

Interview mit Prof. Dr Axel Mühlbacher, dem Leiter der Pilotstudie zur „Conjoint Analyse“ im Auftrag des IQWiG

„Suche nach dem Maß des Gesamtnutzens“

Knapp zwei Jahre arbeitete das IQWiG gemeinsam mit nationalen und internationalen Experten an einer Methode für die Bewertung von Kosten-Nutzen-Verhältnissen. Ein wesentlicher Bestandteil dieser Methode ist die Analyse der Effizienzgrenze, die - so der damalige Institutsleiter Prof. Dr. Peter T. Sawicki - „die für die deutschen Rahmenbedingungen am besten geeignete Methode“ sei. Doch genau die stieß bei der Vorstellung des Methodenpapiers im Oktober 2009 auf harsche Kritik: 29 führende deutsche GesundheitsökonomInnen lehnten damals den Vorschlag des IQWiG als „wissenschaftlich unhaltbar“ und „normativ nicht begründet“ ab. „Nirgendwo in dem Methodenpapier ist ein Algorithmus beschrieben, wie ein Summenparameter berechnet wird, durch den die Gewichtung von Nutzen und Schaden vorgenommen wird“, kritisierte damals Prof. Dr. med. Jürgen Fritze, leitender Verbandsarzt beim PKV-Verband. Dieses Manko ging das Institut mit zwei Generalaufträgen in Form von Pilotstudien an, mit einer „Conjoint Analyse“ (CA) in der Indikation Hepatitis C sowie einem „Analytic Hierarchy Process“ (AHP) in der Indikation Major Depression - leider in zwei unterschiedlichen Indikationsgebieten, so dass der direkte Vergleich nicht möglich ist. Das Ziel, das nun dennoch zum Greifen nahe scheint: Verfahren wissenschaftlich zu beschreiben, die es ermöglichen, einen kardinalen Nutzenwert zu beschreiben, mit dessen Hilfe alle denkbaren Therapiealternativen in einem Indikationsgebiet über alle relevanten Endpunkte hinweg verglichen werden können. Mit dem Leiter der „Conjoint Analyse“-Pilotstudie, Prof. Dr. Axel Mühlbacher (Hochschule Neubrandenburg, IGM Institut Gesundheitsökonomie und Medizinmanagement sowie Mitgründer der Gesellschaft für empirische Beratung mbH), sprach „Monitor Versorgungsforschung“.

>> Herr Prof. Mühlbacher, die Ausgangslage, warum das IQWiG überhaupt in der Bewertung von Kosten-Nutzen-Verhältnissen sowie der dafür nötigen Analyse der Effizienzgrenze einen deutschen Weg, der oft als deutscher Sonderweg abgetan wurde, ging, war darin begründet, dass das Institut die vor allem im angelsächsischen Bereich vorherrschende QALY-Systematik nicht einfach übernehmen wollte.

Der QALY wurde damals vor allem aus ethischen Gründen abgelehnt. Das Konzept wird aber durchaus auch aus methodischen Gründen in Frage gestellt. Die Vorteile des QALY-Konzeptes sind zum einen die vordergründige Plausibilität und zum anderen der breite internationale Erfahrungsschatz. Man bewertet Gesundheitstechnologien unter der Annahme, dass der Patientennutzen durch die Lebenslänge und die Lebensqualität abgebildet werden kann. Bewertungskriterien, welche in klinischen Studien berücksichtigt werden, bleiben außen vor. Konkreter: Studien, welche für die Preisfestsetzung mit Hilfe der Kosten-Nutzen-Bewertung herangezogen werden, haben keinen konkreten Bezug zu Studien, welche für die Zulassung, also der Dokumentation des Nutzen-Risiko- bzw. Nutzen-Schaden-Verhältnisses durchgeführt werden. Beide Studienansätze sind zudem in unterschiedlichen Wissenschaftsdisziplinen angelegt. Die Ökonomen und Psychologen dokumentieren den Nutzen einer Gesundheitstechnologie durch die Lebensqualität bzw. die lebensqualitätsbereinigten Lebensjahre und die Mediziner konkretisieren den medizinischen Nutzen anhand patientenrelevanter Endpunkte bzw. klinischer Effektmaße. Die dazu notwendigen Studiendesigns laufen parallel und bauen nicht aufeinander auf.

Das macht nun die Effizienzgrenze anders.

Stimmt. Mit der Effizienzgrenze, wenn sie denn ein theoretisches Fundament hat, wäre ein Instrumentarium geschaffen, das die Ergebnisse klinischer Studien berücksichtigt bzw. patientenrelevante Endpunkte in eine Nutzenbewertung überführt. Will man die Kosten-Nutzen-Bewertung auf den Ergebnissen der evidenzbasierten Medizin aufbauen, d.h. medizinische und ökonomische Studien durchgängig gestalten, dann ist die Logik der Effizienzgrenze ein zukunftsweisender Ansatz.

Nehmen wir einmal an, in einer klinischen Studie wird das Nutzen-Risiko-Verhältnis einer bestimmten Technologie beleuchtet...

... dann wäre es mit der Effizienzgrenze theoretisch denkbar, auf

Basis der klinischen Effekte und der Nebenwirkungen eine ökonomische Nutzenbewertung durchzuführen, also das Kosten-Nutzen-Verhältnis zu bestimmen. Nach der Zulassung könnte der Entscheidungsträger die Preise von Gesundheitstechnologien auf Grundlage der evidenzbasierten Medizin diskutieren.

Die Einführung der Effizienzgrenzen war bekanntlich 2009 nicht gerade mit Fortune gekrönt. Dennoch scheint sie eine rationale, verlässliche und auch nachvollziehbare Art und Weise zu sein, mit der man Nutzen-Kosten-Bewertungen durchführen könnte.

Die Effizienzgrenze wäre dann rational, verlässlich und nachvollziehbar einsetzbar, wenn die Methode der Nutzenbewertung wissenschaftlichen Ansprüchen genügt. Die Kritiker der Methodenpapiere sahen hier berechtigterweise noch Ergänzungsbedarf. Dem IQWiG muss man aber zu Gute halten, dass die Fragen, mit denen wir uns heute beschäftigen, erst durch dieses Methodenpapier aufgeworfen worden sind.

Die erste Frage war sicher, welche Endpunkte überhaupt genutzt werden können, um Effizienzgrenzen zu berechnen.

Ja. Zu Beginn jeder Bewertung muss man sich für bestimmte Bewertungskriterien entscheiden. In diesem Fall patientenrelevante Endpunkte, was gewiss keine triviale Frage ist und erheblichen Einfluss auf das Ergebnis hat. Noch wichtiger ist allerdings die Frage, wie diese Endpunkte letztendlich bei der Bewertung des Nutzens gewichtet werden. Die große Lücke im Methodenpapier war die Aggregation von unterschiedlichen Endpunkten - also die Berechnung des Gesamtnutzens. Dieser Gesamtnutzen sollte auch der Entscheidung des G-BA zugrunde liegen. Die Gretchenfrage lautet: Gibt es eine Rationale, welche die zu betrachtenden klinischen Effekte, mit den von ihnen generierten Patientennutzen in Beziehung setzt? Und das nicht nur pro Endpunkt, sondern auch bei mehreren Endpunkten, welche sich ja durchaus auch widersprechen können.

Gesamtnutzen heißt, man nimmt alle Effektparameter über alle Endpunkte ...

... und generiert daraus einen Nutzenwert, mit dessen Hilfe alle denkbaren Therapiealternativen in einem Indikationsgebiet über alle relevanten Endpunkte verglichen werden. Damit könnte dann eine Ko-

sten-Nutzen-Bewertung durchgeführt werden. Wir dokumentieren also den Gesamtnutzen anhand der zugrundeliegenden klinischen Evidenz.

Wird hier bereits eine Vergleichsgröße diskutiert, die einen Vergleich über alle Indikationen ermöglicht?

Noch nicht. Der Patientennutzen muss in Deutschland immer nur in einem bestimmten Indikationsgebiet verglichen werden. Für die Pilotprojekte hat dies die Aufgabe vereinfacht. Insofern bedurfte das Maß des Gesamtnutzens keiner indikationsübergreifenden Nutzendimension, sondern nur eines dimensionslosen Nutzenscores oder -index.

Wobei genau gegen diese indikationsspezifische Bewertung die Ökonomen geradezu Amok gelaufen sind.

Verständlich. Aus der Perspektive des Ökonomen wird eine Kosten-Nutzen-Bewertung durchgeführt, wenn Informationen über die optimale Allokation generiert werden sollen. Wenn man nun den Bewertungsrahmen der Effizienzgrenzen einsetzt, um eine indikationsspezifische Entscheidung zu treffen, dann ist das ein Schritt in die richtige Richtung. Aus der Perspektive der Ökonomie ist das aber noch nicht der finale Schritt.

Der finale Schritt wäre der indikationsübergreifende Vergleich.

Die derzeitige Interpretation des SGB V lässt diesen Schritt so nicht zu. Losgelöst von der politischen Debatte ist mit unserem Methodenansatz der indikationsübergreifende Vergleich möglich, d.h. eine optimale Entscheidung über alle Interventionen im Gesundheitswesen durchführbar. Das bedingt aber, dass man sich auf eine Nutzendimension bzw. auf ein Maß des Gesamtnutzens einigt, welcher über alle Alternativen im Gesundheitswesen erhoben werden kann.

Das ist möglich?

Durchaus. Namhafte Ökonomen diskutieren seit längerem monetäre Äquivalente als einen Lösungsansatz. Wem dies zu weit geht, der könnte eventuell auf das Äquivalent gesunder Lebensjahre zurückgreifen. Ich bin optimistisch, dass wir auch im Rahmen der Präferenzmessung zukünftig Methodenansätze diskutieren, welche eine indikationsübergreifende Bewertung von Technologien mit der Effizienzgrenze erlauben. Dies wird aber nur dann geschehen, wenn die Entscheidungsträger danach verlangen. Zudem würde dies die offene Frage der Extrapolation im Rahmen der Effizienzgrenze weiter vorantreiben.

Hätte dies das IQWiG vor der Veröffentlichung des Methodenpapiers wissen müssen oder kann man die Diskussion eigentlich nur auf Basis dieses Methodenpapiers führen?

Wenn man fair ist: letzteres. Ein erstes Methodenpapier abseits des Mainstreams zu veröffentlichen, war ein sehr mutiger Akt. Vor allem auch deshalb, weil es nationale und internationale Experten gibt, welche sich hervorragend mit den traditionellen Methoden auskennen, während es mit der Effizienzgrenze so gut wie keine praktischen Erfahrungen gab. Die Effizienzgrenze gibt einen Rahmen für interessante Methoden der Nutzenbewertung, wie die Discrete-Choice-Analyse oder das Best-Worst-Scaling. Ich bin überzeugt, dass wir parallel in den nächsten vier bis fünf



Jahren sehr spannende alternative Ansätze der Nutzenmessung haben werden.

Warum erst in den nächsten vier bis fünf Jahren?

Zum einen wäre es notwendig, die ökonomischen Grundlagen der Effizienzgrenze noch genauer zu betrachten bzw. zu diskutieren. Bei der Nutzenbewertung kommt es jetzt darauf an, dass wir die Übertragung der klinischen Evidenz in die ökonomischen Modelle der Nutzenbewertung besser verstehen. Zum anderen braucht es sicherlich Zeit, sich mit dem einen oder anderen Vorschlag zur Ermittlung eines endpunkt-basierten Nutzenwertes kritisch auseinander zu setzen. Wir haben das Pilotprojekt ja gerade erst fertig gestellt.

Damit kann die Diskussion wieder beginnen, wie damals bei der Vorstellung der Effizienzgrenze.

Kritische Diskussionen sind sicher nicht immer schön, sie sind aber notwendig, um den Weg weiter erfolgreich fortsetzen zu können. Ein bisschen provokativ formuliert, hatten die QALY-Experten 25 Jahre Zeit, ihre theoretischen und methodischen Probleme zu lösen. Geben wir also der Effizienzgrenze und der multiattributiven Nutzenanalyse auch ein bisschen Zeit.

Solange jedoch das Modell nicht in der Praxis eingesetzt wird...

... ich weiß. Solange haben die QALY-Befürworter noch die schlagkräftigeren Argumente. Doch ich glaube fest daran: Das Interesse des IQWiG und des G-BA vorausgesetzt, ist die oben diskutierte Problematik der Effizienzgrenze mit Hilfe von Präferenzstudien lösbar. Das haben wir im IQWiG-Pilotprojekt versucht aufzuzeigen.

Wenn das IQWiG die Pilotstudien veröffentlichen würde.

Ich halte die Veröffentlichung und die Diskussion für wichtig. Ich kann mir einfach nicht vorstellen, wie wir in Deutschland ohne Kosten-Nutzen-Bewertung auskommen wollten. Das AMNOG hat jetzt zwar erst einmal den Druck heraus genommen. Doch mittelfristig wird die Kosten-Nutzen-Bewertung auch in Deutschland wieder diskutiert werden. Das hängt jedoch auch von künftigen politischen Konstellationen ab. So ist es nach der nächsten Bundestagswahl sicherlich denkbar, dass neben der Finanzierungsdebatte auch die Kosten-Nutzen Bewertung wieder in den Fokus rücken könnte.

Nachdem aber nun der erste Bestandmarkt aufgerufen worden ist, wird alleine schon deshalb das Thema Kosten-Nutzen-Bewertung wieder mehr in den Fokus rücken.

Denkbar ist das. Mit der erstmaligen Nutzenbewertung für Arzneimittel aus dem sogenannten Bestandmarkt, die zur Behandlung des Diabetes mellitus Typ 2 zugelassen sind, gilt es die Wirkstoffe und die Wirkstoffkombinationen zu vergleichen. Auch hier muss der Gesamtnutzen der zur Verfügung stehenden Alternativen bestimmt werden.

Sind wir denn noch weit davon entfernt, alle notwendigen Informationen für eine Kosten-Nutzen Bewertung aus dem Dossier abzuleiten?

Ich denke nicht. Neben der Benennung der zweckmäßigen Vergleichstherapie und dem therapeutischen bedeutsamen Zusatznutzen enthält

das Dossier heute schon die Kosten der Therapie für die GKV. Der Vorteil wäre, dass eine funktionierende Effizienzgrenze eine Rangreihung erlaubt, während dessen zum Beispiel die QALY-Systematik das „The winner takes it all“-Prinzip verfolgt. Was eben noch fehlt, ist eine Methode, die aus unterschiedlichen betrachteten Endpunkten einen patientenseitigen Gesamtnutzen aggregiert.

Das ist ein schrecklicher Begriff.

Wir haben noch keinen besseren gefunden. Ich würde es zukünftig gerne als endpunktbasierten Nutzenscore oder -index bezeichnen. Doch wie immer man diese Aggregation von Endpunkten in eine einheitliche Kennzahl bezeichnet: Sie ist für die Entscheidungsfindung wichtig.

Was steht am Anfang?

Die Frage nach der Identifizierung von Endpunkten. Hier liegt es nahe, dass wir hier die Rolle der Patienten stärken. Warum nicht vor der klinischen Studienphase auch eine Präferenzstudie durchführen, um die Relevanz einzelner Bewertungskriterien aus der Perspektive der Patienten zu klären.

Was macht aus diesen Endpunkten denn nun patientenrelevante?

Eben diese Präferenzstudien. Über die Erhebung der Patientenpräferenzen finden wir heraus, welche Prioritäten Patienten wirklich haben. Wir können auch die gleiche Studie mit Experten durchführen und so Unterschiede bei der Bewertung von einzelnen Endpunkten aufzeigen. Diese Informationen könnten so eine weitere Perspektive - nämlich die der Patienten - in die wissenschaftliche Diskussion der jeweiligen Gremien einbringen. Grundlage des Patientennutzens sollten doch auch die Bewertungskriterien sein, welche der Patient für die Entscheidung heranziehen würde. Fehlt ein patientenrelevantes Bewertungskriterium, dann kann man doch nicht von einer Patientenorientierung oder einem Patientennutzen sprechen.

Die Identifizierung von Endpunkten ist sehr kritisch. Wer heute Studien für den AMNOG-Prozess macht, wird doch nur auf jene Endpunkte abstellen, die in klinischen Studien bereits gut dokumentiert und breit anerkannt sind.

Das ist ein wichtiger Punkt. Genau hier hatten wir am Anfang auch Probleme. Die Ausschreibung des IQWiG hieß ja: Identifizierung, Gewichtung und Priorisierung. Wenn man die Aufgabe ganz pragmatisch betrachtet, ging es nicht darum, einen neuen Endpunkt zu identifizieren, sondern die zu nehmen, welche in klinischen Studien bereits gut dokumentiert sind. Es ging also mehr um die Gewichtung unterschiedlicher Endpunkte und letztendlich um die Aggregation in ein einheitliches Maß des Gesamtnutzens. Damit wäre dann eine Priorisierung erst möglich.

Was für die Nutzen-Kosten-Bewertung als auch für die frühe Nutzenbewertung gilt.

Absolut. Es ist für mich nicht nachvollziehbar, warum der Nutzen einer Technologie bei der frühen Nutzenbewertung unterschiedlich abgeleitet werden sollte als bei der Berechnung einer Effizienzgrenze bzw. bei der Kosten-Nutzen-Bewertung.

Wer die bisherigen Beschlüsse der frühen Nutzenbewertung betrachtet, wird erkennen, dass eigentlich alle patientenrelevante Endpunkte negiert wurden, die nicht mit harten Endpunkten auf Studienbasis validiert worden sind. Die Frage lautet doch: Wie wichtig ist es dem System, was

ihm der Patient rückspiegelt? Oder ist nur das wichtig, was das System aus welchen Gründen auch immer als wichtig erachtet?

Grundlage der Entscheidung ist immer die evidenzbasierte Medizin. Da ich kein Mediziner bin, will ich mir kein Urteil zur Validität von Endpunkten oder Surrogatparametern erlauben. Für mich als Ökonom stellt sich aber folgende Frage: Sind Patienten bereit, zwischen unterschiedlichen Effekten einer Gesundheitstechnologie zu wählen? Konkret: Sind Patienten bereit, für ein Mehr an Wirkung auch ein höheres Risiko einzugehen? Oder: Würden Patienten auf gesunde Lebensjahre verzichten, um das Risiko eines Schadens durch Nebenwirkungen zu reduzieren? Beide Fragen sind relevant, sowohl für die Zulassung und die Preisfestsetzung.

Das sind harte Entscheidungen.

Das ist mir bewusst. Ebenso, dass die Prioritäten der Patienten von Experten nicht immer mitgetragen werden können. Doch wäre es falsch, wenn wir diese Prioritäten der Betroffenen bei der Zulassung und Preisfestsetzung nicht zum Gegenstand der Beratung machen würden. Es sollte klar sein, dass sowohl Patienten, aber auch Experten letztendlich subjektive Entscheidungen treffen. Wenn Patienten bestimmte Eigenschaften vorziehen, spricht dies doch erst einmal dafür, dass eine Technologie mit diesen Eigenschaften auch einen höheren Wertbeitrag, also Patientennutzen liefert.

Das wird bei IQWiG und G-BA derzeit nicht abgebildet.

Wenn die aktuellen Methoden diesen Wertbeitrag nicht abbilden können, müssen wir uns Gedanken machen, ob wohl etwas an den Patienten falsch ist oder an der Methode. Ich tippe auf die Methoden.

Dann wird wohl der Spitzenverband Bund argumentieren: Das können Sie gerne machen, solange Sie das selbst, aber nicht das Solidarsystem zahlen.

Das ist eine spannende und berechtigte Frage. Spannend deshalb, da ich als Bürger meine Entscheidungsrechte an die zuständigen Entscheidungsgremien delegiere. Der Grund liegt in den Informationsasymmetrien zwischen betroffener Person und dem jeweiligen Experten. Ich delegiere aber unter der Annahme, dass mein Wohl als Patient und Versicherter die Handlungsmaxime beziehungsweise das Wertesystem der Entscheidungsträger bestimmt. Der Einwand ist berechtigt, da diese Problematik heftig diskutiert wird und wiederum jetzt die Bevölkerungsprioritäten eine Rolle bei der Entscheidung spielen sollten.

Wobei es ein Unterschied ist, ob man als Versicherter oder als Patient, noch dazu als chronischer denkt.

Exakt. Als Versicherter wünsche ich mir den optimalen Einsatz der knappen Ressourcen, als Patient den maximalen Nutzen aus der Therapie. Beide Anforderungen bedürfen der systematischen Nutzenbewertung beziehungsweise der Kosten-Nutzen-Bewertung. Aber ich kenne kein Argument, warum die Präferenzen der betroffenen Patienten nicht in die Entscheidung eingehen sollten. Dabei sollten wir nicht vergessen, dass wir über Präferenzen einer repräsentativen Gesamtheit sprechen, nicht über die Präferenzen einzelner Personen.

Sie sprechen hier auch die Legitimierung des G-BA an. Es ist sicher ein Unterschied, ob eine Legitimation auf Basis von Präferenzmessungen eines Patientenkollektivs dargestellt wird oder nur durch die Beauftragung seitens der Bundesregierung.

Ich gehe fest davon aus, dass der G-BA im Auftrag der Patienten und Versicherten handelt. Ist das nicht auch das Ziel der Regierung? Deshalb gehe ich ebenso davon aus, dass zukünftig das Interesse an Studien über Patientenpräferenzen steigt. Gerade wenn diese Studien einige wesentliche Probleme der Effizienzgrenze bzw. der Nutzenbewertung lösen und damit die Patientenperspektive bei den Entscheidungen über die Zulassung und die Preisfestsetzung berücksichtigt werden kann.

Kommen wir zurück auf die Bestimmung des Patientennutzens. Sind die Endpunkte einmal identifiziert, müssen sie auch gewichtet werden.

Es ist offensichtlich, dass bei der Berücksichtigung von drei Endpunkten nicht unbedingt jeder Endpunkt mit einem Drittel in die Bewertung eingehen muss. Manche Endpunkte mögen wichtiger, manche unwichtiger sein. Die relative Wichtigkeit sollte ausschlaggebend für die Bewertung sein. Dazu muss man die Frage beantworten können, wie relevant ein Endpunkt im Vergleich mit den anderen Endpunkten ist. Das heißt aber auch: Wir brauchen eine Methode, welche uns ein relatives Gewicht für jeden der Endpunkte zur Verfügung stellt.

Da kann man sich vorstellen, dass es hier sicher endlose Diskussionen gab.

Das war gar nicht so schlimm, denn dafür gibt es wissenschaftlich mehr oder weniger valide Instrumente. Die gleichen Methoden werden zum Beispiel in der Umweltökonomie und den Verkehrswissenschaften eingesetzt, um die Präferenzen, respektive den Nutzen von Umweltmaßnahmen und Infrastrukturentscheidungen zu messen. Eine Gesundheitstechnologie, eine politische Reform oder eine medizinische Dienstleistung kann in die wesentlichen Eigenschaften heruntergebrochen werden. Ähnlich unserer tagtäglichen Entscheidungsprozesse sind diese Eigenschaften ausschlaggebend für die Wahl einer Alternative. In Experimenten können wir diese Entscheidungen reproduzieren und statistisch auswerten. Jede Alternative kann somit anhand der Realisation von unterschiedlichen Ausprägungen bewertet werden. Mit der Aggregation dieser Teilnutzen erhält man dann einen Nutzenwert, der es erlaubt, unterschiedliche Alternativen zu bewerten.

Das klingt dennoch ganz kompliziert...

... ist aber im Endeffekt ganz einfach. Nehmen wir als Beispiel eine Entscheidung über ein Automobil. Eigenschaften eines Autos können die Motorleistung, die Anzahl der Sitze und die Ausstattung sein. Die Ausprägungen der Motorleistung können 100 PS, 200 PS und 400 PS sein. Anzahl der Sitze 2, 4 und 6, die Ausstattung kann das Autoradio, C-Wechsler oder ein komplettes Audiosystem mit Telefonanlage sein. Wenn es nun mit einer Studie gelingt, diesen Ausprägungen einen Wert zuzuordnen, dann kann man jeder denkbaren Kombination einen Gesamtnutzen zuordnen.

Funktioniert das auch im Bereich der Gesundheit?

Sicher. Die Endpunkte kennen wir aus den klinischen Studien. Das sind - übertragen formuliert - nichts anderes als die Eigenschaften eines Produktes. Eine klinische Intervention erzielt bestimmte Effekte und damit werden produktspezifische Ausprägungen realisiert. Nun muss man lediglich über die realistischen Spannbreiten unterschiedliche Produktvarianten konstruieren und Menschen zwischen den Alternativen wählen lassen. Diese Wahl ist für mich die wichtigste Meinungsäußerung und die einzig realistisch messbare. Aus dieser Wahlentscheidung kann man Rückschlüsse auf die Präferenzen ziehen, also auch den Nut-

zen bestimmen.

Wie kann man diese Werte ermitteln?

Wir führen Befragungen mit den betroffenen Bevölkerungsgruppen durch. Wesentlicher Bestandteil ist die Frage der Probanden nach der Wahl zwischen unterschiedlichen Szenarien oder Alternativen. Wir beschreiben also ein Arzneimittel anhand von Endpunkten, wie z.B. Lebensverlängerung, Schmerzreduktion, Wirkungsdauer, Nebenwirkungen. Mit Hilfe verschiedener Ausprägungen dieser Eigenschaften können wir dem Teilnehmer einer Befragung zwei oder mehrere Alternativen zur Auswahl stellen. Mit der Wahl zeigt uns der Patient, welche dieser Alternativen er präferiert. Grundlage der Präferenz ist der wahrgenommene Nutzen aus diesen Alternativen. Diese wird ermittelt durch die statistische Auswertung der Wahlentscheidungen auf Basis der Präferenzen in Bezug auf die zur Auswahl stehenden Alternativen. Aus den berechneten Teilnutzen der einzelnen Eigenschaften bzw. der zugrundeliegenden Ausprägungen wird der Gesamtnutzen abgeleitet. Wichtig ist dabei lediglich, dass diese Gewichtungsfaktoren durch eine Methode bestimmt werden, welche auch wirklich die Präferenzen erheben kann.

Warum ist die Wahlentscheidung so wichtig?

Nehmen wir die Likert-Skala mit fünf Kästchen. 1 steht für nicht-wichtig und 5 für sehr wichtig. Verhandeln Sie mal mit ihrem Sohn oder ihrer Tochter über ein neues Auto. Schreiben Sie dazu alle möglichen Kriterien von PS, Design bis Ausstattung auf. Was meinen Sie, was sie oder er wohl ankreuzen wird? Design: wichtig 5. PS: wichtig 5. Ausstattung: wichtig 5. Aber steht dann nachher ein Porsche vor dem Haus? Eben nicht, weil es eine Budgetbeschränkung gibt und jeder gewohnt ist, sogenannte Trade-Offs zuzulassen. Weil man sich im Leben eben immer entscheiden muss zwischen bestimmten Eigenschaften. Oder nehmen Sie den Vergleich zweier Eigenschaften mit Hilfe einer Skala von 1 bis 9. Sie müssen bei diesem Vergleich angeben, um wie viel mehr sie die Motorleistung gegenüber der Ausstattung bevorzugen.

Können wir das wirklich?

Nein. Wir können Dinge und Eigenschaften in eine Rangordnung bringen, aber es ist in der Realität nicht zu beobachten, dass wir Eigenschaften einen konkreten Wert zuordnen - auch nicht im Vergleich zu einer Alternative. Können Sie die Frage beantworten, wie viel Wert auf einer Skala von 1 bis 9 Sie einer PS-Steigerung von 10 PS bei ihrem Auto im Vergleich zu einem zusätzlichen CD-Wechsler zuordnen? Eine ähnliche Frage müsste der Studienteilnehmer in einem Analytic Hierarchy Process, kurz AHP, beantworten. Mit der Choice-format Conjoint Analyse oder einem Discrete-Choice-Experiment werden dagegen lediglich die Wahlentscheidungen modelliert.

Wie soll das funktionieren?

Ein vereinfachtes Beispiel: Ein bestimmter Wirkstoff verlängert das Leben um 6 Monate, 8 Monate oder 10 Monate - das sind die Eigenschaftsausprägungen. Den Wirkstoff muss man einmal, zweimal oder auch dreimal pro Woche spritzen. Die Nebenwirkung sind starke oder leichte Kopfschmerzen und Müdigkeit. Bei der Alternative 1 erzielt der Wirkstoff eine Lebensverlängerung von 10 Monaten, die Nebenwirkungen sind starke Kopfschmerzen und Müdigkeit und gespritzt werden muss zweimal die Woche. Bei Alternative 2 wird einmal pro Woche gespritzt, der Patient hat nur leichte Kopfschmerzen und Müdigkeit, aber der Wirkungsgrad liegt bei nur 6 Monaten. Das Einzige, das wir von einen

Teilnehmer erfragen: Bitte sagen Sie uns, welche Alternative würden Sie bevorzugen! Bevorzugt er Alternative 1, dann haben wir einen ersten Anhaltspunkt, dass das Kriterium Lebensverlängerung den Patientennutzen stärker beeinflusst, als die Kriterien der Anwendung und Nebenwirkung. Antwortet er mit der Alternative 2, dann haben wir die Information, dass dieser Befragte bereit ist, eine geringere Lebenszeit für eine Verringerung der Nebenwirkungen bzw. eine niedrigere Spritzrate in Kauf zu nehmen. Je mehr Fragen wir in dieser Form stellen, desto genauer wird die statistische Analyse.

Die Kernfrage wird sein: Macht man das mit dem Analytic Hierarchy Process oder der Conjoint Analyse?

Die wenigen Befürworter der AHP argumentieren, dass die AHP kognitiv einfacher von Patienten zu verstehen ist. Zudem soll die Studie mit weniger Leuten durchführbar sein, d.h. der Aufwand soll geringer und damit das Instrument günstiger sein.

Das scheint ein Hammerargument zu sein.

Ich würde diesen Argumenten nicht zustimmen wollen. Nehmen wir es auseinander. Das Frageformat der AHP ist vordergründig einfacher - das stimmt. Ein Teilnehmer muss zwei Eigenschaften vergleichen, also Lebensverlängerung mit der Nebenwirkung Kopfschmerz. Dieser Vergleich beinhaltet nicht, ob es um 10 Jahre oder nur um 10 Monate Lebensverlängerung oder um lähmenden Kopfschmerz oder leichtes Kopfweh geht. Nicht sehr realistisch! Zudem muss er aber angeben, wie stark er eine Eigenschaft präferiert, dies auf einer Skala von 1 bis 9. „Ich präferiere Lebensverlängerung 9 mal mehr als Kopfweh“ - Das ist noch weniger realistisch. Dazu kommt, dass diese Vergleiche hierarchisch aufgebaut werden. Die gesamte Entscheidungssituation wird zu keinem Zeitpunkt für den befragten Patienten offensichtlich. Dies kann zu Verzerrungen bei der Bewertung führen. Darüber hinaus wird in der Literatur diskutiert,

dass es zu sogenannten „rank reversals“ kommt. Das bedeutet, dass sich unter bestimmten Umständen die Rangreihung der besten Alternativen verändert. Dies geschieht jedoch nicht, weil sich die Antworten der befragten Personen verändern, sondern aufgrund der Auswertungslogik. Diese birgt wiederum die Gefahr von Fehlurteilen.

Warum braucht man bei AHP weniger Probanden?

Weil es in der Literatur schlichtweg keine Aussagen dazu gibt. Es wird eine interne Konsistenzprüfung gemacht. Der Schwellenwert wurde aber vom Erfinder der Methode, dem Mathematiker Thomas L. Saaty, ohne weitere Erklärung willkürlich festgelegt.

Wenn dem so ist, kann man das Instrument AHP doch gar nicht im Rahmen der Effizienzgrenzen-Analyse einsetzen.

Ob ja oder nein, wird die Diskussion der nächsten Monate zeigen. Heute würde ich es nicht empfehlen. Mich stört die fehlende theoretische Einordnung und noch mehr das wenig realistische Frageformat.

Ob nun CA oder AHP das Rennen machen wird, bei beiden Ansätzen steht zumindest endlich der Patient im Mittelpunkt.

Das ist die Idee, die beide Arbeitsgruppen der Pilotprojekte angetrieben hat. Versorgung muss letztlich auf den Präferenzen der Betroffenen basieren. Wenn man die Präferenzen erst einmal kennt, wäre es töricht, diese nicht zu berücksichtigen. Die ersten Schritte auf diesem Weg haben wir aber schon einmal hinter uns.

Danke für das Gespräch. <<

Das Interview führte MVF-Chefredakteur Peter Stegmaier

Bitte beachten Sie die Ankündigung des nächsten MVF-Kongresses zum Thema „Patientenrelevante Endpunkte“ auf S. 22

BALD ERHÄLTlich:

Disease Management Programme Statusbericht 2012

Wissenschaftliche und redaktionelle Beiträge der MVF-Fachkongresse „10 Jahre DMP“ und „Versorgung 2.0“



Mit Beiträgen von

L. Altenhofen | C. Bartel | C. Bensing | J. Biermann | J. Böcken | S. Burger | T. Czihal | S. Ein Waldt | M. Ewers | P. Fey | J. Fröhlich | M. Gaßner | C. Gawlik | J. Gebhardt | G. Glaeske | C. Graf | S. Guthoff-Hagen | J. Härtel | B. Häußler | B. Hagen | E. Höfer | F. Hofmann | W. Hoffmann-Eber | D. Horenkamp-Sonntag | L. Kerkemeyer | C. Kiefer | A. Kleinfeld | K. Koch | T. Köhler | K. Köhn | T. König | J. Leinert | R. Linder | U. Moldenhauer | S. Mostardt | H.-D. Nolting | I. Oelmann | H. Rebscher | M. Redaelli | S. Roebbers | R. Roski | A. Rüther | D. Schaeffer | U. Siering | S.G. Spitzer | P. Stegmaier | D. Graf von Stillfried | S. Stock | S. Südhof | E.J. van Lente | M. Wagner | P. Willenborg | K. Wegscheider | K. Zich |

Erscheint demnächst im Verlag von „Monitor Versorgungsforschung“.
Bestellen Sie direkt beim Verlag: eRelation AG - Content in Health - Kölnstraße 119 - 53111 Bonn -
Tel. 0228 - 76 38 28 00 - info@erelation.org - ISBN 978-3-9814519-0-0