Dr. rer. biol. hum. Ch.-M. Dintsios, MPH

Patientengewichtete Endpunkte: Conjoint Analyse (CA) versus Analytic Hierarchy Process (AHP)

Die so genannte Multikriterielle Entscheidungsanalyse (MKEA) kann zur Unterstützung komplexer Technologiebewertungen eingesetzt werden. Ihren Ursprung haben multikriterielle entscheidungsanalytische Verfahren in der Entscheidungstheorie bzw. Operations Research und werden dort eingesetzt, um Entscheidungen unter Unsicherheit zu treffen. Sie unterstützen Entscheidungsträger bei der Bewertung einer endlichen Anzahl alternativer Gesundheitsinterventionen unter Verwendung einer endlichen Anzahl von Leistungskriterien. Zu den am meisten verbreiteten multikriteriellen Entscheidungsanalyseinstrumenten gehört neben der Conjoint-Analyse, dem Elimination and Choice translating Reality-Verfahren (ELECTRE) und der Multiattributiven Nutzwerttheorie (MAUT) (Lootsma/Schujit 1997) der Analytic Hierarchy Process (AHP).

>> Diese Instrumente unterscheiden sich hauptsächlich in ihren Ansätzen, Gewichte für Attribute der untersuchten Technologien zu erheben und präsentieren unterschiedliche statistische Modelle bei der Strukturierung der Präferenzen der Entscheidungsträger zu den zur Verfügung stehenden Alternativen. In dem von uns betrachteten Anwendungsbereich wären primär die befragten Patienten Entscheidungsträger und die jeweiligen Endpunkte innerhalb des betrachteten Krankheitsbildes die Leistungskriterien der zur Anwendung kommenden Gesundheitsinterventionen. Der Theorie nach liegt der Mehrwert dieser Instrumente in der Unterstützung der klinischen bzw. (gesundheits)ökonomischen Bewertung der Gesundheitstechnologien, die als multifaktorieller Prozess in entsprechende Entscheidungsfindungen eingehen.

Nichtsdestotrotz wird die Anwendbarkeit und Validität multikriterieller Entscheidungsanalyse hinterfragt. Beispielsweise liefern verschiedene multikriterielle Entscheidungsanalyseverfahren unterschiedliche Ergebnisse bei Anwendung im Rahmen desselben Entscheidungsproblems (Zanakis et al. 1998). Des Weiteren bleibt ihre prädiktive Validität oft genug spekulativ. Unternommene experimentelle Vergleiche schlussfolgerten, dass jede multikriterielle Entscheidungsanalysemethode ihre eigenen Vor- und Nachteile hat. Diese gilt es in vorliegender Abhandlung für die zwei etablierten Instrumente CA und AHP, die auch im Gesundheitsbereich (de Bekker-Grob et al. 2010; Wallenius et al. 2008) bzw. bei der Erhebung von Patientenpräferenzen immer mehr eingesetzt werden, kurz zu beleuchten und gegenüberzustellen.

Mathematisch-theoretische Grundlagen des AHP

AHP als kompositionelles Verfahren

Die Grundannahme des AHP lautet, dass der berechnete rechte Eigenvektor der zugrundegelegten Matrix A ungefähr dem Vektor der relativen Gewichte entspricht. Es handelt sich somit um ein kompositionelles Verfahren, das einem gewichteten additiven Nutzenmodell folgt (Abb. 1).

Zusammenfassung

Beide Verfahren, Conjoint Analyse (CA) und Analytic Hierarchy Process (AHP), werden zur Erhebung von Patientenpräferenzen (hier Priorisierung und Gewichtung von Endpunkten) eingesetzt. Über ihre theoretische Verortung hinaus (ökonomische Nutzentheorie versus deskriptive Theorie) unterscheiden sie sich hinsichtlich des zugrundegelegten Modells (additives Teilwertmodell versus gewichtetes additives Nutzenmodell), methodischen Ansatzes (dekompositionell versus kompositionell), Bewertungsprozesses (holistisch versus Paarvergleiche), methodischer Schwächen (kognitive Überforderung versus unrealen Entscheidungssituationen) und des Ressourcenaufwands (umfangreich versus gering) nicht unerheblich. CA kann bezogen auf a priori Annahmen zu widersprüchlichen Ergebnissen führen, AHP leidet unter dem so genannten Rank Reversal. Sie generieren allerdings in vergleichenden Studien durchaus ähnliche Ergebnisse und besitzen je nach Fragestellung ihre Vor- und Nachteile.

Schlüsselwörter

Multikriterielle Entscheidungsanalyse, Conjoint Analyse, Analytical Hierarchy Process, Patientenpräferenzen, Endpunktgewichtung, Nutzen-Schaden-Abwägung, Nutzenaggregierung

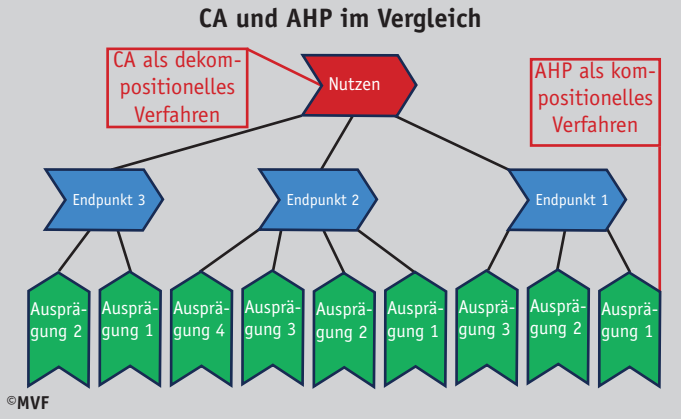


Abb. 1: Dekompositionelle CA versus Kompositioneller AHP.

Rechter Eigenvektor

Das AHP bietet einen Ansatz an, der zur Abschätzung von Gewichten der einzelnen Entscheidungskriterien und zur Priorisierung von zur Verfügung stehenden Alternativen geeignet ist. Diese Quantifizierung wird aus einer Matrix von paarweisen Vergleichen zweier Kriterien bzw. Alternativen hergeleitet. Der paarweise Vergleich der Entscheidungskriterien zeigt die relative Wichtigkeit dieser Kriterien bei der Zielerreichung, der Vergleich der Alternativen die relative Präferenz für die betrachteten Alternativen in Abhängigkeit von einem entsprechenden Entscheidungskriterium. Beide - relative Wichtigkeit der Kriterien und relative Präferenz der Alternativen - beruhen auf einer ordinalen Skala, die den numerischen Wertebereich von 1 für gleiche Wichtigkeit bzw. Präferenz bis zu 9 für extreme Wichtigkeit oder Präferenz innerhalb eines paarweisen Vergleichs einnimmt. Dabei stehen die einzelnen numerischen Ausprägungen der Ordinalskala für folgende Ausprägungen der relativen Wichtigkeit oder Präferenz:

- 1 für indifferente bzw. gleich wichtige Aussagen (Kriterien),
- 3 für etwas wichtiger,
- 5 für viel wichtiger,
- 7 für sehr viel wichtiger und
- 9 für extrem viel wichtiger.
- 2, 4, 6 und 8 werden jeweils als intermediäre Werte eingesetzt.

Die innerhalb des paarweisen Vergleichs betrachteten Kriterien können umgedreht werden, wenn die Richtung der Wichtigkeit sich umkehrt. Es entsteht eine Matrix von paarweisen Vergleichen über n Kriterien, mit A = (aij):

Matrix von paarweisen Vergleichen				
Kriterien	**Kriterium1**	**Kriterium 2**	**...**	**Kriterium n**
Kriterium 1	a_{11}	a_{12}	...	a_{1n}
Kriterium 2	a_{21}	a_{22}	...	a_{2n}
Kriterium 3	a_{31}	a_{32}	...	a_{3n}
...	...		...	
....	...		...	
Kriterium n	a_{n1}	a_{n2}	...	a_{nn}

Tab. 1: Matrix von paarweisen Vergleichen.

Sowohl Reflexivität (aij = 1 auf der Diagonalen) als auch Reziprozität (aij = 1/aji) werden in dieser Matrix vorausgesetzt. Aus diesem Grund müssen nur die Werte über die Diagonale eingetragen werden. Dies bedeutet, dass für n Kriterien nur n(n-1)/2 distinkte paarweise Vergleiche notwendig sind. Wenn im Rahmen des AHP Gruppen mit x Beteiligten befragt werden (x > 1), können die einzelnen Zellen der Matrix mit aggregierten paarweisen Vergleichen ausgefüllt werden, in dem das geometrische Mittel der individuellen Einzelbewertungen gebildet wird:

$$aij = (aij(1) * aij(2) * \ldots * aij(x))(1/x) \text{ für alle } i,j = 1, 2, \ldots, n.$$

Wenn alle paarweisen Vergleiche in perfekter Konsistenz zu den verbleibenden Vergleichen stehen, dann gilt die Transitivität, formal ausgedrückt wie folgt:

$$aik * akj = aij \text{ für alle } i, j, k = 1, 2, \ldots, n.$$

Das AHP liefert ein Konsistenzmaß, um auf mögliche Inkonsistenzen beim Kriterienvergleich hinzuweisen. Dieses Maß fußt auf dem maximalen Eigenwert λ_{max} der Matrix mit den paarweisen Vergleichen. Dieser Eigenwert ist immer größer oder gleich n für positive, reziproke Matrizen und gleich n, wenn die Matrix der paarweisen Vergleiche konsistent ist. Durch Normalisierung mit der Matrixgröße definiert Saaty den sogenannten Konsistenzindex C.I. in folgender Weise:

$$C.I. = (\lambda_{max} - 1)/(n - 1).$$

Saaty (1989) generiert für eine Matrix der Größe n in einem Randomisierungsverfahren zufällig 500 Matrizen und berechnet den durchschnittlichen Konsistenzindex C.I. aus diesen Matrizen, den er Random Index (R.I.) nennt. Die Konsistenzratio (C.R.) wiederum gibt an, wie weit die paarweisen Bewertungen von einer völlig zufälligen Matrix paarweiser Vergleiche abweichen und wird definiert als Verhältnis zwischen Konsistenzindex und Random Index: C.R. = C.I./R.I.

Nach Saaty sind 10% der Inkonsistenz der zufälligen Matrix erlaubt. Dies bedeutet, dass ein Wert für die Konsistenzratio von unter 10% (C.R. < 0,1) toleriert bzw. akzeptiert werden kann. Derselbe Eigenwert der Matrix mit den paarweisen Vergleichen wird nun eingesetzt, um die Gewichte der Kriterien abzuschätzen:

$$wi = (\Sigma n\ j=1\ aij * wj) / \lambda_{max} \text{ für alle } I = 1, 2, \ldots, n.$$

Diese Eigenvektormethode kann interpretiert werden als einfacher Mittelungsprozess, bei welchem die finalen Gewichte die Durchschnitte aller Vergleichsmöglichkeiten der Wichtigkeit der zu bewertenden Kriterien darstellen.

Die CA als dekompositionelles Verfahren

Die CA baut auf die Werttheorie nach Lancaster auf (Lancaster 1966). Der Nutzen eines „Gutes“ leitet sich aus den Eigenschaften und nicht aus dem „Gut“ selbst ab. Sie folgt somit der Theorie des stochastischen Nutzens (Random Utility Theory), wonach das Prinzip der Nutzenmaximierung gilt: Individuen entscheiden sich jeweils für die Alternative mit dem voraussichtlich höchsten Nutzen. Bei unvollständigen Informationen kann der Nutzen nicht vollständig erfasst werden. Er setzt sich additiv zusammen aus der beobachtbaren deterministischer Komponente (z.B. Merkmale/Attribute der Technologie, persönliche Eigenschaften) in Form einer linearen Nutzenfunktion (Schätzung der Gewichtung der Attribute) und aus einer nicht beobachtbaren stochastischer (Zufalls-)Komponente. Je nach Verteilungsannahmen erfolgt die Schätzung über ein Probit- oder Logit-Modell. Es handelt sich somit um ein dekompositionelles Verfahren, das einem additiven Teilwertmodell folgt (Tab. 1).

Logistische Regression

Die Conjoint Analyse folgt in der hier primär im Fokus stehenden Discrete Choice Variante einem (logistischen) Regressionsmodell, wonach für den zu ermittelnden Nutzen die Beziehung gilt:

$$U_{ij} = V_{ij} + \varepsilon_{ij}$$

(U_{ij} = Nutzen („utility“), ε_{ij} = stochastische unbeobachtbare Komponente (u.a. Unterbewusstsein, körperliche Verfassung, Gemütszustand), V_{ij} = deterministische Komponente (Merkmale des Gutes/der Leistung und des Individuums), X = Merkmale (Attribute) des Gutes/der Leistung j; z = Merkmale des Individuums i; β = Maß für die Stärke der Gewichtung der Merkmale (Attribute) des Gutes/der Leistung j und γ = Maß für die Stärke der Gewichtung der Merkmale des Individuums i.

Die Gleichungen $V_{ij} = V(X_{ij}, z_{ij})$ und $V_{ij} = \beta X_{ij} + \gamma z_{ij}$ setzen dann die deterministische Komponente in Relation zu den Attributen des Gutes und den Merkmalen des Individuums.

Die Schätzung der relativen Bedeutung (Gewichtung) eines Attributes (hier Endpunktes) erfolgt dann im Rahmen einer logistischen Regression nach den folgenden Gleichungen zu den jeweiligen deterministischen Komponenten, wobei die β-Koeffizienten die Endpunktgewichte liefern:

(i) $V(A) = \alpha + \beta_1 X_1^A + \ldots + \beta_n X_n^A$

(ii) $V(B) = \alpha + \beta_1 X_1^B + \ldots + \beta_n X_n^B$

und

(iii) $\ln {}^{p(A)}/_{p(B)} = \beta_1(X_1^A - X_1^B) + \ldots + \beta_n(X_n^A - X_n^B)$

Theoretische Verortung: ökonomische Nutzentheorie (CA) versus deskriptive Theorie (AHP)

Zwar fußen beide Verfahren auf den Axiomen der Nutzentheorie, aber beim AHP wird aufgrund der Schwierigkeiten präziser Präferenzbeurteilung durch Anwender auf restriktive Annahmen der Konsistenz und Transitivität der Präferenzäußerungen verzichtet. Somit werden wegen der weniger strengen Annahmen hinsichtlich der Präferenzstruktur auch teils irrationale nicht als falsche Entscheidungen identifiziert. Das AHP-Verfahren erfüllt somit wesentliche Anforderungen der ökonomischen Nutzentheorie nicht und gilt nutzentheoretisch als nicht fundiert. Saaty (1990) bezeichnet die dem AHP zugrunde

liegenden Theorie als *deskriptive* - von der *normativen Nutzentheorie* abzugrenzende - Theorie.

Methodische Gütekriterien CA versus AHP

Bei den aus der Sozialempirie, Psychometrie und Biometrie etablierten methodischen Gütekriterien wie Augenscheinvalidität, interne Validität, Konvergenzvalidität, prädiktive Validität und Modellpassung schneiden die beiden Methoden zwar in den jeweiligen Studien unterschiedlich ab, ohne dass sich aber dadurch ein Trendvorteil für eine Methode ergibt. Lediglich bei verschiedenen Dominanztests, bei denen die Übereinstimmung des Ergebnisses in Relation zur von den Befragten vorab als dominierend bezeichneten Alternative gemessen wird, scheint die CA etwas unterlegen zu sein. CA kann bezogen auf a priori Annahmen somit zu widersprüchlicheren Ergebnissen führen. Dominanztest sind ein Kriterium für die prädiktive Validität. Beispielhaft aus der hierzu vorhandenen Literatur seien zwei Studien zitiert (Mulye 1998; Scholl et al. 2005). Triantafyllou (2000) definiert allerdings in seiner Arbeit zwei Kriterien zur Bewertung der methodischen Güte: 1. Akkurate Methoden in multidimensionalen Entscheidungsfindungen sollen auch in eindimensionalen Fragestellungen erfolgreiche Lösungen liefern. Dies wird in den meisten Fällen von AHP (als gewichtetes Summenmodell) erfüllt. 2. Die Reihenfolge der Alternativen hin zur besten soll nicht geändert werden, wenn eine Alternative (nicht die beste) durch eine andere ersetzt wird, wobei die Attributsgewichte unverändert bleiben.

Hier tritt allerdings das Phänomen des Rank Reversal bei AHP auf, also der Nichteinhaltung genau dieses zweiten Gütekriteriums. Als Ursache des Rank Reversal wird die Ermittlung der relativen Gewichtungen mittels der Eigenwertmethode diskutiert. Obwohl das AHP-Verfahren auf den Axiomen der Nutzentheorie basiert, wird aufgrund der Schwierigkeiten einer präzisen Präferenzbeurteilung durch die Anwender auf die restriktiven Annahmen der Konsistenz und Transitivität der Präferenzäußerungen verzichtet. Somit werden wegen der weniger strengen Annahmen hinsichtlich der Präferenzstruktur auch teils irrationale nicht als falsche Entscheidungen identifiziert (Saaty & Vargas 2001). Das AHP-Verfahren erfüllt somit wesentliche Anforderungen der ökonomischen Nutzentheorie nicht und gilt nutzentheoretisch als nicht fundiert. Allerdings kann dieser Argumentation von Seiten der Verhaltensökonomie auch entgegengehalten werden, dass die auf den Annahmen des Homo economicus basierende Nutzentheorie einer nutzenmaximierend agierenden, der ökonomischen Rationalität folgenden individuellen Verhaltensweise nach dem Standardmodel der Ökonomie nicht dem realen menschlichen Verhalten in ökonomischen Entscheidungen entspricht (Tversky/Kahneman 1974).

Bezogen auf die *Rationalität* als *komplexe Verhaltenstheorie* leitet der Nobelpreisträger McFadden (1998) spezifisch für die Aspekte *Präferenzen, Perzeptionen* und *Prozess* über entsprechende Experimente her, dass die meisten kognitiven Anomalien aus Fehlern in der *Perzeption* herrühren, die damit zusammenhängen, wie Informationen gespeichert, abgerufen und verarbeitet werden bzw. aus Fehlern im Entscheidungsprozess selber, die zu einer Formulierung von - kognitiv betrachtet - inkonsistenten Entscheidungsproblemen, zumindest wenn man Rationalität in ihrer engen, nutzentheoretisch fundierten Definition wahrnimmt, führen. Shafir und LeBoef (2002) zeigen auch auf, dass Individuen oft Rationalitätsannahmen zu logischen Schlussfolgerungen, Beurteilungen und Wahlmöglichkeiten in ihren Entscheidungen abweichend vom normativen Verhaltensmodell verletzen. Sie unterstützen die mittlerweile in den Sozialwissenschaften aufkommende Kritik an der Rationalität als axiomatische Verhaltensform.

Methodenvergleich CA versus AHP		
	Conjoint Analyse	**AHP**
Voraussetzung	**Dekompositionell** Vorzugsweise Unabhängigkeit der Attribute, arbiträr sind Kombinationen von Attributsleveln (Ausprägungen) möglich	**Kompositionell** Vorzugsweise Unabhängigkeit der Attribute
Eingesetzte Skalen des Inputs	Ordinal oder Intervallskaliert	Semantisch deskriptive 9-Punkte-Verhältnisskala
Skalen des Outputs	Intervallskala	Verhältnisskala
Anwendbarkeit	Bis zu 6 Attributen mit 2-4 Level, Viele Stimuli	Viele Attribute, bis zu 7-8 Attribute bzw. Attributslevel (Auprägungen)
Methodischer Ansatz	Ermittlung der Teilnutzenbeiträge von Eigenschaften und deren Ausprägungen auf Basis empirisch erhobener Gesamtnutzenurteile	Direkte Bewertung der Eigenschaften und deren Ausprägungen mit anschließender Gesamtbewertung
Bewertungsprozess	Ganzheitlich: reale Entscheidungssituation wird simuliert	Paarvergleichsbewertung: Komplexitätsreduktion durch Dekomposition und hierarchische Strukturierung
Methodische Schwächen	Schnelle kognitive Überforderung der Probanden bei Vielzahl von Eigenschaften und deren Ausprägungen	Bewertung mittels Paarvergleichen entspricht nicht der realen Entscheidungssituation, z. T. mangelnde Stabilität der Ergebnisse
Ressourcenaufwand (Patienten, Zeit, etc.)	Patientenzahl bei repräsentativer Stichprobe oft über 100, digital und auf Papier machbar, mindestens 6 Wochen	Patientenzahl bei repräsentativer Stichprobe ab 25, Digital und auf Papier machbar, ca. 2 bis 3 Wochen

Tab. 2: Methodenvergleich: CA versus AHP.

Kognitive Überforderung (CA) versus unreale Entscheidungssituationen (AHP)

Einem fiktiven Beispiel mit fünf Attributen (N=5) und jeweils drei Ausprägungen bzw. Levels (L=3) für ein zu bewertendes Gut mit Alternativen folgend, ergeben sich nach den weiter oben beschriebenen mathematischen Grundlagen beider Verfahren beim AHP n (n-1)/2 = 10 Paarvergleiche (Trade-offs) und bei der CA 5^3 = 125 Szenarien, was bei einem simplifizierenden sogenannten orthogonalem Design, wonach die Annahme gilt, dass keine Interaktionseffekte zwischen den Attributen bestehen und somit keine Interaktionseffekte gemessen werden müssen mit der Konsequenz, dass jede Ausprägung eines Attributs mit jeweils einer Ausprägung eines anderen Attributes in einem Szenario vorkommen, 12 Szenarien pro befragten Patient ergibt. Berücksichtigt man die Trade-offs innerhalb der Szenarien erscheint die kognitive Herausforderung, ja sogar Überforderung bei der CA selbstevident im Vergleich zum AHP. Andererseits wird die Versorgungsrealität allerdings viel besser über solche Szenarien abgebildet, als durch abstrakte paarweise Vergleiche beim AHP. Abbildung 2 zur paarweisen Befragung beim AHP und Abbildung 3 zu den präferierten Szenarien im Rahmen einer CA, beide patientenseitige Endpunktgewichtungen zum Krankheitsbild Hepatitis C betreffend, sollen die jeweiligen Herausforderungen verdeutlichen.

Vergleichende Studien CA versus AHP

Ijzerman et al (2012) vergleichen AHP und CA bei der Bewertung von Alternativen in der Rehabilitation von Schlaganfällen. In den Gewichtungen der betrachteten Attribute dieser Alternativen - wobei aus

methodischen Gründen die Attribute des einen Verfahrens für das andere zusammengelegt werden müssen - kommen sie zum Ergebnis, dass es eine Übereinstimmung der jeweiligen Gewichte bei vergleichbaren Mittelwertstandartabweichungen für drei Attribute gibt (AHP versus CA: klinischer Outcome 0,30 versus 0,33; Einfachheit der Anwendung 0,16 mit beiden Methoden, Komplikationen 0,20 versus 0,19), Abweichungen dagegen bei zwei Attributen (AHP versus CA: Kosmetik und Komfort 0,28 versus 0,05; Einfluss der Behandlung operationalisiert als Dauer und Invasivität 0,06 versus 0,28). Der Vergleich zeigt darüber hinaus Unterschiede im Ressourcenverbrauch und in der kognitiven Beanspruchung der Befragten. Die CA ist aufwendiger und komplizierter.

Andere Vergleiche wiederum, wie die Studie von Ijzerman et al. aus dem Jahre 2008 zur Erhebung von Präferenzen bei Patienten mit neurologischen Störungen, zeitigen relativ ähnliche Attributsgewichte, wobei nicht-medizinische Faktoren wie soziale Akzeptanz und Anwenderfreundlichkeit der jeweiligen unterstützenden Behandlungen eine große Rolle zu spielen scheinen. In dieser Studie unterschieden sich die Gewichte zwischen Patienten und befragten klinischen Experten unabhängig von der angewandten Methode. Ein Phänomen, auf das man auch in anderen Indikationen wie beispielsweise der Depressionsbehandlung trifft (Hummel et al. 2012; Danner et al. 2011). In einer weiteren Studie (Benaïm et al. 2010) verwenden die Autoren CA und AHP zur Gewichtung der 6 Items der London Handicap Scale, eines Erhebungsinstruments der Beeinträchtigung, das bei der Schlaganfallprävention oft Anwendung findet. Mit beiden Methoden wird die physische Unabhängigkeit am höchsten gewichtet, Unterschiede zeigen sich jedoch in der Reihenfolge der anderen Items. Bei Anwendung der sogenannten „10 binary choices“ Methode nach Schoenemaker & Waid (1982), wonach 10 zufällige fiktive Patientenprofilpaare mit ähnlichen Summenscores gewählt werden und die Konkordanz zwischen schlechteren Profilen mit AHP & CA Ergebnissen verglichen wird (ein Kappa-Wert > 0,40 führt zu einer guten Übereinstimmung, ein Kappa-Wert ≤ 0 zeigt eine sehr schlechte Übereinstimmung), sind bei der CA in 3 von 10 Fällen gute Übereinstimmungen erzielt worden, in 1 eine moderate und in 6 ein schlechte Übereinstimmung, wohingegen beim AHP in 6 von 10 Fällen eine gute und in jeweils 2 eine moderate und eine schlechte Übereinstimmung erzielt wurden.

Verwertbarkeit von CA und AHP für das IQWiG

In ihrem Diskussionspapier zum Methodenvergleich bei der Gewichtung multipler patientenrelevanter Endpunkte entwickeln Neidhardt et al. (2012) einen Kriterienkatalog. Dieser orientiert sich an (i) den gesetzlichen Vorgaben, (ii) den Anforderungen aus der IQWiG-Methodik (Effizienzgrenzkonzept) und (iii) den allgemeinen Anforderungen an das wissenschaftliche Vorgehen. Unter den gesetzlichen Vorgaben subsumieren die Autoren die Transparenz und ein international etabliertes Verfahren der Gesundheitsökonomie. Zu den Anforderungen aus der IQWiG-Methodik zählen sie die approximative Kardinalität des aggregierten Nutzenmaßes, die Flexibilität und die Möglichkeit zur Aggregation

Literatur

Benaïm, C./Perennou, D. A./Pelissier, J. A./Daures, J. P. (2010): Using an analytical hierarchy process (AHP) for weighting items of a measurement scale: A pilot study. In: Revue d'Épidémiologie et de Santé Publique 2010, 58: 59–63

Danner, M./Hummel, J. M./Volz, F./van Manen, J. G./Wiegard, B./Dintsios, C. M./Bastian, H./Gerber, A./Ijzerman, M. J. (2011): Integrating patients' views into health technology assessment: Analytic hierarchy process (AHP) as a method to elicit patient preferences. Int J Technol Assess Health Care 2011, 27, 4: 369-75

De Bekker-Grob, E./Ryan, M./Gerard K. (2012): Discrete Choice Experiments in Health Economics: A review of the literature. In: Health Econ 2012, 21: 145–172

EMA (2010): Benefit-risk methodology project. Work package 2 report: Applicability of current tools and processes for regulatory benefit-risk assessment. In: http://www.ema.europa.eu/docs/en_GB/document_library/Report/2010/10/WC500097750.pdf (abgerufen am 16.01.2013)

FDA (2012): Guidance for Industry and Food and Drug Administration Staff - Factors to Consider When Making Benefit-Risk Determinations in Medical Device Premarket Approvals and De Novo Classifications. In: http://www.fda.gov/downloads/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/UCM296379.pdf (abgerufen am 16.01.2013)

Hummel, M. J./Volz, F./van Manen, J. G./Danner, M./Dintsios, C. M./Ijzerman, M. J./Gerber, A. (2012): Using the analytic hierarchy process to elicit patient preferences: prioritizing multiple outcome measures of antidepressant drug treatment. In: Patient 2012, 5, 4:225-37

IJzerman, M. J./van Til, J. A./Bridges, J. (2012): A Comparison of Analytic Hierarchy Process and Conjoint Analysis Methods in Assessing Treatment Alternatives for Stroke Rehabilitation. In: Patient 2012, 5, 1: 45-56

IJzerman, M. J./van Til, J. A./Snoek, G. J. (2008): Comparison of Two Multi-Criteria Decision Techniques for Eliciting Treatment Preferences in People with Neurological Disorders. In: Patient 2008, 1, 4: 265-272

IQWiG (2011): Allgemeine MethodenVersion 4.0. In: https://www.iqwig.de/download/IQWiG_Methoden_Version_4_0.pdf (abgerufen am 16.01.2013)

Lancaster, K. (1966): A New Approach to Consumer Theory. In: Journal of Political Economy 1966, 74:132-57

Lootsma, F.A./Schuijt, H. (1997): The multiplicative AHP, SMART and ELECTRE in acommon context. In: Journal of Multi-Criteria Decision Analysis 1997, 6: 185-196

McFadden, D. (1999): Rationality for Economists? In: Journal of Risk and Uncertainty 1999, 19, 1-3: 73-105

Mulye, R. (1998): An empirical comparison of three variants of the AHP and two variants of conjoint analysis. In: Journal of Behavioral Decision Making 1998, 11, 4: 263-280

Neidhardt, K./Wasmuth, T./Schmid, A. (2012): Die Gewichtung multipler patientenrelevanter Endpunkte - Ein methodischer Vergleich von Conjoint Analyse und Analytic Hierarchy Process unter Berücksichtigung des Effizienzgrenzenkonzepts des IQWiG. In: Diskussionspapier 02-12 Februar 2012. Universität Bayreuth, Wirtschaftswissenschaftliche Diskussionspapiere.

Saaty, T. L./Vargas, L. G. (2001): Models, methods, concepts & applications of the analytic hierarchy process. In International series in operations research & management science 2001, 34, Boston: Kluwer Academic Publishers

Saaty, T.L. (1990): An exposition of the AHP in reply to the paper "Remarks on the analytic hierarchy process". In: Management Science 1990, 36, 3: 259-268

Saaty, TL. (1989): Group decision making and the AHP. In: Golden, B. L./Wasil, E. A./Harker, P. T. (Hrsg.) (1989). The Analytic Hierarchy Process: Applications and Studies. Berlin; Springer

Schoenemaker, P./Waid, C. (1982): An experimental comparison of different approaches to determining weights in additive utility models. In: Management science 1982, 28: 182-196

Scholl, A./Manthey, L./Helm, R./Steiner, M. (2005): Solving multiattribute design problems with analytic hierarchy process and conjoint analysis: An empirical comparison. In: European Journal of Operational Research 2005, 164, 3: 760-777

Shafir, E./LeBoeuf, R. A. (2002): Rationality. In: Annu. Rev. Psychol. 2002, 53: 491–517

Triantaphyllou, E. (2000): Multi-Criteria Decision Making Methods: A Comparative Study. Boston: Kluwer Academic Publishers

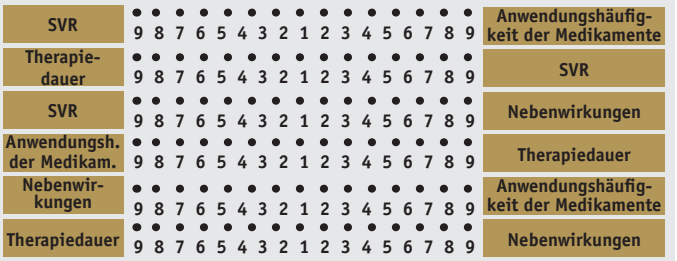
Tversky, A./Kahneman, D. (1974): Judgment under uncertainty: heuristics and biases. In: Science 1074, 185:1124-1131

Wallenius, J.7Dyer, J. S./Fishburn, P. C./Steuer, R. E./Zionts, S./Deb, K. (2008): Multiple criteria decision making, multiattribute utility theory: Recent accomplishments and what lies ahead. In: Management Science 2008, 54, 7: 1336-1349

Zanakis, S. H./Solomon, A./ Wishart, N./ Dublish, S (1998): Multi-attribute decision making: a simulation comparison of select methods. In: European Journal of Operational Research 1998, 107: 507-529

Gewichtung von Aspekten der Hepatitisbehandlung

Jeder Aspekt soll im Folgenden von Ihnen als behandelnder Arzt gewichtet werden, d.h. wenn Sie an die Zielsetzung der Hepatitis-Behandlung ihrer Patienten denken, welchem Aspekt würden Sie eine größere Bedeutung beimessen? „Welcher Aspekt der Hepatitis-Behandlung ist Ihnen wichtiger, und um wie viel wichtiger?“



©MVF

Abb. 2: Paarweise Vergleiche zur Endpunktgewichtung bei Hepatitis C; Legende: 1 = gleich wichtig, 2= liegt dazwischen, 3 = etwas wichtiger, 4 = liegt dazwischen, 5 = viel wichtiger, 6 = liegt dazwischen, 7 = sehr viel wichtiger, 8 = liegt dazwischen, 9 = extrem wichtiger.

CA-Szenarien zur Endpunktgewichtung bei Hepatitis C

	Therapie A	Therapie B
Dauer der antiviralen Therapie	24 Wochen	24 Wochen
Das Medikament (Interferon) muss gespritzt werden	1x alle 14 Tage	1x pro Woche
Anhalten grippeähnlicher Symptome nach der Injektion	an 3 Tagen nach der Inj.	an 1 Tag nach der Inj.
Die Wahrscheinlichkeit, dass Magen-Darm-Beschwerden auftreten, liegt bei	25 von 100 Pers. (25 %)	45 von 100 Pers. (45 %)
Die Wahrscheinlichkeit, dass psychische Nebenwirkungen auftreten, liegt bei	55 von 100 Pers. (55 %)	45 von 100 Pers.(45 %)
Die Wahrscheinlichkeit, dass Hautprobleme und/oder Haarausfall auftreten, liegt bei	55 von 100 Pers. (55 %)	45 von 100 Pers.(45 %)
Die Wahrscheinlichkeit der anhaltenden Virusfreiheit 6 Monate nach Therapieende liegt bei	55 von 100 Pers. (55 %)	45 von 100 Pers. (45 %)
Bitte entscheiden Sie sich für eine Therapie (kreuzen Sie an)	☐	☐

Tab. 3: CA-Szenarien zur Endpunktgewichtung bei Hepatitis C.

von Präferenzen. Als allgemeine Anforderungen an das wissenschaftliche Vorgehen definieren sie die nutzentheoretische Fundierung, die Erfüllung von vorbestimmten Gütekriterien, die Wirtschaftlichkeit und Praktikabilität des Ansatzes und die Vollständigkeit der Bewertung mittels der eingesetzten Methoden. Die Bewertung der Methoden anhand dieses Kriterienkatalogs schlussfolgern die Autoren, konnte zeigen, dass weder CA noch AHP alle Anforderungen erfüllen, die an eine Methode zur Endpunktgewichtung im Rahmen des Effizienzgrenzenansatzes unter Berücksichtigung der gesetzlichen Vorgaben, der IQWiG-Methodik und der allgemeinen Anforderungen an wissenschaftliches Vorgehen zu stellen sind und sie empfehlen die Durchführung weiterer empirischer Vergleichsstudien, die beide Methoden gezielt im Hinblick auf die Fragestellung der Verwertbarkeit zur Endpunktgewichtung im Rahmen der IQWiG-Methodik überprüfen. Wenn man allerdings die bereits vorliegenden und auch hier vorgestellten Vergleichsarbeiten hinsichtlich des vorgestellten Kriterienkatalogs analysiert, gelangt man zur selben Schlussfolgerung. Letztlich müssen die Anwender die Kriterien des vorgestellten Kriterienkatalogs anhand ihrer eigenen Präferenzen priorisieren, um sich für das eine oder andere Verfahren zu entscheiden (Tab. 3).

Patient-weighted endpoints: Conjoint Analysis vs. Analytic Hierarchy Process

Both techniques, Conjoint Analysis (CA) and Analytic Hierarchy Process (AHP), are used to elicit patient preferences (in this case to prioritize and weight endpoints). Beyond their theoretical basis (economical utility theory versus descriptive theory) they differ significantly concerning the underpinned model (additive partial utility model versus versus weighted additive utility model), the methodological approach (decompositional versus compositional), the assessment process (holistic versus pair-wise comparisons), methodological weaknesses (excessive demand of cognition versus unreal decision scenarios) and the resource use (extensive versus low). CA may lead in relation to a priori assumptions to conflicting results. The AHP's drawback is a possible rank reversal. Both techniques are generating in comparing studies quite similar results and show depending on the research question they are used pros and cons respectively.

Keywords

Multicriterial decision analysis, Conjoint Analysis, Analytical Hierarchy Process, Patients' preferences, Endpoint weighting, Benefit-Risk trade off, Benefit aggregation

Ausblick

Multikriterielle Entscheidungsverfahren zur Patientenpräferenzerhebung bei Endpunktgewichtung gewinnen immer mehr an Bedeutung in Gesundheitssystemen. In Deutschland sind sie für die gesundheitsökonomische Evaluation nach dem vom IQWiG vorgeschlagenen Effizienzgrenzenansatz eine Prämisse, um endpunktbezogenen Nutzen zu aggregieren. Das IQWiG nennt in seinem aktuellen Methodenpapier (Version 4.0) CA und AHP auch als eine Möglichkeit, endpunktbezogene Nutzen- und Schadensaspekte gegenüberzustellen (IQWiG 2011).

Die FDA (2012) sieht in ihrer Guidance for Industry and FDA Administration Staff 03/2012 unter der Rubrik „Factors to Consider When Making Benefit-Risk Determinations“ Patientenpräferenzen hinsichtlich der Nutzen- und Schadensaspekte von medizinischen Anwendungen als nützliche Informationsquelle. Aber auch die EMA (2010) hat den Stellenwert der Ermittlung von Patientenpräferenzen durch multikriterielle Entscheidungsanalyseverfahren zur Nutzen-Schaden-Abwägung in ihrem Benefit-Risk Methodology Project erkannt. Es bleibt abzuwarten, ob auch Entscheidungsträger wie der G-BA sich Methoden wie die CA und das AHP bei ihren Beschlussfassungen zum Zusatznutzen von neuen Arzneimitteln im Rahmen der frühen Nutzenbewertung für die Saldierung von Nutzen- und Schadensaspekten anstelle intuitiver Heuristiken bedienen werden. Patientenrelevante Endpunkte, die den Nutzen und Schaden für die betroffenen Patienten abbilden, sollten zur Herleitung einer Gesamtaussage hinsichtlich eines vorliegenden Zusatznutzens in Zeiten von Shared-decision making auf der Arzt-Patienten-Entscheidungsebene von den betroffenen Patienten auch auf Systemebene gewichtet werden. Dies verlangt auch, dass das System dem mündigen Patienten Vertrauen schenkt, und zwar nicht nur auf der rein rhetorisch-verbalen Ebene, sondern auch im Rahmen der unmittelbar die betroffenen Patienten tangierenden Entscheidungsfindung. <<

Dr. rer. biol. hum. Ch.-Markos Dintsios, MPH

ist seit 2011 Senior Referent für Gesundheitsökonomie beim Verband der forschenden Arzneimittelhersteller (vfa) in Berlin. Der Pharmazeut, Wirtschaftswissenschaftler und Master in Public Health war 1994-1997 Patientenvertreter am Klinikum Nürnberg, bis 2005 wissenschaftlicher Mitarbeiter am Institut für Epidemiologie, Sozialmedizin & Gesundheitssystemforschung der Medizinischen Hochschule Hannover und von 2005-2011 beim IQWiG tätig. Kontakt: c.dintsios@vfa.de

