Nutzen zu finden. In seine Überlegungen gehen neben vielen anderen Aspekten Kontraindikationen, Unverträglichkeiten, Komorbiditäten, Krankheitsphase, erwartete Therapiedauer, die Lebensumstände des Patienten, die Motivation und Befähigung des Patienten und dessen Präferenzen ein.

Für diesen Prozess ist ein großes Spektrum von therapeutischen Optionen vorteilhaft. Behandlungen mit identischer Wirkung (z. B. „Me-too“-Präparate) bereichern die therapeutische Vielfalt möglicherweise nicht. Therapien mit gleichem kollektivem Durchschnittsnutzen können dagegen durchaus für einzelne Patienten von besonderer Bedeutung sein. Ebenso hängt der Nutzen von z. B. Screeningverfahren, DMP und Krankenhausaufenthalten stark von der individuellen Situation wie z. B. von persönlichem Risiko, persönlicher Einstellung und psychischer Verfassung (Screening), von der Routineversorgungssituation im eigenen Umfeld und den beteiligten Behandlern, vom persönlichen Unterstützersystem des Patienten und der eigenen Motivationslage (DMP) sowie von den Begleitumständen eines Krankenhausaufenthalts für Patienten und Angehörige ab. Diese Zusammenhänge sind Gegenstand der Versorgungsforschung. Und es sind diese Zusammenhänge, die Patienten und Therapeuten neben dem therapeutischen Erfolg besonders interessieren. Die sehr nah am potenziellen Interesse der Leser arbeitenden Zeitungsjournalisten haben das erkannt und richten ihre Krankenhausrankings entsprechend aus, leider im Allgemeinen mit Rückgriff auf willkürlich gesammelte Daten und vor allem ohne zureichende Methodik. Demgegenüber ist wissenschaftlich fundierte Versorgungsforschung in der Lage, mit eigenen systematisch erhobenen Studiendaten und solider Methodik stärker belastbare hilfreiche Aussagen zur Einrichtungs- und Therapiewahl zu machen. Für die Nutzenbewertung in der Versorgungsforschung bedeutet das ganz allgemein, dass für jede Therapie, jedes Programm und jede Einrichtung eine Charakterisierung der Patienten- oder Nutzergruppe erfolgen sollte, welche daraus einen Nutzen ziehen könnte. Diese Gruppe kann bei nutzlosen Therapien natürlich auch sehr klein oder nicht vorhanden sein.

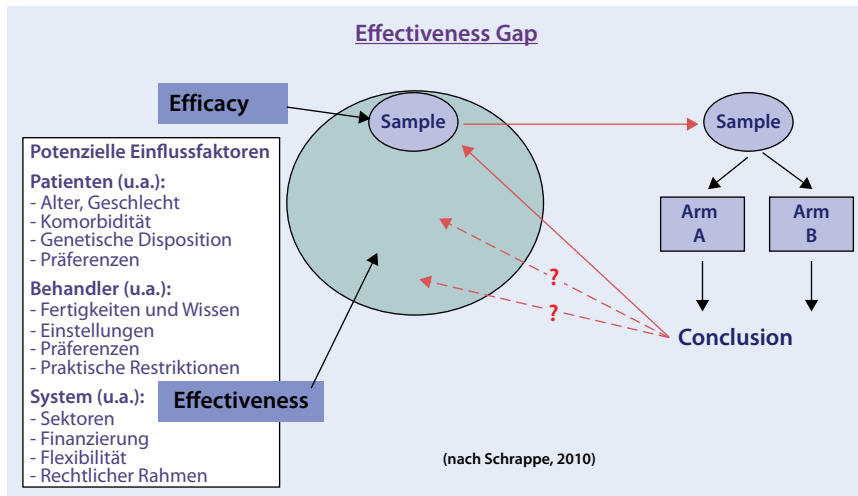
### Nutzen ist vorläufig

Eine weitere wichtige Determinante des Nutzens, die ins Kalkül gezogen werden sollte, ist die **Zeit**. Das gilt zunächst für den Nutzer von Angeboten bzw. den

Patienten, dessen Situation sich im Laufe des Lebens verändert. Lebensumstände wechseln, Krankheiten durchlaufen Phasen und hinterlassen Spuren. Was heute dem Patienten Nutzen bringt, kann morgen nutzlos geworden sein. Von den Screeningverfahren etwa können junge Menschen mit niedriger Krebsinzidenz und alte Menschen mit großen konkurrierenden Risiken nicht profitieren. In gleichem Maße wie die Nutzer verändern sich aber auch Programme und Einrichtungen selbst, etwa bei Personalwechsel, wirtschaftlicher Schräglage oder Missmanagement. Und es ändert sich das Umfeld, in dem Nutzer auf Versorgungsangebote treffen. Demografische Veränderungen haben Rückwirkungen auf die Versorgungslandschaft und verändern damit z. B. den Nutzen der Telemedizin. Wissenschaft und Forschung bringen Innovationen hervor, die Märkte verändern. Bessere Therapien erhöhen zunächst und reduzieren dann den Nutzen von Screeningmaßnahmen. Über einen längeren Zeitraum betrachtet haben Therapien und Versorgungsmodelle eine begrenzte Lebensdauer. Der Nutzenbewertung haftet damit häufig auch etwas prinzipiell Vorläufiges an. Ergebnisse müssen deshalb zeitnah zur Verfügung stehen und sich flexibel Veränderungen anpassen. Dies gilt umso mehr, je komplexer und vielgliedriger eine zu bewertende Intervention ist.

Die Einrichtung eines Versorgungsprogramms (z. B. DMP, Impfprogramme) ist eine solche **komplexe** Intervention [15]. Sie verlangt Investitionen, Integration in Bestehendes, Umorganisationen, Ablaufveränderungen sowie Zeitplanung, und dieses gelingt nur, wenn genügend Zeit und Raum für Erprobungen, Verbesserungen, weitere Innovationen, abweichende Varianten und lokale Rücksichtnahmen ist. Sie findet im politischen Raum und damit in einer Atmosphäre ständigen Agierens und ständigen Wandels statt. Für die Versorgungsforschung stellt diese Situation eine schwierige Herausforderung dar. Empirische Studien brauchen Zeiten relativer Konstanz der Bedingungen, konstante Fragestellungen, Randbedingungen, Einschlusskriterien, definierte Endpunkte u. Ä. In Bereichen mit hoher politischer Anteilnah-

me oder hoher Innovationsgeschwindigkeit ist unter diesen Umständen Evidenzfindung und Nutzenbewertung erschwert. Dabei ist gerade in dieser Situation eine empirische Erdung der Aktivitäten besonders wichtig. Herkömmliche Modelle der Evidenzfindung und Nutzenbewertung wie Phase-III-Studien und Metaanalysen von Studien mit höchster Qualitätsstufe, die nur auf einen Endpunkt schauen, sind hierfür zu eingeschränkt im Fokus und zu langsam im Ablauf. Sie nutzen die in den Studien angesammelte Information zu ängstlich und zu vorsichtig, sie beschränken sich zu oft auf Gesichertes, wo das Wahrscheinliche vorläufig ausreichen würde. Mit dem wachsenden Bedarf nach empirisch begründeten Entscheidungshilfen hat sich eine neue Forschungsrichtung unter der Überschrift „Comparative Effectiveness Research (CER)“ gebildet, die sich speziell der vergleichenden Nutzenbewertung unter den geschilderten Umständen zuwendet und sich mit der (Weiter-)Entwicklung von Methoden auseinandersetzt, die den besonderen Anforderungen gerecht werden [16, 17]. **Abb. 3** zeigt programmatisch die Aufgabenstellung: Auf der Basis eines in klassischer Form erbrachten Wirksamkeitsnachweises (efficacy) werden Nutzenstudien mit unterschiedlichem Design aufgebaut, die die praktische Bewährung der Intervention unter unterschiedlichen Alltagsbedingungen zum Ziel haben (effectiveness), um damit die von vielen empfundene Lücke zwischen dem Wirksamkeitsnachweis in Zulassungsstudien und dem praktischen Nutzen in der klinischen Routine (efficacy-effectiveness gap, s. **Abb. 3**) zu schließen. CER versteht sich dabei als Informant und Unterstützer von Entscheidungsträgern, die zu bestimmten Zeitpunkten politische Entscheidungen treffen müssen und nicht auf das Endergebnis großer Studien warten können. Sie müssen zum aktuellen Zeitpunkt die bestmögliche Evidenz verfügbar haben, die stärker vorläufig ist als unangenehm sein dürfte und nur teilweise durch Studien der höchsten Evidenzklasse gestützt wird. CER ist zwar nicht beschränkt auf Versorgungsforschungsfragestellungen, kommt aber den Anforderungen an die Nutzenbewertung in der Versorgungsforschung sehr entgegen.



**Abb. 3** ▲ Efficacy-effectiveness gap. In großen Wirksamkeitsstudien wird die efficacy für eine speziell ausgewählte Untergruppe von Patienten untersucht. Ob die Ergebnisse auf den Nutzen (effectiveness) übertragen werden können, hängt von vielen Einflussfaktoren ab (linker Kasten). Zur Erkundung der effectiveness müssen Studien an anderen Qualitätskriterien ausgerichtet werden als Wirksamkeitsstudien (Comparative Effectiveness Research) [18].

## Designs und Auswertungsmethoden für die erweiterte Nutzenbewertung

Die Zielsetzung und Methodik der CER ist in der Literatur ausführlich beschrieben [16, 17, 19]. Kurz gefasst wird im Unterschied zu den auf Wirksamkeitsnachweise zielenden Phase-III-Studien auf eine hochgradige Standardisierung und Qualifizierung der untersuchten Verfahren zu Studienzwecken verzichtet, um ein möglichst zutreffendes Bild der Effectiveness zu erhalten. Es werden Vergleiche zwischen zwei oder mehreren aktiven Versorgungsangeboten durchgeführt, nicht ein einzelner Vergleich mit einer inaktiven Kontrolle. Es werden mehrere patientennahe Endpunkte simultan betrachtet, heterogene Kollektive ausdrücklich zugelassen. CER favorisiert eine Vielfalt von Studienansätzen und problemangepassten Designs zur vergleichenden Evaluation von Interventionsergebnissen. Mittlerweile wird Patient Centered Outcome Research (PCOR) als eine wichtige Erweiterung der CER verstanden. In PCOR wird die Patientenzentrierung explizit einbezogen, sowohl hinsichtlich Präferenzen und Entscheidungshilfen als auch bei der Berücksichtigung von Patientencharakteristika [17].

Einige methodische Ansätze seien hier aufgeführt:

1. **Head-to-head-Studien** sind individuell randomisierte Studien zwischen konkurrierenden Versorgungsangeboten. Nach dem Intention-to-treat-Ansatz ausgewertet gelten sie bei hinreichender Vollständigkeit der Daten als Königsweg zur Evidenzfindung beim Nutzenvergleich [20]. Sie setzen allerdings voraus, dass für jeden Teilnehmer die zu vergleichenden Angebote auch tatsächlich zur Verfügung stehen. Sie sind zudem eventuell schwer zu realisieren, weil sie die Kooperation konkurrierender Hersteller oder Anbieter verlangen. Die Finanzierung muss häufig aus öffentlichen Quellen erfolgen. Es gibt entsprechend zu wenige Studien dieses Typs.
2. Versorgungsprogramme sind häufig Interventionen, die jeweils ein begrenztes Gebiet umfassen, z. B. einen Kreis, eine Gemeinde oder eine Arztpraxis (sog. Cluster). Der einzelne Teilnehmer hat dann keine Wahlmöglichkeit und kann nicht individuell randomisiert werden. Hier können **clusterrandomisierte Studien** weiterführen, bei denen die Einführung von Versorgungsprogrammen zufällig einzelnen Clustern zugeteilt wird [21, 22]. In diesen Studien erfolgt die Intervention häufig auf Clusterebene, die Ergebnismessung aber auf Patientenebene. Die Auswertung dieser
3. Alternativ zur Clusterrandomisierung bieten sich bei verzögerter schrittweiser Einführung von Programmen keilförmige **Stepped-Wedge-Designs** an, bei denen die Reihenfolge der Einführung einer Intervention zwischen verschiedenen Zentren oder Regionen randomisiert werden kann [23, 24]. In diesem Design können saisonale Effekte und andere Kalendereffekte kontrolliert werden, sodass neben dem Vergleich der Zentren zu festen Zeitpunkten für jedes Zentrum bzw. jede Region ein für die Zeiteffekte adjustierter Vergleich der Perioden vor und nach Programmeinführung möglich ist.
4. Die Begleitung eines Entwicklungs- und Anpassungsprozesses bzw. ein fortgesetztes Monitoring lässt sich mit **adaptiven Designs** oder **bayesianischen Methoden** durchführen. Sie ermöglichen Modellierungen bei komplexen Unternehmungen der oben beschriebenen Art. Bayesianische Methoden verwenden das vorhandene Vorwissen zur Stabilisierung von Schätzverfahren und sind deshalb für Vergleiche mit begrenzten Stichprobenumfängen oder vorläufige Aussagen bei laufenden Studien besonders gut geeignet. Sie beruhen auf der Erkenntnis, dass ein kleiner Bias häufig zu kleineren Fehlern führt als ein großer zufälliger Fehler. Die Methoden sind mathematisch anspruchsvoll und in der Versorgungsforschung bisher wenig eingesetzt [25].
5. Gut geführte prospektiv geplante **Register** und **Datenbanken** sowie geplante **Kohortenstudien** erlauben die Modellierung von zeitabhängigen Prozessen unter Verwendung unterschiedlicher epidemiologischer Designs (z. B. nested case control studies, case cohort studies) auch bei nachträglich aufgetretenen Fragestellungen (siehe auch [26, 27]). Zur Aus-

Studien wird bevorzugt mit Mehrebenenmodellen durchgeführt. Das setzt allerdings auf Forscherseite die Auseinandersetzung mit neuer Methodik voraus. In der Praxis stellt zudem die Motivierung der Kontrollgruppe zur Studienteilnahme eine besondere Herausforderung dar.



wertung sollten ebenfalls Mehrebenenmodelle mit festen und zufälligen Effekten bzw. Überlebenszeitmodelle verwendet werden. Die Darstellung der Ergebnisse, die auch prozessbegleitend möglich ist, kann in Form von Control Charts erfolgen. Leider gibt es die hierfür erforderlichen Datenquellen in Deutschland in deutlich geringerer Zahl als in vielen anderen Ländern.

6. Randomisierte Studien, Kohortenstudien und Register werden immer häufiger im **Verbund** durchgeführt. So können Register als Rekrutierungshilfe für klinische Studien verwendet werden. Diese sogenannten „nested registries“ erlauben zudem eine Einschätzung, wie repräsentativ die Patientenauswahl der randomisierten Studie ist, und einen Vergleich der Langzeitverläufe von Studienpatienten im Vergleich zu Patienten, die nicht an der Studie teilgenommen haben. Bei Interventionen, bei denen viele Patienten klare Präferenzen aufweisen (z. B. Akupunktur, Naturheilverfahren), kann man im Verbund von Kohortenstudie und randomisierter Studie den Patienten Wahlfreiheit lassen, ob sie sich offen für eine Option entscheiden oder sich randomisieren lassen wollen. Diese Verbundstudien lassen eine Beurteilung zu, welchen Anteil die Erwartungshaltung aufgrund der eigenen Präferenz an der Wirkung einer Therapie hat.
7. Bei nachhaltigen Schwierigkeiten, konkurrierende therapeutische Vorgehensweisen in einer Studie zu vereinen, die einen Direktvergleich ermöglicht, bietet es sich an, eine Erweiterung der klassischen Metaanalyse auf mehr als zwei zu vergleichen: die sog. **Netzwerk-Metaanalysen**, die indirekte Vergleiche zwischen Armen aus verschiedenen Studien ermöglichen.

## Fazit

**Nutzenbewertungen von Versorgungsangeboten werden entsprechend gesetzlicher Vorgaben im Rahmen standar-**

**disierter Verfahren in Deutschland vom IQWiG auf der einen Seite, aber auch von Versorgungsforschung und Epidemiologie auf der anderen Seite vorgenommen. G-BA und IQWiG bewegt bei der frühen Nutzenbewertung gemäß ihrem Auftrag in erster Linie die Sorge, dass insuffiziente Innovationen Marktzugang erhalten und damit die Solidargemeinschaft finanziell belasten könnten. Die Versorgungsforschung bewegt hauptsächlich die Sorge, dass die Anpassung der Versorgung an den besonderen Bedarf der Patienten nicht gelingen könnte. Wie in den gesetzlichen Regeln und Verfahrensanweisungen festgelegt, beschränkt sich die Nutzenbewertung des IQWiG auf Quellen mit hohem Evidenzgrad und pflegt ein striktes Regelwerk, das unzureichenden Versorgungsangeboten den Weg versperren soll. Versorgungsforschung und Epidemiologie wertschätzen die Vielfalt der Quellen und der Therapieangebote, ohne dabei Qualitätsunterschiede zu ignorieren. Bei seriöser Arbeit nutzen sie ihr reiches Methodenspektrum, um in einer komplexen Welt belastbare Aussagen machen zu können; indem sie den Nutzen von Versorgungsangeboten unter verschiedenen Umständen und Bedingungen vergleichen, geben sie Hinweise auf Verbesserungsmöglichkeiten und zur individuellen Therapie- oder Therapeutenwahl. G-BA und IQWiG auf der einen und Versorgungsforschung und Epidemiologie auf der anderen Seite weisen somit zwar verschiedene Aufgabenstellungen und deshalb verschiedene Prioritäten bei der Zielsetzung und Herangehensweisen bei der Methodik auf, haben aber ähnliche Qualitätskriterien und fühlen sich auch den Zielen des jeweils anderen verpflichtet.**

## Korrespondenzadresse

**Prof. Dr. K. Wegscheider**  
 Institut für Medizinische Biometrie und Epidemiologie  
 Universitätsklinikum Hamburg-Eppendorf  
 Martinistraße 52 (Gebäude W34)  
 20246 Hamburg  
 k.wegscheider@uke.uni-hamburg.de

## Einhaltung ethischer Richtlinien

**Interessenkonflikt.** K. Wegscheider ist Mitglied der Arzneimittelkommission der deutschen Ärzteschaft (AkdÄ) und des Kompetenznetzes Vorhofflimmern (AF-Net). Er gibt an, dass seine Institution neben Drittmitteln für öffentlich geförderte Studien auch Drittmittel für die biometrische Betreuung und statistische Auswertung von der Industrie kofinanzierter investigator initiated trials oder von der Industrie gesponsorter Studien zu bereits auf dem Markt befindlichen Hilfsmitteln erhalten hat oder erhält. Beteiligt sind die Firmen Biotronik, Medtronic und Resmed. Unabhängig davon erhält K. Wegscheider persönlich Honorare für statistische Fortbildung (Boston Consulting, Resmed).

A. Drabik, C. Bleich und H. Schulz geben an, dass kein Interessenkonflikt besteht.

Dieser Beitrag beinhaltet keine Studien an Menschen oder Tieren.

**Open Access.** Dieser Artikel unterliegt den Bedingungen der Creative Commons Attribution License. Dadurch sind die Nutzung, Verteilung und Reproduktion erlaubt, sofern der/die Originalautor/en und die Quelle angegeben sind.

## Literatur

1. Gemeinsamer Bundesausschuss (2014) Verfahrensordnung des Gemeinsamen Bundesausschusses (in der Fassung vom 18. Dezember 2008, zuletzt geändert am 19. Juni 2014). [https://www.g-ba.de/downloads/62-492-922/VerFO\\_2014-06-19.pdf](https://www.g-ba.de/downloads/62-492-922/VerFO_2014-06-19.pdf). Zugegriffen: 26. Nov. 2014
2. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (2014) Allgemeine Methoden – Version 4.2 vom 18.06.2014. [https://www.iqwig.de/download/IQWiG\\_Methoden\\_Entwurf-fuer-Version-4-2.pdf](https://www.iqwig.de/download/IQWiG_Methoden_Entwurf-fuer-Version-4-2.pdf). Zugegriffen: 26. Sept. 2014
3. Pfaff H, Abholz H, Glaeske G et al (2011) Versorgungsforschung: unverzichtbar bei Allokationsentscheidungen – eine Stellungnahme. Dtsch Med Wochenschr 136(48):2496–2500
4. Wikipedia, die freie Enzyklopädie (2014) Seite „Versorgungsforschung“ – Bearbeitungsstand: 19.06.2014. <http://de.wikipedia.org/wiki/Versorgungsforschung>. Zugegriffen: 26. Sept. 2014
5. Goodman SN (2002) The mammography dilemma: a crisis for evidence-based medicine? Ann Intern Med 137(5 Pt 1):363–365
6. Olsen O, Gøtzsche PC (2001) Cochrane review on screening for breast cancer with mammography. Lancet 358(9290):1340–1342
7. Moyer VA, U.S. Preventive Services Task Force (2012) Screening for prostate cancer: U.S. Preventive Services Task Force recommendation statement. Ann Intern Med 157(2):120–134
8. Mottet N, Bastian PJ, Bellmunt J et al (2014) Guidelines on prostate cancer – European Association of Urology 2014. [http://www.uroweb.org/gls/pdf/1607%20Prostate%20Cancer\\_LRV3.pdf](http://www.uroweb.org/gls/pdf/1607%20Prostate%20Cancer_LRV3.pdf). Zugegriffen: 26. Sept. 2014
9. Swiss Medical Board Fachgremium (2013) Systematisches Mammographie-Screening – Bericht vom 15.12.2013. [http://www.medical-board.ch/fileadmin/docs/public/mb/Fachberichte/2013-12-15\\_Bericht\\_Mammographie\\_Final\\_rev.pdf](http://www.medical-board.ch/fileadmin/docs/public/mb/Fachberichte/2013-12-15_Bericht_Mammographie_Final_rev.pdf). Zugegriffen: 26. Sept. 2014

10. Biller-Andorno N, Jüni P (2014) Abolishing mammography screening programs? A view from the Swiss Medical Board. *N Engl J Med* 370(21):1965–1967
11. Wegscheider K, Drabik A, Kaduszkiewicz H, Schäfer I, van den Bussche H (2014) Disease-Management-Programme – was wurde bisher erreicht? In: Jonitz G et al (Hrsg) Ergebnisverbesserung durch Qualitätsmanagement. Report Versorgungsforschung, Bd 8. Deutscher Ärzte-Verlag, Köln, S 145–154
12. Wegscheider K (2004) Methodische Anforderungen an Einrichtungvergleiche („Profiling“) im Gesundheitswesen. *Z Ärztl Fortbild Qualitätssich* 98(8):647–654
13. Wegscheider K, Romes C, Gawlik C (2006) Evaluation von Disease Management Programmen in Deutschland. In Göppfarth D, Greß S, Jacobs K, Wasmann J (Hrsg) Risikostrukturenausgleich 2006 Zehn Jahre Kassenwahlfreiheit. Asgard, St. Augustin
14. Flintrop J, Osterloh F (2013) Krankenhäuser: Qualität ganz oben auf der Agenda. *Dtsch Arztebl* 110(59):A-2406/B-2120/C-2044
15. Medical Research Council (2008) Developing and evaluating complex interventions: new guidance. <http://www.mrc.ac.uk/documents/pdf/complex-interventions-guidance/>. Zugriffen: 29. Sept. 2014
16. Tunis SR, Benner J, McClellan M (2010) Comparative effectiveness research: policy context, methods development and research infrastructure. *Stat Med* 29(19):1963–1976
17. Methodology Committee of the Patient-Centered Outcomes Research Institute (2012) Methodological standards and patient-centeredness in comparative effectiveness research: the PCORI perspective. *JAMA* 307(15):1636–1640
18. Schrappe M (2010) Nutzendimensionen II. Vortrag auf dem Symposium „Onkologie im Brennpunkt“ der Deutschen Krebsgesellschaft. Berlin, 27.10.2010. [http://www.schrappe.com/ms2/index\\_htm\\_files/dkrebsg10.pdf](http://www.schrappe.com/ms2/index_htm_files/dkrebsg10.pdf). Zugriffen: 29. Sept. 2014
19. Peduzzi P, Kyriakides T, O'Connor TZ, Guarino P, Warren SR, Huang GD (2010) Methodological issues in comparative effectiveness research: clinical trials. *Am J Med* 123:e8–e15
20. Lehmacher W, Wolff S (2011) Die Bedeutung von Head-to-Head-Studien für die versorgungsnahe klinische Forschung. *Z Evid Fortbild Qual Gesundhewes* 105(9):639–645
21. Sönnichsen AC, Winkler H, Flamm M (2010) The effectiveness of the Austrian disease management programme for type 2 diabetes: a cluster-randomised controlled trial. *BMC Fam Pract* 11:86
22. Lando HA, McGovern PG, Barrios FX, Etringer BD (1990) Comparative evaluation of American Cancer Society and American Lung Association smoking cessation clinics. *Am J Public Health* 80(5):554–559
23. Mdege ND, Man MS, Tylor Nee Brown CA, Torgerson DJ (2011) Systematic review of stepped wedge cluster randomized trials shows that design is particularly used to evaluate interventions during routine implementation. *J Clin Epidemiol* 64(9):936–948
24. Brown CA, Lilford RJ (2006) The stepped wedge trial design: a systematic review. *BMC Med Res Methodol* 6:54
25. Zhou X, Liu S, Kim ES, Herbs RS, Lee JJ (2008) Bayesian adaptive design for targeted therapy development in lung cancer – a step forward personalized medicine. *Clin Trials* 5(3):181–193
26. Neugebauer AM, Icks A, Schrappe M (2010) Memorandum III: Methoden für die Versorgungsforschung (Teil 2). *Gesundheitswesen* 72(10):739–774
27. Lauer MS, D'Agostino (2013) The randomized registry trial – the next disruptive technology in clinical research? *N Engl J Med* 369:1579–1581