Multikriterielle Entscheidungsfindung im deutschen Gesundheitswesen

Von der Wirtschaftswissenschaftlichen Fakultät der
Gottfried Wilhelm Leibniz Universität Hannover
zur Erlangung des akademischen Grades

Doktorin der Wirtschaftswissenschaften
– Doctor rerum politicarum –

genehmigte Dissertation

von

M.Sc. Katharina Schmidt
geboren am 27.09.1988 in Hamburg

2018

Referent: Prof. Dr. J.-Matthias Graf von der Schulenburg

Korreferent: Prof. Dr. Wolfgang Greiner

Tag der Promotion: 20. Dezember 2017

Zusammenfassung

Die multikriterielle Entscheidungsfindung (Multi Criteria Decision Making, kurz MCDM) kann zur transparenten Erörterung eines Entscheidungsproblems beitragen und wird in den Bereichen Umwelt und Energie bereits angewendet. In Deutschland wird die MCDM derzeit im Rahmen der frühen Nutzenbewertung durch den Gemeinsamen Bundesausschuss (GBA) diskutiert, jedoch bisher nicht in den Entscheidungsfindungsprozess integriert. Aufgrund der Zunahme von chronischen Erkrankungen in der Bevölkerung in den vergangenen Jahren rücken für die Bewertung von Medikamenten oder Therapien neben Mortalität und klinischen Parametern vermehrt patientenrelevante Endpunkte, wie z.B. die gesundheitsbezogene Lebensqualität (Health Related Quality of Life, kurz HRQoL), in den Fokus. Ein direkter Vergleich dieser unterschiedlichen Endpunkte ist jedoch nicht möglich, sodass Gewichtungsmethoden etabliert werden müssen, um eine finale Bewertung der vorliegenden Alternativen erhalten zu können.

Diese Dissertation prüft in acht Modulen die Übertragbarkeit der MCDM auf gesundheitsökonomische Fragestellungen. Hierbei werden zunächst Methoden der qualitativen und quantitativen Forschung identifiziert, anhand derer relevante Eigenschaften (Attribute oder Kriterien) der Entscheidungsprobleme ermittelt werden können. Dabei ist eine systematische Literaturrecherche in der Vorstudie von Präferenzmessungen als unverzichtbar anzusehen. Außerdem wird herausgestellt, dass sich Likert-Skalen Bewertungen und auch der Analytische Hierarchieprozess (Analytic Hierarchy Prozess, kurz AHP) zur Identifikation relevanter Attribute eignen. Des Weiteren wird das Discrete Choice Experiment (DCE) und die Willingness to pay (WTP) Methode im gesundheitsökonomischen Kontext angewendet und ihre Umsetzbarkeit bei der Entscheidungsfindung in diesem Bereich demonstriert. Zuletzt werden anhand dreier Module methodische Unsicherheiten und Herausforderungen bei der Anwendung der MCDM in der Gesundheitsökonomie herausgearbeitet. Die Ergebnisse weisen darauf hin, dass der AHP als Alternative zum DCE aufgrund bisher mangelnder methodischer Standards nachranging behandelt wurde. Es kann allerdings aus den Modulen geschlussfolgert werden, dass je nach Fragestellung und Anwendungsfeld sowohl das DCE als auch der AHP geeignet sind, wenn methodische Ausgestaltungen begründet berichtet werden.

Diese Arbeit zeigt, dass die MCDM dazu beitragen kann, transparente Entscheidungen sowohl in der Versorgungsforschung als auch bei Arzneimittelbewertungsverfahren zu treffen. Perspektivisch sind für den AHP standardisierte Leitlinien gefordert, wie sie für das DCE bereits vorliegen. Weitere MCDM Projekte sollten zukünftig Entscheidungen im

Gesundheitswesen informieren, indem sie eine Studiendurchführung und –auswertung basierend auf etablierten Standards und unter Offenlegung des Vorgehens erproben, berichten und etablieren.

Schlagwörter

Multikriterielle Entscheidungsfindung, Präferenzmessung, Discrete Choice Experiment, Analytic Hierarchy Process, Willingness to pay

Abstract

Multi criteria decision making (MCDM), with its potential to improve the transparency of decision processes, has become well established in environmental and energy policy. In Germany, the Federal Joint Committee discusses MCDM in the context of early benefit assessments, although it has not been formally included in decision processes so far. Due to the increasing prevalence of chronic diseases over the past years, not only mortality and clinical parameters, but also patient relevant outcomes (e.g., health-related quality of life) have become important factors in the evaluation of pharmaceuticals. Since these outcomes cannot be compared directly, weighting methods can be employed to support the final evaluation of the available alternatives.

The present doctoral thesis examined the applicability of MCDM in health economics decision making in eight modules. Specifically, we tested qualitative and quantitative methods to identify the relevant attributes of a decision problem. Initially, we showed that a systematic literature review is crucial in the preliminary study phase. Additionally, we pointed out how the quantitative Likert-scale method and the analytic hierarchy process (AHP) could be used for attribute identification. In the following step, we conducted discrete choice experiments (DCE) and willingness to pay (WTP) analyses in health economics settings. Both methods demonstrated their applicability in decision making in the context considered. Then, we analyzed in three modules the methodological uncertainties of MCDM in health economics applications. The results showed that AHP has received less attention due to the lack of methodological standards as compared to DCE. However, once the choices regarding potential methodological issues are explicitly reported, both methods could be used in accordance with the research question and the field of application.

Overall, our work showed that MCDM can support transparent decision making in health economics, especially in health care research and drug assessment. In the future, standardized guidelines are needed for the AHP method, like the ones established for the DCE method. Moreover, an increase in the number of evidence-based MCDM projects carried on would enhance decision making. In this sense, transparent reporting and publication of these studies for testing and establishing MCDM methods in health economics become important.

Key words

Multi Criteria Decision Making, Preferences, Discrete Choice Experiment, Analytic Hierarchy Process, Willingness to Pay

Inhaltsverzeichnis

1 Motivation und Zielsetzung

Während politische Entschlüsse im Bereich Umwelt und Energie bereits seit vielen Jahren durch die Methoden der multikriteriellen Entscheidungsfindung (Multi Criteria Decision Making, kurz MCDM) unterstützt werden [1, 2], hat nun auch die Gesundheitsökonomie das Potential dieser Methoden entdeckt [3]. Bei der MCDM wird ein Entscheidungsproblem durch alle seine relevanten Eigenschaften (qualitativ oder quantitativ) beschrieben und diese im Anschluss gegeneinander abgewogen [3]. Hierbei bezeichnet MCDM sowohl den Vorgang der Entscheidungsunterstützung als auch zusammenfassend die Methoden [4].

Diese Dissertation untersucht die Anwendung der MCDM im gesundheitsökonomischen Kontext und eruiert hierbei die folgenden Fragen:

1. *Wie können relevante Eigenschaften des Entscheidungsproblems identifiziert werden und welche (qualitative oder quantitative) Methoden eignen sich hierfür?*
2. *Sind die Methoden der MCDM für Entscheidungen im gesundheitsökonomischen Kontext geeignet?*
3. *Welche Herausforderungen oder Unsicherheiten ergeben sich bei der Anwendung in Bezug auf die Methoden?*

Zunächst wurden Methoden der MCDM im Rahmen von Beschaffungsentscheidungen im Krankenhausbereich und zur Ressourcenallokation im Gesundheitswesen genutzt [5, 6] und später zunehmend bei klinischen Diagnose- oder Therapieentscheidungen und Health Technology Assessments [7]. In Deutschland wird die MCDA derzeit im Rahmen der frühen Nutzenbewertung durch den Gemeinsamen Bundesausschuss (GBA) erstmals in die Diskussion eingebracht [8, 9], jedoch bisher nicht in den Entscheidungsfindungsprozess integriert. Durch die 2011 mit dem Arzneimittelneuordnungsgesetz (AMNOG) eingeführte Bestimmung des (Zusatz-)Nutzens bei Markteintritt neuer Arzneimittel liegt hier ein klassisches Entscheidungsproblem für eine Gewichtung unterschiedlicher Kriterien vor. Neben Effekten wie Mortalität und die Messung von klinischen Parametern rücken auch die

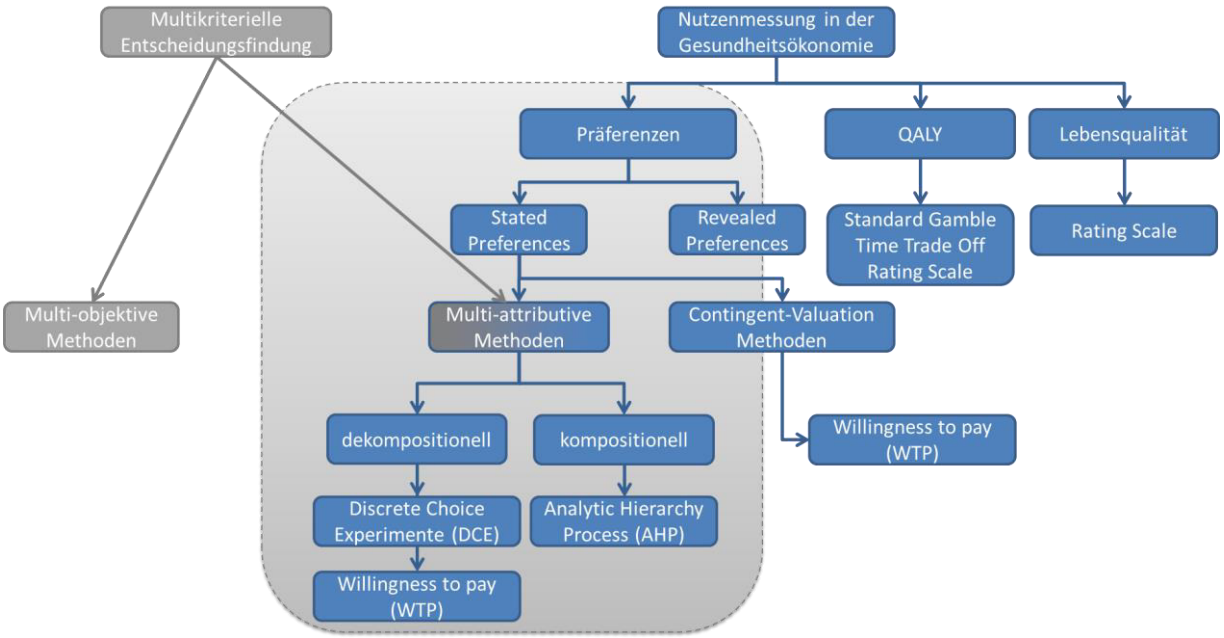
patientenrelevanten Outcomes, wie z.B. die gesundheitsbezogene Lebensqualität (Health Related Quality of Life, kurz HRQoL), für die Beurteilung des Arzneimittelnutzens in den Fokus. Besonders bei chronischen Erkrankungen, die in den vergangenen Jahren zunehmend das Krankheitsbild der deutschen Bevölkerung bestimmen [10], kann keine Heilung durch die Therapie erreicht werden. Somit werden bei Therapieentscheidungen andere Aspekte relevant, wie beispielsweise die Schmerzreduktion, die Aufrechterhaltung von Mobilität und Teilhabe am sozialen Leben. Diese individuellen Patientenpräferenzen gilt es zu erheben und idealerweise im Rahmen der Therapiemöglichkeiten umzusetzen. Die Schwierigkeit hierbei ist, dass die Messung der einzelnen Kriterien in unterschiedlichen Einheiten erfolgt. Die zusätzlich gewonnene Überlebenszeit durch ein bestimmtes Medikament kann beispielsweise in Monaten oder Jahren gemessen werden. Dahingegen kann die HRQoL nicht direkt gemessen werden, weshalb hier validierte Fragebögen zum Einsatz kommen. Bei der Messung der HRQoL gibt es mittlerweile zahlreiche etablierte Instrumente, bei denen die Ergebnisse auf Basis repräsentativer Bevölkerungsbefragungen mit Referenzwerten des jeweiligen Landes unterlegt werden können und somit international vergleichbare qualitätsadjustierte Lebensjahre (Quality Adjusted Life Years, kurz QALY) kalkuliert werden können [11, 12]. Für weitere Endpunkte, wie zum Beispiel Schmerzen, können die Studienergebnisse nicht direkt in ein Verhältnis zum Überleben gesetzt werden, sondern bilden qualitative Kriterien. Es muss also eine Umrechnungsformel gefunden werden, wenn qualitative und quantitative Kriterien nachvollziehbar in ein Verhältnis gesetzt werden sollen. Dazu könnte zukünftig die MCDM beitragen.

Die Gewichtung von Kriterien unterschiedlicher Messniveaus kann anhand eines Beispiels aus dem Sport verdeutlicht werden. In der Leichtathletik ist der Zehnkampf die Königsdisziplin, da hier die unterschiedlichsten Anforderungen in den Bereichen Schnelligkeit, Kraft und Ausdauer abverlangt werden. Um einen Gesamtsieger ermitteln zu können, reicht es nicht aus, in einer einzelnen Disziplin der Sieger zu sein. Daher wurde hier im Jahr 1984 ein Punktesystem entwickelt, um die Leistungen vergleichbar zu machen und eine Gesamtpunktzahl für jeden Athleten ermitteln zu können [13]. Dieses System sorgt

dafür, dass sowohl Weiten (Weitsprung, Kugelstoßen, Diskuswurf, Speerwurf), Höhen (Hochsprung, Stabhochsprung) als auch Zeiten (100m, 110m Hürden, 400m, 1.500m) miteinander verglichen werden können. Ein ähnliches Vorgehen wird bei der MDCD angewendet, da auch hier aus unterschiedlichen Kriterien eine „Gesamtpunktzahl“ für die Gesamtbewertung gefunden werden muss. Für diese Gewichtung mehrerer Kriterien können Methoden der Präferenzmessung oder Prioritätenschätzung genutzt werden.

Die Methodik der MCDM lässt sich in multi-objektive (Multi-Objective Decision Making, kurz MODM) und multi-attributive Entscheidungsanalysen (Multi-Attribute Decision Making, kurz MADM) unterteilen [14]. Bei ersteren besteht – im Gegensatz zur MADM – keine feste Anzahl an Alternativen, die verglichen werden. Da im Gesundheitswesen in den meisten Fällen eine bestimmte Anzahl an Alternativen verglichen bzw. evaluiert werden sollen, liegt der Fokus dieser Arbeit auf den MADM. Parallel dazu gibt es auch in der gesundheitsökonomischen Präferenzmessung den Begriff der multi-attributiven Methoden (vgl. Abbildung 1). In der neoklassischen Gesundheitsökonomie sind Präferenzen „das Ergebnis der relativen subjektiven Bewertung von Alternativen durch die Abwägung der Kosten und des Nutzens in einem Entscheidungs- und Bewertungsprozesses“ [15], S 160. In der Präferenzmessung fallen unter die multi-attributiven Methoden lediglich die kompositionellen und dekompositionellen Methoden, während bei der entscheidungsanalytischen Betrachtung auch noch multikriteriell betrachtete „revealed preferences“ und „Contigent-Valuation“ Methoden zusammengefasst werden können. Dabei stehen die „revealed preferences“ für aufgedeckte Präferenzen, die sich durch den tatsächlichen Kauf eines Produktes oder der Inanspruchnahme einer Dienstleistung ausdrücken. Die Ermittlung dieser Art von Präferenzen ist indirekt, da nur die Folge beobachtet werden kann, jedoch nicht die Gründe für diese Entscheidung. Dahingegen stellen die „stated preferences“ eine direkte Form der Präferenzmessung dar, indem hier hypothetische Produkte oder Dienstleitung direkt bewertet werden.

Abbildung 1: Methodenübersicht



Quelle: eigene Darstellung nach [16, 17]

Bei den multi-attributiven Methoden der Präferenzmessung werden (in Einklang mit den MADM) die Produkte oder Dienstleistungen anhand ihrer Eigenschaften (Attribute oder Kriterien) und Ausprägungen (Level) bewertet. Hierbei spielt es eine Rolle, ob die Präferenzermittlung über das gesamte Produkt getroffen und daraus die Teilnutzen ermittelt werden (dekompositionell) oder ob die Teilnutzen der einzelnen Eigenschaften ermittelt werden, die dann zum Gesamtnutzen aggregiert werden (kompositionell). Ein Beispiel für die dekompositionelle Nutzenbewertung ist das Discrete Choice Experiment (DCE), das die angewendete Methode in den Modulen 4 und 5 darstellt. Ein Beispiel für die direkte Nutzenmessung der einzelnen Eigenschaften eines Produktes ist der Analytische Hierarchieprozess (Analytic Hierarchy Process, kurz AHP). In der Literatur wird darauf hingewiesen, dass im AHP ermittelte Prioritäten im engeren Sinne nicht als Präferenzen angesehen werden dürfen, da sie nicht auf der Erwartungsnutzentheorie basieren [18]. Allerdings findet in der Literatur kein konsistenter Umgang mit den Begrifflichkeiten statt. Ebenso sollte darauf hingewiesen werden, dass sämtliche in Abbildung 1 dargestellten

Begriffe nicht trennscharf abgrenzbar sind und in der Literatur uneinheitlich verwendet werden. Im Folgenden bezieht sich diese Arbeit auf den enger gefassten Begriff aus der Präferenzmessung.

Im Folgenden wird kurz auf den Ablauf einer multi-attributiven Studie im Allgemeinen eingegangen, da sich die Module dieser Dissertation ebenfalls an den Prozessschritten orientieren (siehe Abbildung 2). Zuerst wird die Problemdefinition vorgenommen und das Ziel der Studie festgelegt. Hierbei ist es ebenfalls relevant, wer das Entscheidungsproblem beantworten soll. Dieser Aspekt ist in Bezug auf das Forschungsziel zu sehen [19] und kann unterschiedliche Sichtweisen berücksichtigen und vergleichen (z.B. Experten, Patienten, Angehörige, Allgemeinbevölkerung).

Abbildung 2: Studienablauf Präferenzmessung

Quelle: Eigene Darstellung in Anlehnung an [4, 20, 19]

In der Vorstudie wird das Entscheidungsproblem final strukturiert, indem Attribute oder Kriterien näher definiert und mögliche Level festgelegt werden (qualitativ, quantitativ, literaturbasiert) [19]. Für die Identifikation von Attributen gibt es bisher keinen Goldstandard,

sodass in Studien unterschiedliche Methoden eingesetzt werden. Außerdem sollte in diesem Studienabschnitt das experimentelle Design festgelegt werden. Im nächsten Schritt der Hauptstudie wird daraufhin ein Fragebogen erstellt und die Befragung durchgeführt. Die MCDM lässt sich in diesem Schritt in die Messung der Performance und in die Gewichtung der Eigenschaften einteilen. Die Performance bezieht sich auf die Bewertung der Alternativen anhand der Kriterien, während sich die Gewichtung auf die Präferenzen der Kriterien im Vergleich zueinander bezieht [4]. Anhand eines Beispiels können die beiden Begriffe wie folgt differenziert werden: Die Performance ist die Reduktion der empfundenen Schmerzen für die Medikamente A und B; die Gewichtung erfolgt über die Beurteilung der Schmerzreduktion im Vergleich zu der gewonnen Überlebenszeit (unabhängig vom Medikament). Bei reinen Präferenzstudien werden in den meisten Fällen nur die Gewichtungen evaluiert und die Performance aus bereits vorhandenen (klinischen) Erhebungen ergänzt. Bei der Auswertung können abschließend eine Vielzahl von Analysen durchgeführt werden, die sich aus der Datenverteilung ergeben. Dabei sollten jedoch die Ergebnisse auf Verzerrungen und ihre Repräsentativität überprüft werden (Sensitivitätsanalysen) [4].

Ziel dieser Dissertation ist es, unterschiedliche Methoden zur Präferenzmessung gegenüberzustellen und begründete Einschätzungen zur Anwendbarkeit in der Gesundheitsökonomie zu erörtern. Dabei werden unterschiedliche Planungsschritte detailliert betrachtet (Kapitel 2.1: Identifikation von relevanten Entscheidungskriterien; Kapitel 2.2: Studiendurchführung in unterschiedlichen Anwendungsfeldern) und Vor- und Nachteile abgeleitet. Des Weiteren werden aktuelle methodische Herausforderungen bei der Studiendurchführung des AHP betrachtet (Kapitel 2.3). In Kapitel 3 werden Schlussfolgerungen gezogen werden, wie bestehende Gesundheitsleistungen oder -produkte „präferenzgerecht“ unter Anwendung der MCDM Methoden optimiert werden können, um knappe Ressourcen im Gesundheitswesen effizienter verteilen zu können.

2 Beitrag der vorliegenden kumulativen Dissertationsarbeit

2.1 Identifikation von relevanten Entscheidungskriterien

Diese kumulative Dissertation befasst sich mit unterschiedlichen Methoden der MCDM, mit dem Schwerpunkt auf den multiattributiven Messungen mittels DCE und AHP. Zunächst wird in diesem Rahmen die Wahl und Durchführung von Vorstudien anhand der Module 1 bis 3 betrachtet.

Sämtliche Methoden der Präferenzmessung basieren im ersten Schritt auf der Identifikation relevanter Eigenschaften, die den Forschungsgegenstand instrumentalisieren. In Ermangelung eines Goldstandards für die Identifikation der Attribute wurde dieser essentielle Schritt in der Vergangenheit häufig vernachlässigt [21, 22]. Im Rahmen des Center for Health Economics Research (CHERH) konnte die Förderung für ein vierjähriges Projekt zu der Ermittlung der Präferenzen von Lungen- und Darmkrebspatienten in Bezug auf ihre Therapie erreicht werden. In diesem Projekt sollten die Patienten abwägen, welche Nebenwirkungen und Wirkungen der Chemotherapie sie präferieren.

In Vorbereitung dieser Studie wurde eine systematische Literaturrecherche durchgeführt, um den Stand der Forschung zu den Präferenzen bezüglich der Therapie von Lungenkrebs abbilden zu können und relevante Attribute für das anschließende DCE zu identifizieren (Modul 1: „Preferences of lung cancer patients for treatment and decision-making: a systematic literature review“). Es konnten sowohl qualitative als auch quantitative Studiendesigns ermittelt werden, die aus Sicht unterschiedlicher Akteure (Patienten, Ärzte, Gesunde) und Subgruppen (Frauen, ältere Patienten, nach Krankheitsstadien) die Präferenzen erhoben. Aus den Studien ergaben sich konträre Einstellungen der befragten Akteure zur Wichtigkeit von Lebenszeitverlängerung in Abwägung zur HRQoL während der Therapie. Des Weiteren untersuchte dieses Review, ob und wie die Patienten in die Entscheidungsfindung bezüglich ihrer Therapie einbezogen wurden. Es zeigte sich, dass die Lungenkrebspatienten in den identifizierten Studien tendenziell eine passive Rolle bei der Therapieentscheidung einnehmen wollten. Allerdings beruhten diese Erkenntnisse auf

lediglich vier Studien, die teilweise kleine Patientenfallzahlen einschlossen. Es stellt sich nun die Frage, warum die Krebspatienten nicht aktiv an ihrer Therapiegestaltung teilnehmen wollten, obwohl sich ein genereller Trend in der Gesundheitsversorgung hin zur informierten Entscheidung oder gemeinsamen Entscheidungsfindung (Shared Decision Making, kurz SDM) zeigt. Dieser Frage wurde unter anderem in Modul 4 nachgegangen.

Häufig werden bei quantitativen Erhebungen die sogenannten Likert- oder Fünferskalen zur Bewertung von Eigenschaften unterschiedlicher Konstrukte genutzt. Hierbei ist es jedoch von großer Bedeutung, dass keine Abwägungsentscheidung (Trade-Off Entscheidung) stattfindet, sondern eine losgelöste Bewertung einer individuellen Eigenschaft. Derartige Entscheidungen finden sich in Modul 2, in dem Komponenten eines pneumologischen Rehabilitationsprogrammes von Teilnehmenden auf einer elfstufigen Skala (0 bis 10) bewertet werden sollten. Im Vordergrund der Analysen standen hier Einflussfaktoren auf die Bewertungen der Wichtigkeit einzelner Rehabilitationskomponenten. Da diese Daten nicht normalverteilt vorlagen, wurde aufgrund der vorliegenden Verteilung ein Beta-Regressionsmodell gewählt. Außerdem wurden Faktorenanalysen (Hauptkomponentenanalyse mit Varimaxrotation) durchgeführt, um eine mögliche Aggregation von mehreren Rehabilitationskomponenten zu einer neuen Rehabilitationskategorie zu prüfen. In der Faktorenanalyse zeigte sich, dass aufgrund der Ladungen vier der Rehabilitationskomponenten zu einer neuen Rehabilitationskategorie (den Standortfaktoren) aggregiert werden konnten. Während sich im Gesamtüberblick über alle Eigenschaften von Rehabilitationsprogrammen nur geringe Unterschiede in der Wichtigkeit herausstellten (zwischen 7 und 9 Punkten im Mittel), konnten hingegen unterschiedliche Einflussfaktoren auf die Bewertungen identifiziert werden. Die Ergebnisse dieser Analysen wurden im Folgenden genutzt, um einerseits die Überlappung von Attributen für das DCE zu verhindern und andererseits wichtige soziodemografische und krankheitsspezifische Informationen in einen zukünftigen DCE-Fragebogen zu integrieren. Diese Arbeit sollte als Vorstudie eines weiteren DCEs zu den Präferenzen von Rehabilitanden in Bezug auf die Eigenschaften eines hypothetischen Rehabilitationsprogrammes dienen, welche zum

jetzigen Zeitpunkt noch nicht abgeschlossen ist. Die Relevanz dieses Moduls ergibt sich somit durch die Nutzung alternativer Methoden (in diesem Falle Faktorenanalysen und Beta-Regressionen bei Likert-Skalen-Bewertungen) zur Identifikation von Attributen für DCEs und die transparente Darstellung einer Vorstudie.

Eine weitere innovative Methode, mit der relevante Eigenschaften des Entscheidungsproblems erörtert werden können, wurde In Modul 3 verwendet. Unter Berücksichtigung der Erkenntnisse aus den methodischen Modulen (Module 6 bis 8) wurde der AHP in einer Vorstudie zu den Standortfaktoren von ambulanten Hausärzten in Niedersachsen angewendet. Diese Publikation untersuchte in einer zweiteiligen Erhebung die Prioritäten bei möglichen Standortfaktoren mittels AHP und einer quantitativen Befragung bei Stadt- und Landärzten. Hierbei wurde die AHP-Methodik für die Identifikation von Standortfaktoren genutzt, die im zweiten Studienteil anhand einer Querschnittsbefragung näher untersucht wurden. Die zahlreichen aus der Literatur identifizierten Standortfaktoren wurden mittels AHP in individuellen persönlichen Interviews gewichtet. Es zeigte sich, dass der AHP auch in individuellen Interviews sehr gut durchführbar war. Durch eine anschließende quantitative Querschnittsbefragung konnten Unterschiede in der Wichtigkeit von Standortfaktoren zwischen Land- und Stadtärzten belegt werden. Die Eignung der AHP-Methodik als Vorstudie ist somit gegeben und konnte die Fragen der Hauptstudie auf relevante Kriterien reduzieren. Somit konnte der Schwerpunkt auf die im AHP identifizierten Kriterien gleichzeitig begründet und belegt sowie der Aufwand für Studienteilnehmer reduziert werden.

2.2 Studiendurchführung in unterschiedlichen Anwendungsfeldern

Die Erkenntnisse zu den relevanten Eigenschaften aus den Vorstudien wurden im weiteren Verlauf für die Hauptbefragungen verwendet. In Modul 4 („Therapy Preferences of Patients with Lung and Colon Cancer: A Discrete Choice Experiment“) wurden anhand eines DCEs die Präferenzen bezüglich der Chemotherapie von Lungen- und Darmkrebspatienten identifiziert. Außerdem wurden in dem Fragebogen soziodemographische und

krankheitsbezogene Charakteristika abgefragt, um subgruppenspezifische Analysen durchführen zu können. In den multivariaten Berechnungen wurden Conditional Logit Modelle (CLM), Generalized Linear Mixed Logit Modelle (GLMM) sowie Latent Class Mixed Logit Modelle (LCMLM) durchgeführt. Letztere sollten dazu genutzt werden, datenbasiert Klassen anhand bestimmter Charakteristika und Präferenzen herausstellen zu können. Durch die Berechnungen zu den LCMLMs wurden drei Patientenklassen ermittelt, die sich in ihren Präferenzen bezüglich der Therapie unterschieden. Die Ergebnisse wiesen darauf hin, dass nicht die soziodemografischen Charakteristika der Patienten die Therapiepräferenzen beeinflussten, sondern die krankheitsspezifischen Merkmale und die Information durch den behandelnden Arzt. Die Studienresultate ließen die Schlussfolgerung zu, dass die ärztliche Beratung die Präferenzen hin zu einem Abwägen von Lebenszeit und HRQoL beeinflussten. Thematisch konnte diese Publikation somit einen Wissenszuwachs für Ärzte und die Versorgungsforschung bieten, da bisher subgruppenspezifische Merkmale von Patienten in Bezug auf die Präferenzen nicht vorlagen. Gleichzeitig wurde mit dieser Publikation die Methode des DCE weiterentwickelt, weil in der Vergangenheit hauptsächlich CLM durchgeführt und somit kein Schwerpunkt auf latente Klassenunterschiede für die Präferenzen gelegt wurden.

Da bisher lediglich die Nutzenseite der Präferenzmessung im Vordergrund stand, wurde ein Modul eingeschlossen, das zusätzlich die Kostenseite betrachtet. Modul 5 zeigt eine spezifische Ausgestaltung des DCE, welches ein Attribut zur Zahlungsbereitschaft berücksichtigte. Diese Arbeit beschäftigte sich mit den Präferenzen von Testeigenschaften zur Ganzgenomsequenzierung. Ein Attribut bezeichnete die Testkosten, die mit 500 Euro, 1.000 Euro oder 1.500 Euro veranschlagt wurden. Diese Testkosten basierten auf einer Literaturrecherche und waren als realistisch einzuschätzen. Aufgrund der trade-off Entscheidungen ließen sich einerseits die Präferenzen für das Attribut der Testkosten ermitteln und andererseits Abwägungsentscheidungen zwischen den Testkosten und den übrigen Attributen abschätzen. Somit konnten die Präferenzen der weiteren Attribute in

Geldeinheiten umgerechnet werden. Ziel hierbei war es, eine vergleichbare Einheit zu wählen, sodass sich die Studienergebnisse auch extern vergleichen bzw. messen ließen.

2.3 Methodische Aspekte bei der Studiendurchführung

Der AHP ist eine anerkannte Methode der MCDM, hat jedoch in der Gesundheitsökonomie bisher eine geringe Aufmerksamkeit erfahren. Die folgenden Module untersuchten deshalb, in welchen Anwendungsbereichen bereits Publikationen zu dieser Methodik vorhanden sind und welche methodischen Ausgestaltungen vorlagen. Ziel der Module 6 bis 8 war es, Schlussfolgerungen zum Einsatz des AHP als Alternative zum DCE bei Entscheidungen im Gesundheitswesen ziehen zu können.

In Modul 6 wurde eine systematische Literaturrecherche zur Anwendung des AHP in der Versorgungsforschung durchgeführt. Eine systematische Analyse dieses Feldes lag lediglich für den Zeitraum bis 2011 durch eine Arbeit von Hummel und Ijzerman vor [23], sodass eine Aktualisierung für die Entwicklung und den aktuellen Stand der Forschung von Interesse war. Die Betrachtung der eingeschlossenen Publikationen im Zeitverlauf zeigte, dass – obwohl die erste Publikation eines AHP im Gesundheitskontext bereits 1981 veröffentlicht wurde – ein Anstieg der Veröffentlichungen erst in den vergangenen 10 Jahren erfolgte. In den Jahren 2011 und 2012 lagen 9 Publikationen pro Jahr vor, bis hin zu 20 Veröffentlichungen im Jahr 2015 bis zum Oktober. Hier werden die Relevanz der Aktualisierung der systematischen Literaturrecherche und ebenfalls die zunehmende Bedeutung des AHP deutlich. Des Weiteren wurde eine Qualitätsbewertung der vorliegenden Artikel durchgeführt, die sich auf die Vollständigkeit der Berichterstattung in den vorliegenden AHP-Artikeln bezog. Infolgedessen kann aus Modul 6 abgeleitet werden, dass ein inkonsistenter Einsatz der Methodik in Bezug auf Aggregation der Einzelbewertungen, Kalkulation der benötigten Studienteilnehmer, dem Umgang mit inkonsistentem Antwortverhalten und der Durchführung von Sensitivitätsanalysen vorlag. Die Qualitätsbewertung der Studien ergab, dass ein Großteil der Studien nicht über alle relevanten Aspekte bei der Durchführung eines AHP berichtete. Die Studienqualität verbesserte sich in den letzten drei Jahren nicht, was die

dringend notwendige Festlegung und Etablierung von Qualitätsstandards und Leitlinien für die AHP-Durchführung offenbarte.

Resultierend aus den Ergebnissen des Moduls 6 wurde ein AHP durchgeführt, das sich mit der Wichtigkeit von Informationskriterien für seltene Erkrankungen beschäftigte (Modul 7: „Measuring patients' priorities using the Analytic Hierarchy Process in comparison with Best-Worst-Scaling and rating cards: methodological aspects and ranking tasks"). Bei diesem Modul handelt es sich um einen methodischen Vergleich der etablierten BWS- und Ranking Card-Verfahren mit dem AHP. Vorherige Untersuchungen bezogen sich auf einen Vergleich von DCE und AHP bzw. DCE und BWS, sodass hier aufgrund der zunehmenden Bedeutung der AHP-Methodik ein Nachholbedarf bestand. Zudem wurden Unsicherheiten im Umgang mit der AHP-Methodik geprüft, die aus Modul 6 resultierten. Modul 7 trägt dazu bei, den Appell nach einer einheitlichen Methodik beim AHP zu unterstützen. Dies bedeutet im Detail, dass geometrische Mittelwerte genutzt werden sollten und die Teilnehmenden je nach Forschungsfrage als eine Einheit (Aggregation of individual judgments, AIJ) oder als eigenständig Bewertende (Aggregation of individual priorities, AIP) angesehen werden und sich die Aggregationsmethode daran orientieren sollte. Zuletzt wurde ein Methodenvergleich vollzogen, bei dem die Bewertungen der Informationskriterien mittels AHP, BWS und Ranking Cards gegenübergestellt wurden. Es konnten moderate bis starke Korrelationen zwischen AHP und BWS festgestellt werden, sodass dadurch Hinweise auf die Validität der AHP-Methodik abgeleitet werden konnten.

Ein weiteres Modul untersuchte im selben Projekt wie Modul 7 die Effekte von individuellen Entscheidungen gegenüber Gruppenentscheidungen beim AHP (Modul 8). Der AHP wurde zunächst für die Konsensfindung in einer Gruppe entwickelt, wurde später jedoch auch für individuelle Entscheider eingesetzt, deren Einzelergebnisse im Anschluss zu einem Gruppenergebnisse aggregiert wurden. Hierbei stellte sich die Frage, ob der Prozess der Konsensfindung einen Unterschied in den Ergebnissen im Vergleich zu aggregierten Einzelentscheidungen hervorruft. In Bezug auf die Konsensfindung bei der

Gruppenentscheidung konnte herausgefunden werden, dass seltener Extremwerte gewählt wurden als bei Einzelentscheidungen und die Ergebnisse signifikant abwichen. Außerdem wurde untersucht, wie sich die Festlegung des Konsistenz-Ratios (Consistency-Ratio, kurz CR) auf allen Ebenen der Hierarchie im Vergleich zum gleichen CR lediglich auf der zweiten Ebene auswirkte. Die CR gibt hierbei an, wie konsistent Studienteilnehmer über mehrere Fragen hinweg geantwortet haben. Durch die Festlegung des akzeptierten CR Levels konnten die Ergebnisse beeinflusst werden, da Personen mit inkonsistentem Antwortverhalten (CR über 0,2 oder 0,1) ausgeschlossen werden mussten. Die Anzahl der auszuschließenden Personen variierte hierbei sehr stark, je nachdem welcher CR herangezogen wurde. Hierdurch entstand ein hohes Verzerrungspotenzial.

3 Zusammenfassung der Ergebnisse und Ausblick auf den weiteren Forschungsbedarf

Die Durchführung und Publikation von Studien im Bereich der MCDM unterliegt methodischen Herausforderungen, die mit dieser Dissertation aufgezeigt werden konnten. Gleichzeitig wird deutlich, dass Aspekte der Transparenz im Bereich der Studiendurchführung und Ergebnisdarstellung umgesetzt werden können. Da es bisher keinen etablierten Goldstandard bei dieser Art von Studien gibt, leistet die vorliegende Dissertation hierfür einen wichtigen Beitrag. Im ersten Schritt ist besonders die fundierte Auswahl einer Vorstudie von hoher Relevanz, um die Identifikation von geeigneten Attributen oder Kriterien gewährleisten zu können.

Somit können bezogen auf die erste untersuchte Forschungsfrage

1. *Wie können relevante Eigenschaften des Entscheidungsproblems identifiziert werden und welche (qualitative oder quantitative) Methoden eignen sich hierfür?*

die folgenden Antworten abgeleitet werden: Zur transparenten Darstellung der Attributsfindung für Präferenzstudien bieten sich unterschiedliche qualitative oder quantitative Methoden an. Aus Modul 1 (und im methodischen Kontext auch Modul 6) konnte

die Erkenntnis gewonnen werden, dass eine systematische Literaturrecherche in der Vorstudie von Präferenzstudien als unverzichtbar anzusehen ist, da auf diese Weise zunächst der Stand der Forschung herausgearbeitet wird. Dabei ist vor allem die Literatur aus der nahen Vergangenheit (fünf bis zehn Jahre) von Relevanz. Mit einer tabellarischen Aufbereitung der Ergebnisse können darüber hinaus ein Überblick über die Qualität und mögliche Limitationen bisheriger Studien identifiziert werden.

Als Vorstudie sind neben den in der Übersichtsarbeit untersuchten Studiendesigns auch Faktorenanalysen und Likert-Skalen Bewertungen geeignet; dennoch lässt sich anhand des Moduls 2 erkennen, dass Bewertungen anhand von Likert-Skalen bei multikriteriellen Entscheidungen nicht sinnvoll und zielführend sind, da sich lediglich geringe Unterschiede in den Bewertungen ergaben. Außerdem wurde durch Modul 3 gezeigt, dass sich zur Attributsreduktion der AHP anbietet, da hier Abwägungsentscheidungen zwischen jeweils zwei Kriterien vorgenommen werden und letztendlich eine Rangfolge der Kriterien aufgestellt werden kann. Ein weiterer Vorteil des AHP ist die Durchführbarkeit bei einer geringen Teilnehmerzahl, da auch individuelle Kriteriengewichtungen vorgenommen werden können. Zusammenfassend lässt sich schlussfolgern, dass vor allem die Nachvollziehbarkeit der Attributsfindung gewährleistet werden muss und dies über hochwertige Studien mittels qualitativen, quantitativen oder literaturbasierten Methoden möglich ist.

Die zweite Forschungsfrage

2. *Sind die Methoden der MCDM für Entscheidungen im gesundheitsökonomischen Kontext geeignet?*

kann aufgrund der folgenden Erkenntnisse bejaht werden. In Modul 4 konnten die Präferenzen von Lungen- und Darmkrebspatienten für ihre Therapie mittels DCE ermittelt werden. Anhand dieser Studie wurde zunächst deutlich, dass eine Aufbereitung des DCEs in einem Fragebogen und die Erhebung der Präferenzen auf diese Weise möglich waren. Bei der Translation der Ergebnisse in Gesprächen mit Onkologen konnte die Validität der Präferenzerhebung unterstützt werden. Die WTP-Studie zeigte ebenfalls, dass einerseits

subgruppenspezifische Präferenzen festgestellt und zum anderen diese Präferenzen in Geldeinheiten umgerechnet werden konnten. Dieses Vorgehen ist bei vielen gesundheitsbezogenen Fragestellungen umstritten, da eine Informationsasymmetrie zwischen den Leistungserbringen und Patienten bezüglich der tatsächlichen Kosten vorliegt. Somit sind Patienten im deutschen Gesundheitswesen selten mit den tatsächlichen Kosten für ihre Leistungs- oder Arzneimittelinanspruchnahme vertraut und können diese daher auch ungenügend einordnen bzw. gegenüber anderen Eigenschaften abwägen. In diesem Fall war eine WTP-Studie jedoch durchaus geeignet, da die Kosten von prädiktiven Ganzgenomsequenzierungstests bisher privat gezahlt werden müssen. Dahingegen könnten bei einem derartig beschaffenen DCE auch Limitationen aufgrund von Framing-Effekten, Protestantworten und realitätsferner Einschätzungen der tatsächlichen Kosten durch die Teilnehmenden auftreten, die bereits in der Literatur diskutiert wurden [24–27].

Allerdings ergaben sich aus den beiden DCE-Befragungen einige Schwierigkeiten, die im Folgenden diskutiert werden. Zunächst empfiehlt es sich, für ein DCE-Projekt aufgrund der benötigten Vorstudie und Teilnehmerrekrutierung ausreichend Zeit und Ressourcen zu berücksichtigen. Bei der Rekrutierung von Teilnehmenden ist sowohl ein ressourcenschonendes als auch standardisiertes Vorgehen erforderlich. Es sollte auch besonderer Wert darauf gelegt werden, dass die Attribute – je nach Fähigkeiten der Teilnehmer – umfassend definiert und verständlich sind, um die Validität der Erkenntnisse zu gewährleisten.

Die zukünftige Nutzung der Ergebnisse aus Präferenzstudien im Gesundheitswesen obliegt ebenfalls einigen Restriktionen. So wäre es beispielsweise denkbar, mit den Ergebnissen aus Präferenzstudien klinische Studien zu ergänzen und Arzneimittelbewertungen umfassend zu informieren. Es stellt sich jedoch die Frage, wer die Präferenzstudien durchführt. Zum einen können die pharmazeutischen Unternehmen die relevanten Endpunkte nicht umfassend in der klinischen Phase abschätzen, zum anderen stehen nach der Dossiereinreichung nur drei Monate bis zur Bewertung des Dossiers durch das IQWiG

und den GBA zur Verfügung. Somit steht bei der frühen Nutzenbewertung im Rahmen des AMNOGs hier die zeitliche Komponente in direkter Konkurrenz zur evidenzbasierten Studiendurchführung. Die Studien zu den Modulen aus dieser Arbeit nahmen zwischen zwei und vier Jahren in Anspruch, was nicht mit dem Zeitraum von drei Monaten im AMNOG-Prozess in Einklang zu bringen ist. Es gibt jedoch Pilotstudien, die eine Studiendurchführung in einem kürzeren Zeitraum getestet haben und mit entsprechenden Ressourcen erfolgreich durchgeführt werden konnten [28]. Allerdings diskutieren die Autoren, dass bei einer Erkrankung mit geringerer Prävalenz oder limitiertem Zugang die Erhebung erschwert gewesen wäre.

Die Methoden der Präferenz- oder Prioritätenabschätzung sollte jedoch zukünftig auch in anderen gesundheitsökonomischen Bereichen verstärkt verfolgt werden. Die Ergebnisse aus der Therapiepräferenzstudie (Modul 4) könnten beispielsweise nicht nur Entscheidungen zwischen Ärzten und Patienten unterstützen, sondern auch bei der Weiterentwicklung von Arzneimitteln eine wichtige Rolle einnehmen. Einfache Bewertungen ohne trade-off Entscheidungen sind hingegen in den meisten Fällen realitätsfern und sollten in Zukunft abgelöst werden, wenn evidente Entscheidungen getroffen werden sollen. Die MCDA kann somit einen wichtigen Beitrag dazu leisten, transparente Entscheidungen zu treffen und weg von Entscheidungen durch Konsens oder Experten zu gelangen.

Diese Empfehlungen, besonders in Bezug auf den AHP, unterliegen jedoch weiteren methodischen Restriktionen, die in den Modulen 6 bis 8 näher betrachtet wurden und die folgende Forschungsfrage beantworten konnten:

3. *Welche Herausforderungen oder Unsicherheiten ergeben sich bei der Anwendung in Bezug auf die Methoden?*

Während für das DCE bereits Leitlinien zur Studiendurchführung existieren [19], sind beim AHP noch keine entsprechenden Hilfen vorhanden. Hierzu können jedoch die durchgeführten AHP-Studien einen Beitrag leisten. Mit dem Modul 6 wurde die Forschungslücke zum aktuellen Anwendungsstand des AHP im Gesundheitswesen

geschlossen. Dadurch wurde deutlich, dass der AHP im Vergleich zum DCE noch aufgrund mangelnder methodischer Standards keine ebenbürtige Alternative in der MCDM darstellt. Der in der Vergangenheit häufig aufgetretenen Problematik von hohen Inkonsistenzen konnte in dem Projekt zu den Informationskriterien für Menschen mit seltenen Erkrankungen durch eine Rangreihung der Attribute begegnet werden. Die Befragten erhielten zu Beginn die Aufforderung, Kärtchen mit den Kriterien der Wichtigkeit nach zu sortieren. Auf diese weiterhin sichtbare Rangreihung konnte dann bei der AHP-Befragung zurückgegriffen werden. Anhand der Publikationen der Module 7 und 8 konnte die Empfehlung abgegeben werden, dass bei Aggregation geometrische Mittelwerte genutzt werden sollten und weitere methodische Ausgestaltungen in Bezug auf die dahinterstehende Frage vollzogen werden sollten. Ein wichtiger Aspekt ist zudem die Vergleichbarkeit der AHP-Ergebnisse mit denen aus dem BWS. Ebenfalls sollte berichtet werden, welche CR zugrunde gelegt wurde, da somit die Anzahl der Teilnehmenden stark beeinflusst werden konnte und dadurch ebenfalls Verzerrungspotential auftreten könnte. Außerdem wurden Erkenntnisse zur Aggregationsmethode gewonnen: Bei Entscheidungen zu rivalisierenden Gütern sollten Gruppenentscheidungen bevorzugt werden, da hier ein offener Diskurs über die gegensätzlichen Argumente geführt und ein Konsens gefunden werden kann. Bei intimen Entscheidungsproblemen oder unangenehmen Entscheidungen, bei denen individuelle Prioritäten von Interesse sind, können dementsprechend Einzelbefragungen durchgeführt werden.

Der Vergleich von AHP und DCE zeigte, dass beiden Methoden anwendbar sind und verlässliche Ergebnisse liefern können. Das Anwendungsfeld für jede Methode lässt sich über die zugrundeliegende Forschungsfrage ermitteln: Während der AHP bei Entscheidungen mit dem Ziel einer Rangfolge oder bei Konsensfindung einer Gruppe geeignet erscheint, lässt sich das DCE vorranging in der Präferenzmessung mit Gewichtung aller Attribute anwenden. Bei beiden Methoden muss jedoch auf die methodische Umsetzung Wert gelegt und auf die Transparenz der Ergebnisdarstellung geachtet werden.

Auch wenn mit dieser Arbeit ein wichtiger Beitrag zur Präferenzforschung geleistet wurde, bleiben weitere Fragen noch offen. So wird beispielsweise diskutiert, wer die Abwägung zwischen Eigenschaften bzw. Endpunkten vornehmen sollte. Bei der Entscheidungsfindung für eine Therapiealternative konnte gezeigt werden, dass Patientenpräferenzen subgruppenspezifisch vorlagen (Modul 4). Allerdings wurde in der systematischen Literaturrecherche zur Entscheidungsfindung bei Lungenkrebspatienten (Modul 1) herausgestellt, dass die Krebspatienten bevorzugt eine passive Rolle bei der Entscheidungsfindung einnehmen wollten. Demgegenüber steht die derzeit zunehmende Forderung nach Patientenbeteiligung (§ 140f Absatz 2 SGB V) [29] und Umsetzung des SDM. Bei Betrachtung der Ergebnisse aus der Therapiepräferenzstudie kann geschlussfolgert werden, dass der behandelnde Arzt aufgrund der zur Verfügung gestellten Informationen die Präferenzen beeinflussen kann. Daher sollte in Zukunft vermehrt auf die angemessene Information der Patienten geachtet werden. Nur so können die Patienten die vorhandenen Alternativen verstehen und Entscheidungen informiert treffen. Eine Übertragung der Forderung auf die Makro-Ebene für gesundheitspolitische Entscheidungen gilt gleichermaßen: Die hinzugezogenen Kriterien und die Gewichtung der Kriterien sollten offengelegt werden.

Referenzen

1. Greening LA, Bernow S. Design of coordinated energy and environmental policies. Energy Policy 2004; 32: 721–735
2. Wang J-J, Jing Y-Y, Zhang C-F, Zhao J-H. Review on multi-criteria decision analysis aid in sustainable energy decision-making. Renewable and Sustainable Energy Reviews 2009; 13: 2263–2278
3. Baltussen R, Niessen L. Priority setting of health interventions: the need for multi-criteria decision analysis. Cost effectiveness and resource allocation : C/E 2006; 4: 14
4. Department for Communities and Local Government (Hrsg.). Multi-criteria analysis. London: Communities and Local Government Publications, 2009
5. Nobre FF, Trotta LTF, Gomes LFAM. Multi-criteria decision making - an approach to setting priorities in health care. Statist. Med. 1999; 18: 3345–3354
6. Diaby V, Campbell K, Goeree R. Multi-criteria decision analysis (MCDA) in health care. Operations Research for Health Care 2013; 2: 20–24
7. Adunlin G, Diaby V, Xiao H. Application of multicriteria decision analysis in health care: a systematic review and bibliometric analysis. Health expectations : an international journal of public participation in health care and health policy 2015; 18: 1894–1905
8. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Analytic Hierarchy Process (AHP) – Pilotprojekt zur Erhebung von Patientenpräferenzen in der Indikation Depression. https://www.iqwig.de/de/projekte-ergebnisse/projekte/gesundheitsoekonomie/ga10-01-pilotstudie-analytic-hierarchy-process-in-der-indikation-majore-depression.1409.html#overview (letzter Zugriff am: 01.04.2017)
9. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) (Hrsg.). Allgemeine Methoden: Version 4.2. Köln: Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, 2015
10. Robert-Koch Institut (RKI). Demografische Alterung und Folgen für das Gesundheitswesen. https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=2&cad=rja&uact=8&ved=0ahUKEwiY_fL_i4PTAhUCFywKHYkcDXgQFggIMAE&url=https%3A%2F%2Fwww.rki.de%2FDE%2FContent%2FGesundheitsmonitoring%2FGesundheitsberichterstattung%2FGBEDownloadsK%2F2012_2_Demografischer_Wandel_Alterung.pdf%3F__blob%3DpublicationFile&usg=AFQjCNHMWtUJ-WM6rDQEtklZsvZUhjtRrA&sig2=V6QlP-8rY3PYThldzpyFlQ&bvm=bv.151325232,d.bGs (letzter Zugriff am: 01.04.2017)
11. Rasanen P, Roine E, Sintonen H, Semberg-Konttinen V, Ryynanen O-P, Roine R. Use of quality-adjusted life years for the estimation of effectiveness of health care: A systematic literature review. International journal of technology assessment in health care 2006; 22: 235–241
12. Schöffski O, Greiner W. Das QALY-Konzept als prominentester Vertreter der Kosten-Nutzwert-Analyse. In: Schöffski O, Graf von der Schulenburg J-M (Hrsg.). Gesundheitsökonomische Evaluationen. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012: 71–110
13. International Association of Athletics Federations. IAAF Scoring Tables for Conbined Events. https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwjumsnEyZrTAhUGiRoKHeSjBnoQFggfMAA&url=https%3A%2F%2Fwww.iaaf.org%2Fdownload%2Fdownload%3Ffilename%3D53f7d332-be0c-434c-8467-

1d9078966147.pdf%26urlslug%3DIAAF%2520Scoring%2520Tables%2520for%2520Combined%2520Events&usg=AFQjCNFjXsNdimPJi2cAmUuXEuqlVFSoyA&sig2=AA9TGSIm3pieZAqCI_o-9A&cad=rja (letzter Zugriff am: 10.04.2017)
14. Hwang CL, Yoon K. Multiple Attribute Decision Making: Methods and Applications A State-of-the-Art Survey: Springer Berlin Heidelberg, 2012
15. Mühlbacher A, Bethge S, Tockhorn A. Präferenzmessung im Gesundheitswesen. Gesundh ökon Qual manag 2013; 18: 159–172
16. Mühlbacher A, Bethge S, Ekert S, Tockhorn A, Nübling M. Der Wert von Innovationen im Gesundheitswesen: Spielen Patientenpräferenzen überhaupt eine Rolle? Recht und Politik im Gesundheitswesen 2008; 14: 53–62
17. Mühlbacher A, Kaczynski A. Making Good Decisions in Healthcare with Multi-Criteria Decision Analysis: The Use, Current Research and Future Development of MCDA. Applied health economics and health policy 2016; 14: 29–40
18. Saaty TL. Fundamentals of Decision Making and Priority Theory With the Analytic Hierarchy Process: RWS Publications, 2000
19. Bridges JFP, Hauber AB, Marshall D, et al. Conjoint analysis applications in health--a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value in health : the journal of the International Society for Pharmacoeconomics and Outcomes Research 2011; 14: 403–413
20. Geldermann, J, Lerche, N. Leitfaden zur Anwendung von Methoden der multikriteriellen Entscheidungsunterstützung. https://www.google.de/url?sa=t&rct=j&q=&esrc=s&source=web&cd=1&ved=0ahUKEwiqzcnbwYPTAhWDVSwKHf6fAtkQFggaMAA&url=https%3A%2F%2Fwww.uni-goettingen.de%2Fde%2Fdocument%2Fdownload%2F285813337d59201d34806cfc48dae518.pdf%2FMCDA-Leitfaden-PROMETHEE.pdf&usg=AFQjCNGE3Aez1_GhHT73dmFzVPxZRALjBQ&sig2=kYzLKqD0LXryAr3v21KZNw&bvm=bv.151325232,d.bGg&cad=rja (letzter Zugriff am: 01.04.2017)
21. Coast J, Horrocks S. Developing attributes and levels for discrete choice experiments using qualitative methods. Journal of health services research & policy 2007; 12: 25–30
22. Hiligsmann M, van Durme C, Geusens P, et al. Nominal group technique to select attributes for discrete choice experiments: an example for drug treatment choice in osteoporosis. Patient preference and adherence 2013; 7: 133–139
23. Hummel M, IJzerman M. The past and future of the AHP in health care decision making. XI International Symposium on the Analytic Hierarchy Process (ISAHP)
24. Ratcliffe J. The use of conjoint analysis to elicit willingness-to-pay values. Proceed with caution? International journal of technology assessment in health care 2000; 16: 270–275
25. Slothuus Skjoldborg U, Gyrd-Hansen D. Conjoint analysis. The cost variable: an Achilles' heel? Health economics 2003; 12: 479–491
26. Auspurg K, Liebe U. Choice-Experimente und die Messung von Handlungsentscheidungen in der Soziologie. Köln Z Soziol 2011; 63: 301–314
27. Meyerhoff J, Liebe U. Do protest responses to a contingent valuation question and a choice experiment differ? Environ Resource Econ 2008; 39: 433–446
28. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG). Präferenzmessung bei Parodontopathien. https://www.iqwig.de/download/GA15-01_Arbeitspapier_Praeferenzmessung-bei-Parodontopathien_V1-1.pdf (letzter Zugriff am: 10.04.2017)

29. Sozialgesetzbuch (SGB) Fünftes Buch (V): § 140f Beteiligung von Interessenvertretungen der Patientinnen und Patienten, In der Fassung der Bekanntmachung vom 20. Dezember 1988 (BGBl. I S. 2477), zuletzt geändert durch das Gesetzes zur Stärkung der Heil- und Hilfsmittelversorgung vom 04.04.2017 (BGBl. I S. 778) 2017

Module der kumulativen Dissertation

Identifikation von relevanten Entscheidungskriterien

Modul 1
Schmidt K, Damm K, Prenzler A, Golpon H, Welte T. Preferences of lung cancer patients for treatment and decision-making: a systematic literature review. European journal of cancer care 2016; 25 (4), S. 580–591.

Modul 2
Lingner H, **Schmidt K**, Aumann-Suslin I, Wittmann M, Schuler M, Schultz K. Reha-Komponenten im Patientenfokus – Sekundärdatenauswertung der Befragungen von COPD- und Asthma-Rehabilitanden. ZEFQ 2018. Online first, DOI: 10.1016/j.zefq.2018.03.008.

Modul 3
Schmidt K, Marten O, Kühne C, Zeidler J, Frank M. Einflussfaktoren auf die Standortwahl von hausärztlichen Land- und Stadtärzten in Niedersachsen. Gesundheitsökonomie & Qualitätsmanagement 2017. Online first. DOI: 10.1055/s-0043-103092.

Studiendurchführung in unterschiedlichen Anwendungsfeldern

Modul 4
Schmidt K, Damm K, Vogel A, Golpon H, Manns MP, Welte T, von der Schulenburg J-M. Therapy Preferences of Patients with Lung and Colon Cancer: A Discrete Choice Experiment. Patient Preference and Adherence 2017; 11, S. 1647–1656.

Modul 5
Plöthner M, **Schmidt K**, Schips C, Damm K. Which attributes of whole genome sequencing tests are most important for the general population? Results from a German preference study. Pharmacogenomics and Personalized Medicine 2018;11, S. 7–21.

Methodische Aspekte bei der Studiendurchführung

Modul 6
Schmidt K, Aumann I, Hollander I, Damm K, von der Schulenburg J-M. Applying the Analytic Hierarchy Process in healthcare research: A systematic literature review and evaluation of reporting. BMC Med Inform Decis Mak 2015; 15: 112.

Modul 7
Schmidt K, Babac A, Pauer F, Damm K, Schulenburg J-M. Measuring patients' priorities using the Analytic Hierarchy Process in comparison with Best-Worst-Scaling and rating cards: methodological aspects and ranking tasks. Health Economics Review 2016; 6(1): 50.

Modul 8
Pauer F, **Schmidt K**, Babac A, Damm K, Frank M, von der Schulenburg J-M. Comparison of different approaches applied in Analytic Hierarchy Process - an example of information needs of patients with rare diseases. BMC medical informatics and decision making 2016; 16 (1): 117.

Modul 1

Preferences of lung cancer patients for treatment and decision-making: a systematic literature review

Schmidt, Katharina

Damm, Kathrin

Prenzler, Anne

Golpon, Heiko

Welte, Tobias

European journal of cancer care 2016;

25 (4), S. 580–591.

European Journal of Cancer Care

Feature and Review Paper

Preferences of lung cancer patients for treatment and decision-making: a systematic literature review

K. SCHMIDT, MSC, RESEARCH ASSOCIATE, *Leibniz University Hannover, Center for Health Economics Research Hannover (CHERH), Hannover*, K. DAMM, PHD, SENIOR RESEARCH ASSOCIATE, *Leibniz University Hannover, Center for Health Economics Research Hannover (CHERH), Hannover, and Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), Member of the German Center for Lung Research (DZL), Hannover,* A. PRENZLER, PHD, SENIOR RESEARCH ASSOCIATE, *Leibniz University Hannover, Center for Health Economics Research Hannover (CHERH), Hannover, and Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), Member of the German Center for Lung Research (DZL), Hannover,* H. GOLPON, DR. MED., PROFESSOR, SENIOR PHYSICIAN, *Hannover Medical School, Clinic for Pneumology, Hannover, and Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), Member of the German Center for Lung Research (DZL), Hannover,* & T. WELTE, DR. MED, PROFESSOR, *Hannover Medical School, Clinic for Pneumology, Hannover, and Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), Member of the German Center for Lung Research (DZL), Hannover, Germany*

SCHMIDT K., DAMM K., PRENZLER A., GOLPON H. & WELTE T. (2016) *European Journal of Cancer Care* **25**, 580–591
Preferences of lung cancer patients for treatment and decision-making: a systematic literature review

The consideration of patient preferences in decision-making has become more important, especially for life-threatening diseases such as lung cancer. This paper aims to identify the preferences of lung cancer patients with regard to their treatment and involvement in the decision-making process. We conducted a systematic literature review from 12 electronic databases and included studies published between 2000 and 2012. A total of 20 studies were included in this review. These revealed that lung cancer patients do have preferences that should be considered in treatment decisions; however, these preferences are not homogenous. We found that patients often consider life extension to be more important than the health-related quality of life or undesirable side effects. This preference seems to depend on patient age. Nausea and vomiting are the most important side effects to be avoided; the relevance of other side effects differs highly between subgroups. The majority of lung cancer patients, nevertheless, seem to prefer a passive rather than an active role in decision-making, although the self-reported preferences differed partly from the physicians' perceptions. Overall, we identified an urgent need for larger studies that are suitable for subgroup analyses and incorporate multi-attributive measurement techniques.

Keywords: lung cancer, patient, preference, treatment, decision-making, systematic review.

INTRODUCTION

Worldwide, lung cancer is one of the most common malignancies and is the leading cause of cancer-related deaths, with a poor 5-year survival rate of approximately 15% (Ferlay *et al.* 2010). In 2008, more than 1.6 million new lung cancer cases and 1.3 million deaths were estimated. Since tumours are often diagnosed at advanced stages,

Correspondence address: Katharina Schmidt, Leibniz University Hannover, Center for Health Economics Research Hannover (CHERH), Otto-Brenner-Str. 1, D-30159 Hannover, Germany (e-mail: ks@cherh.de).

Accepted 11 November 2015
DOI: 10.1111/ecc.12425

treatments are intended to prolong survival time and palliate symptoms (Reck 2009). With regard to treatments, predominantly meaning chemotherapy, chemoradio- or radiotherapy, there is a general trade-off between improved survival duration (while accounting for toxicity) and palliative care without life extension. Therefore, decisions about treatment strategies involve such trade-offs between uncertain risks and benefits (Blinman *et al.* 2010, 2011).

In this context, many authors have emphasised the need to consider the patient and his or her attitudes and needs in such treatment decisions. The consideration of patient preferences during oncology treatment is seen as an indicator of quality in modern health care (Oliver & Greenberg 2009). Since a preference-based therapy offers a high fit to the individual needs of the patients, better compliance might be implied – a key success factor in tumour treatment.

Furthermore, in the context of reimbursement decisions, both the U.S. Food and Drug Administration and European institutions have underscored the importance of patient-reported or patient-relevant outcomes (Food and Drug Administration (FDA) 2009; Federal Joint Committee (G-BA) 2012; Institute for Quality and Efficiency in Health Care (IQWiG) 2013). However, it is necessary to first gather patient preferences in order to assess and interpret these outcomes (Mühlbacher *et al.* 2009).

Accordingly, before establishing health care for lung cancer patients, that is patient-centred and incorporates the patients' preferences three questions arise: (1) Do lung cancer patients have preferences regarding their treatment? (2) Which treatment attributes are most important to them (efficacy, side effects, administration form, etc.)? (3) Do they wish to participate in the decision-making process regarding the choice of a treatment (shared decision-making) and contribute their preferences here?

Hence, the aim of this systematic review is to identify the preferences of lung cancer patients, with regard to their treatment (which treatment attributes are important) and their involvement in the decision-making process in general.

METHODS

Relevant publications were identified through a structured search of 12 electronic databases, including Cochrane Central and Cochrane Reviews, DARE, EMBASE and EMBASE Alert, INAHTA, SOMED, MEDLINE, NHSEED, AMED, BIOSIS, and SciSearch, which were accessed through the German Institute of Medical Documentation and Information. The search terms were deliberately broad and included combinations of the English and German words for 'lung or pulmonary or bronchial' and 'cancer or carcinoma or neoplasm or tumour' in combination with 'patient' and 'preference or willingness'. Additionally, we performed manual research. These steps are undertaken in parallel to the publication for preferences of colorectal cancer patients by Damm *et al.* (2014).

In order to focus on current publications, the results were limited to studies that were published in English or German between 2000 and September 2012. Furthermore, the studies must have fulfilled the following inclusion criteria:

1. Preferences must have been stated by lung cancer patients; the opinions of others such as relatives were not included.
2. Publications must have referred to preferences concerning the actual treatment (not rehabilitation or follow-up) or the decision-making process.
3. Studies that analysed the general preferences of cancer patients were only included if the results for lung cancer patients were presented separately.
4. Only original research was included, i.e. qualitative interviews or quantitative studies; review articles were excluded.

Since the decision-making for or against different types of treatments (chemotherapy, chemoradiotherapy or radiotherapy) is in all cases a trade-offs between uncertain risks and benefits we did not determine on a specific treatment. First, two independent reviewers screened the title and abstract of the resulting studies with regard to their relevance. Second, they read the remaining full texts and checked the articles for the inclusion criteria. Disagreements between reviewers were discussed and were resolved by consensus, reached by re-reviewing the respective papers and discussing them with a third reviewer.

RESULTS

We have illustrated the results of our literature search process in Figure 1. Overall, we identified 8961 articles in our initial database search. Afterwards, we excluded 2675 duplicates. The review process of screening title and abstract resulted in 95 studies. In the next step, two independent reviewers read the full-text articles and discussed disagreements with a third reviewer. The systematic literature review resulted in 20 publications that concerned the preferences of lung cancer patients, including one study that was identified via manual research. Fourteen studies focused on the treatment preferences of lung can-

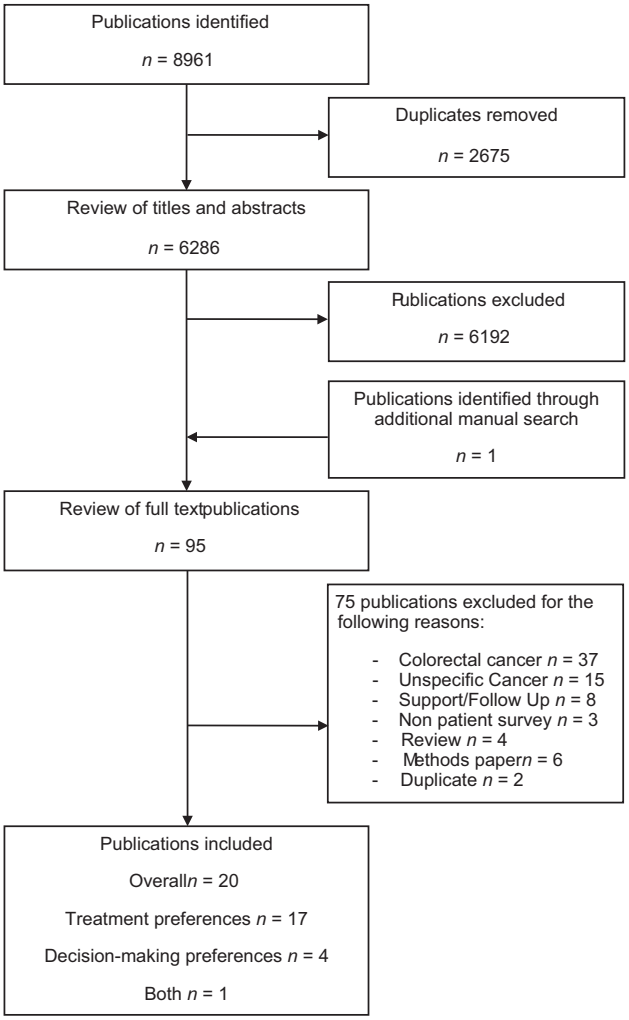


Figure 1. Flow chart of study selection.

cer patients, and two studies analysed the treatment preferences of cancer patients in general but reported separate results for lung cancer patients. Three publications surveyed the decision-making preferences of lung cancer patients, and one study surveyed both the decision-making and treatment preferences. The results of the latter study will be discussed separately in both the treatment preference and the decision-making involvement sections.

Treatment preferences

The results of the 17 publications that referred to the treatment preferences of lung cancer patients are summarised in Table 1. The majority of the studies (n = 15) analysed the trade-offs between uncertain risks (symptoms and treatment-related side effects) and benefits (survival time). Of these, three studies were conducted using qualitative methods (semi-structured or focus group interviews) and therefore included small numbers (between 5 and 13) of participants (Dorman *et al.* 2009; Thornton *et al.* 2011; Gerber *et al.* 2012). The results demonstrated the high relevance of survival benefits, the importance of 'doing something' or 'buying time', and the wish to try any treatment that might prolong life. The health-related quality of life (HRQoL), symptoms and side effects were also relevant to the patients to a lesser extent.

Quantitative analyses incorporated a wide range of different methods. Bridges *et al.* (2012) and Osoba *et al.* (2006) conducted discrete choice experiments (DCE) to survey the most important treatment attributes or the negative HRQoL effects to be avoided. The study by Bridges *et al.* (2012), which included 89 patients, showed that improvements in progression-free survival and symptom severity were the most important. Fatigue was considered the worst side effect. The study by Osoba *et al.* (2006) included 99 non-small cell lung cancer (NSCLC) patients and applied DCE to the domains and symptoms of the European Organisation for Research and Treatment of Cancer quality of life core (C30) questionnaire. In contrast to Bridges *et al.* (2012), the authors found that nausea/vomiting, pain and negative effects on emotional role functioning were the most important side effects to avoid.

Tang *et al.* (2008) used a one-choice-per-participant method to survey the preferences of patients with unresectable NSCLC with regard to shorter or longer palliative radiotherapy schedules, which were described by different attribute levels for duration, survival, distress, symptom control and costs. Of the 92 participants, 55% chose the longer duration schedule because of longer survival and better local control and 45% chose the shorter duration because of the shorter treatment duration, lower costs and better symptom control.

Dubey *et al.* (2005) used Likert scales to analyse the relevance of chemotherapy side effects. Of the 464 participants in this study, 75% considered side effects to be an important factor when choosing a particular regimen. Nausea was considered to be the worst side effect by the majority of patients. Subgroup analyses showed that female patients rated infections and hair loss as more important than did men; parents rated fatigue, hair loss and numbness as more important than did patients without children.

Two studies by Hirose *et al.* (2005, 2009) assessed patient preferences regarding chemotherapy and chemoradiotherapy respectively. In these studies, 73 patients with advanced NSCLC and 120 patients with other respiratory diseases rated the minimal benefits that would make both intensive and less intensive treatments acceptable. Patients with lung cancer were significantly more likely

Table 1. Medical treatment preferences: study characteristics, methods and results

Author, country, Year	Aim of study	Patients *N*; type of LC; comparison group C	Methods	Study design	Main results	Relevant subgroup results	Pharma sponsored
Bernard *et al.*, FR/BE, 2011	The perceived impact of alopecia in chemotherapy treatment	*N*: 135; NSCLC; C: no	WTP for chemotherapy that cut the risk from 40% to 5% via contingent valuation	Quant.	Mean WTP: €83.40 ± 10.2 (€median 37.50), representing 2.1% of total income; 27% of patients had no WTP	Sig. positive predictors: women, annual incomes	No
Bridges *et al.*, GB, 2012	Benefits that patients judge sufficient to compensate for the risks associated with therapy	*N*: 89; NSCLC, 98% non-adv.; C: no	Online DCE; attributes: progression-free survival, pain/cough/shortness of breath, rash, diarrhoea, fatigue, nausea/vomiting, fever, application	Quant.	Increases in progression-free survival together with improvements in symptom severity were most important; fatigue was most important risk, followed by diarrhoea, nausea/vomiting, fever/infection, rash. Oral administration preferred to infusion	No	Yes
Brundage *et al.*, CA, 2000	Evaluation of a decision aid: combined-modality treatment vs. radiotherapy alone	*N*: 18; adv. NSCLC; C: no	Treatment trade-off method: thresholds by varying (1) the probability of surviving 3 years and (2) median survival time	Quant.	14 participants chose combined-modality treatment for an absolute 3-year survival advantage of 5%. 12 patients chose combined-modality treatment for a median survival advantage of 10 weeks	No	No
Chu *et al.*, Asia, Europe, Latin America, 2007	Differences between patient self-assessed treatment preference vs. external-assessed preference by their physicians	*N*: 1884; adv. NSCLC C: their physicians	Questionnaires used to categorise: (A) 'max. extension of survival with acceptance of high toxicity', (B) 'max. extension of survival only if coupled with normal life style', and (C) 'relief of symptoms'	Quant.	(A): patients 60%, physicians 39% (B): patients 26%, physicians 33% (C): patients 14%, physicians 29%	Differences between countries and regions	Yes
Dorman *et al.*, GB, 2009	Patients' treatment goals and preferences	*N*: 9; NSCLCwith brain metastases C: no	Semi-structured interviews (themes: e.g. HRQoL vs. LoL, HRQoL factors, desire to try everything, fear)	Qual.	HRQoL is important to patients, but many are keen to try any treatment which might prolong their life	No	No
Dubey *et al.*, US, 2005	Patient concerns about chemotherapy SE	*N*: 464; adv. NSCLC; C: no	Mailed questionnaire, Likert scales for: symptom, nausea/vomiting, hair loss, fatigue, numbness/tingling, (risk of) infection	Quant.	75% consider SE important for decision on a particular regimen; nausea/vomiting was the most important SE	Sig. differences: women: infection risk, hair loss; with children: fatigue, hair loss, numbness; with chemotherapy: nausea, infection	?

Table 1. *Continued*

Author, country, Year	Aim of study	Patients *N*; type of LC; comparison group C	Methods	Study design	Main results	Relevant subgroup results	Pharma sponsored
Gerber *et al.*, US, 2012	Patients' perceptions of maintenance chemotherapy	*N*: 13; adv. NSCLC; C: no	Focus group interviews (main themes: pros/cons of maintenance chemotherapy, importance of HRQoL, experiences, consideration of maintenance chemotherapy)	Qual.	Key themes for maintenance chemotherapy: (1) survival benefits, disease control, 'buying time'; (2) importance of 'doing something'; (3) HRQoL; (4) role of provider opinion; (5) importance of logistics	No	No
Gironés *et al.*, ES, 2012	Relationship between preferences and chemotherapy use in patients ≥70 years	*N*: 83; NSCLC or SCLC; C: no	One choice between four hypothetical treatments (intensive, less intensive chemotherapy, palliative radiotherapy, no treatment)	Quant.	Most patients chose active treatment (38.6% intensive with most survival benefit, 18% less intensive and survival benefit), 31.3% chose no treatment	Sign. more likely to accept aggressive chemotherapy: elderly (note: contrary within publication), lower performance status, independent, not depressed	No
Hirose *et al.*, JP, 2005	Impact of survival, response rate, and symptom relief against toxicity of chemotherapy	*N*: 73; adv. NSCLC; C: 120 with other respiratory diseases	Rate min. benefit to make 2 (intensive, less intensive) treatments acceptable, for 'chance of cure,' 'response not cure,' and 'symptom relief'	Quant.	Patients with NSCLC significantly more likely to accept either intensive or less intensive treatments for a potentially small benefit for 'chance of cure,' 'response but not cure,' and 'symptom relief'	Sig. positive predictors for less toxicity tolerance: age	No
Hirose *et al.*, JP, 2009	Impact of survival, response rate, and symptom relief against toxicity of chemoradiotherapy	(see Hirose *et al.* 2005)	(see Hirose *et al.* 2005)	Quant.	(see Hirose *et al.* 2005)	No	No
Jensen *et al.*, DK, 2008	Oral or iv Vinorelbine	*N*: 61; advanced NSCLC; C: no	One choice (oral/iv) after both experiences	Quant.	75% prefer oral vinorelbine	No	Yes
Lang *et al.*, TW, 2010	WTP for a hypothetical cure	*N*: 294; lung cancer; C: no	Contingent valuation (double-bounded dichotomous choice and open-ended questions)	Quant.	NTD 7416 or NTD 7032 per month	Sig. positive predictors: female, income, having family; Sig. negative: Karnofsky Performance Status	No
Leighl *et al.*, CA, 2006	WTP for oral EGFR TKI vs. Docetaxel	*N*: 57; adv. NSCLC; C: 54 healthy	Contingent valuation bidding game: therapies described by response, symptom improvement, SE	Quant.	Patients: median CAD 100 per month (range 0–5000); Healthy subjects: median CAD 100 per month (range 0–3000)	Sig. positive predictor: prior chemotherapy	No

Table 1. *Continued*

Author, country, Year	Aim of study	Patients *N*; type of LC; comparison group C	Methods	Study design	Main results	Relevant subgroup results	Pharma sponsored
Meropol *et al.*, US, 2008	Cancer patient values regarding HRQoL and LoL	*N*: 83; adv. LC; C: 376 with adv. cancer	Computer-based survey. One choice from: HRQoL is all that matters, HRQoL is more important but LoL matters, LoL is more important but HRQoL matters, LoL is all that matters	Quant.	LC patients preferred HRQoL over LoL more often; more than half of the patients preferred both equally	Sig.ly positive associated with a preference for HRQoL: older age, male gender, higher education	No
Osoba *et al.*, US, 2006	Cancer patients preferences for HRQoL functional domains and symptoms	*N*: 99; NSCLC; C: 276 with cancer	Ranking questions and DCE; attributes: physical, role, social, emotional, and cognitive function, chemotherapy-related SE, financial difficulties	Quant.	Most important effects to avoid for NSCLC patients: nausea/vomiting, pain, emotional, role functioning and dyspnoea	Late-stage patients: dyspnoea, nausea/vomiting, role, pain, emotional functioning	Yes
Tang *et al.*, AU, 2008	Preferred radiotherapy schedule (longer vs. shorter) in palliation	*N*: 92; unresectable NSCLC; C: no	One choice between two schedules with different attribute levels (survival, distress, symptom control, duration, cost)	Quant.	55% chose longer duration due to longer survival (90%) and better local control (12%), shorter duration (45%) was chosen for shorter duration (80%), cost (61%) and better symptom control (20%)	No	No
Thornton *et al.*, GB, 2011	Factors that influence patients' choice of treatment during the oncologist-consultation	*N*: 5; adv. NSCLC; C: no	Semi-structured interviews (themes: communication, information, hope, other factors)	Qual.	Relevant factors for choice of treatment: survival, HRQoL, symptoms, SE; patients preferred chemotherapy due to greatest chance of survival, radiotherapy due to less toxic effects	No	No

Adv., advanced; CAD, Canadian dollar; DCE, discrete choice experiment; EGFR TKI, epidermal growth factor receptor tyrosine kinase inhibitors; EORTC, European Organisation for Research and Treatment of Cancer; HRQoL, health-related quality of life; LC, lung cancer; LoL, length of life; max., maximum; min., minimum; NSCLC, non-small cell lung cancer; NTD, new Taiwan dollar QLQ, quality of life questionnaire; SCLC, small cell lung cancer; SE, side effect(s); sign., significant(ly) ($P < 0.05$); WTP, Willingness to pay; quant., quantitative study design; qual., qualitative study design.

to accept both the intensive and less intensive treatments. Low tolerance for toxicity correlated with high patient age.

Another study conducted by Chu *et al.* (2007) used a questionnaire to categorise the treatment preferences of patients into three groups: (A) 'maximum extension of survival with acceptance of high toxicity', (B) 'maximum extension of survival only if coupled with a normal life style' and (C) 'relief of symptoms'. A total of 1884 patients with advanced NSCLC, as well as their physicians, participated in this study. The physicians were asked to assess their patients' treatment preferences. The study demonstrated that the physicians' perceptions differ from the actual patient preferences. However 60%, 26% and 14% of the patients self-assessed their treatment needs as (A), (B) and (C), respectively, 39%, 33% and 29% of the physicians assessed their patients' treatment preferences as (A), (B), and (C) respectively.

Gironés *et al.* (2012) analysed the treatment preferences of 83 elderly (≥70 years) NSCLC and small cell lung cancer patients. In this survey, the participants were asked to choose one of four hypothetical treatments (intensive chemotherapy, less intensive chemotherapy, palliative radiotherapy or no treatment). Most patients chose an active treatment, with 38.6% choosing the intensive chemotherapy with the highest survival benefit; 18%, the less intensive chemotherapy with a lower survival benefit; and 31.3%, no treatment. Elderly patients with lower performance status and non-depressive patients were significantly more likely to accept more aggressive chemotherapy.

Meropol *et al.* (2008) performed a computer-based survey of 83 advanced lung cancer patients to analyse the importance of length of life (LoL) compared to HRQoL. Participants chose one out of the following four statements: the HRQoL is all that matters, the HRQoL is more important but the LoL matters, the LoL is more important but the HRQoL matters and the LoL is all that matters. Approximately 30% of the participants preferred HRQoL to LoL, approximately 50% valued these outcomes equally, and 20% preferred LoL to HRQoL. Older age and male gender were positively associated with a preference for HRQoL.

The surveys by Leighl *et al.* (2006) and Bernard *et al.* (2011) and the treatment trade-off study by Brundage *et al.* (2000) had different focuses compared to the previously described studies. The latter evaluated a decision support aid that was designed to reveal the patient's outcome preferences. Eighteen patients rated their 3-year and median survival advantage thresholds to choose a more intensive treatment [combined-modality treatment (CMT)] over less intensive radiotherapy. As a result, 14 of 18 and 12 of 18 patients chose CMT for a 3-year survival advantage of 5% and a median survival advantage of 10 weeks respectively.

Bernard *et al.* (2011) focused on the impact of alopecia. A total of 135 NSCLC patients were asked to rate the positive effects of a chemotherapy regimen that would reduce the risk of alopecia from 40% to 5% via a Contingent Valuation technique (a technique to reveal the willingness to pay (WTP)]. The mean WTP for a 3-week cycle was 83.40 EUR ± 10.2 (2.1% of the total income). About 27% of the participants, mostly men (77%), were unwilling to pay additionally for the lower risk of alopecia. Of those patients with a WTP, female patients were willing to pay more. The patient's annual income correlated positively with the WTP. The mean WTP for the question of threshold for certain not to pay for product B is 173.9 EUR (±18.8).

Leighl *et al.* (2006) conducted a Contingent Valuation study and asked 57 lung cancer patients and 54 healthy subjects to state their WTP for oral epidermal growth factor receptor tyrosine kinase inhibitors versus docetaxel, described by their toxicities, route of administration and benefits. Both groups were willing to pay a median amount of 100 Canadian Dollars (1 CAD ≈ 0.737 EUR) per month, although the range differed (lung cancer patients: 0–5000 CAD; healthy subjects: 0–3000 CAD).

In contrast to the studies that analysed outcome and side effect trade-offs, Jensen *et al.* (2008) focused their survey on the application preferences for vinorelbine. This is a semisynthetic vinca-alkaloid cytotoxic drug that can also be given by the oral route. At the end of a crossover trial with 61 lung cancer patients who received both orally and intravenously administered vinorelbine, three of four preferred the oral application rather than the intravenous one.

A study conducted by Lang (2010) evaluated the WTP for a hypothetical cure in 294 lung cancer patients using the Contingent Valuation technique. The authors noted a WTP of 7032 New Taiwanese Dollars (NTD; 100 NTD ≈ 2.63 EUR) per month. The positive predictors of WTP were female gender, income and having a family; a negative predictor was the Karnofsky Performance score, which measures the general well-being and ability to perform daily life activities of cancer patients (0%: death, 100%: no complaints).

Preferences for involvement in treatment decision-making

The results of the four studies that concern the decision-making process are summarised in Table 2. All four used

Table 2. Involvement in treatment decision-making: study characteristics, methods and results

Author, Country, Year	Aim of study	Patients *N*; type of LC; comparison group C	Methods	Study design	Results	Relevant subgroup results	Pharma sponsored
Brundage *et al.*, CA, 2000	Evaluation of a decision aid: combined-modality treatment vs. radiotherapy alone	*N*: 18; adv. NSCLC; C: no	Control preferences scale	Quant.	7 patients favoured active, 3 favoured passive, 8 patients favoured shared role	No	No
Hotta *et al.*, JP, 2010	Preferences for involvement in treatment decisions and extent of concordance between patients and physicians	*N*: 28; relapsed NSCLC; C: physicians	Control preferences scale	Quant.	Four patients favoured active, 7 a passive, 17 a shared role	Depression and anxiety status (via HADS) sign. associated with preferred roles	No
Pardon *et al.*, BE, 2009	Preferences for participating in decision-making concerning treatment options	*N*: 128; adv. NSCLC; C: no	Control preferences scale	Quant.	15% favoured active, 63% a passive, 22% a shared role	Sign. more likely to prefer shared or personal control: patients living alone; no pain	Yes
Pardon *et al.*, BE, 2012	Changing preferences for participating in decision-making concerning treatment options	*N*: 66; adv. NSCLC; C: no	Control preferences scale; three times, over 4 months	Quant.	In t_1: 15% favoured an active, 62% a passive, 23% a shared role. No change at t_2 and t_3: 50%. Doctor control stays most preferred	Sign. more likely to prefer more participation over time: non-religious patients; more pain	No

HADS, Hospital Anxiety and Depression Scale; NSCLC, Non-small cell lung cancer.

Table 3. Control preference scale statements

Patients' decision role preference
(a) I prefer to make the final selection of which treatment I will receive
(b) I prefer to make the final selection of my treatment after seriously considering my doctor's opinion
(c) I prefer that my doctor and I share responsibility for deciding which treatment is best for me
(d) I prefer that my doctor makes the final decision about which treatment will be used, but seriously considers my opinion
(e) I prefer to leave all decisions regarding my treatment to my doctor

Source: Own representation based on Degner and Sloan (1992)

the Control Preference Scale, an instrument based on the Degner *et al.* (1997) card sort method. In this method, patients choose a decision role statement that best describes their preferences (Table 3). The five statements represent three categories of patients: those who wish to take an active role (statements a and b), a shared role (statement c) or a passive role (statements d and e).

A study performed by Brundage *et al.* (2000) evaluated 18 Canadian NSCLC patients (55% men, mean age: 68). In this study, 39%, 44% and 17% of patients favoured active, shared and passive roles respectively. Hotta *et al.* (2010) evaluated 28 Japanese patients (79% men, mean age: 68) and found that 14%, 61% and 25% favoured active, shared and passive roles respectively. Pardon *et al.* (2009) studied 128 patients (80% men, mean age: 64) with NSCLC. Among other questions, the authors asked the participants to choose their preferred role in both general medical decisions and specific treatment decisions. With regard to general medical decisions, 9%, 42% and 49% favoured active, shared and passive roles respectively. Patients with a low level of education or those who had regular contact with physicians were more likely to prefer shared or active roles. With regard to specific treatment decisions, 15%, 22% and 63% favoured active, shared and passive roles respectively. Patients who lived alone or did not experience pain were more likely to prefer a shared or active role. In a follow-up study, Pardon *et al.* (2012) analysed whether these preferences changed over time. Sixty-seven of the above-mentioned 128 patients were asked three times to choose their preferred roles in medical and treatment decisions. During a 4-month period, 50% of the participants changed their favoured roles to indicate a preference for either more or less participation. However, the majority of patients still preferred a passive role.

DISCUSSION

To our knowledge, this is the first study to analyse the available evidence regarding the preferences of lung cancer

patients for treatment and decision-making. Thus, we complement a former MEDLINE-based review by Blinman *et al.* (2010). However, the previous authors only searched for chemotherapy-related studies. That review included five publications, and two of those studies were also included in our review since they were published between 2000 and 2012 (Hirose *et al.* 2005, 2009). Furthermore, since treatment regimes, benefits and side effects as well as patient self-images change over time, it is important to update the available knowledge regarding the preferences of lung cancer patients.

In the following section, we will summarise and discuss the preferences of lung cancer patients regarding the trade-off between the length and quality of life, the importance of certain side effects and decision-making preferences.

With regard to the trade-off between the length and quality of life, qualitative studies have shown that survival benefits, 'buying time', and the wish to try anything that might prolong life seem to be more important than the HRQoL, symptoms and side effects. These findings were encouraged by a quantitative analysis by Tang *et al.* (2008), who found that 'longer survival' was the main reason for choosing a particular treatment regimen. Chu *et al.* (2007) also showed that a majority of patients wanted the 'maximum extension of survival with acceptance of high toxicity'. Additionally, Hirose *et al.* (2005, 2009) found a higher willingness to accept intensive treatments among lung cancer patients than among patients with other respiratory diseases. A similar effect was shown by Gironés *et al.* (2012), who only included patients ≥70 years old; again, a high proportion of the patients accepted intensive treatment because of the survival benefits. We only found one study in which more participants preferred HRQoL to the LoL in a direct comparison; however, a majority of participants valued both benefits equally (Meropol *et al.* 2008).

Some studies in this review conducted subgroup analyses with respect to age. Hirose *et al.* (2005, 2009) and Meropol *et al.* (2008) found that elderly patients were more likely to choose less toxic treatments and to prefer HRQoL over the LoL respectively. Gironés *et al.* (2012) also conducted age-related subgroup analyses; however, the results within the publication were contrary and hence no conclusion could be drawn.

Even if the results were not completely homogenous, the following conclusion can be drawn: life extension often is more important than either HRQoL or undesirable side effects, although some other studies identified the same importance. This preference might, however, depend on the patients' age. The earlier review by Blinman *et al.* (2010) confirmed our findings. The authors also concluded that the survival benefits of a toxic treatment need only be moderate in order to be chosen by lung cancer patients. One question arising here concerns the health economic evaluation of different treatment interventions. The frequently used quality-adjusted life years calculation, which multiplies the number of life years gained by the HRQoL in an unweighted way, must be questioned.

With respect to the importance of side effects, both Osoba *et al.* (2006) and Dubey *et al.* (2005) reported that nausea and vomiting were the side effects that contributed the most to the choice of chemotherapy. However, it became apparent that subgroup analyses are crucial. Dubey *et al.* (2005) and Bernard *et al.* (2011) showed the high relevance of gender-specific preferences. These authors found that alopecia was more relevant to female patients. Another factor is the familial situation. Dubey *et al.* (2005) showed that having children also influenced the relevance of side effects, as fatigue and hair loss were more important to parents than to patients without children.

In summary, this review indicates that nausea and vomiting are important side effects to be avoided. Some of the studies suggested that the relevance of other side effects (e.g. fatigue, hair loss and dyspnoea) could differ between subgroups (gender, familial situation and age). With respect to treatment-related nausea and vomiting, these preferences might result from the patient's perception of the severity of adverse effects. Griffin *et al.* and Coates *et al.* found that cancer patients generally perceived nausea and/or vomiting to be the most severe (Coates *et al.* 1983; Griffin *et al.* 1996). This might be explained by the pathophysiology and relevance of these two side effects (Hickok *et al.* 2003; Shelke *et al.* 2004). Hence, there might be a correlation between the severity of a side effect and the preference to avoid this specific side effect.

With regard to decision-making preferences, we identified only four studies, all of which used the Control Preferences Scale by Degner *et al.* (1997). Overall, the results of these studies indicate that the majority of lung cancer patients would rather choose a passive role than an active role. On the one hand, this is an interesting phenomenon, since Chu *et al.* (2007) stated that patients highly value survival benefits, whereas physicians strongly emphasise on toxicity and associated symptoms. Davidson *et al.* (2011), who reviewed the literature regarding the influence of physician and patient factors when determining lung cancer chemotherapy, also identified a mismatch between the physicians' perceptions and the patients' preferences. This finding should encourage patients to express their preferences.

On the other hand, lung cancer patients have special characteristics that need to be considered. For example, authors who have analysed the general preferences of cancer patients found that lung cancer patients were more likely to prefer a passive role in decision-making than were other cancer patients (Davidson *et al.* 1999; Tariman *et al.* 2010). An explanation for this finding could be that the decision-making preference depends on the disease severity. Ende *et al.* (1989), who examined the preferences of patients for decision-making and information seeking, found a negative association between disease severity and the desire to make decisions. Since lung cancer patients have a poor prognosis with regard to survival in comparison with other cancer patients (Ferlay *et al.* 2010), the severity of the disease might result in the preference for a more passive role. We also expected that education would have a positive influence on the desire to participate in the decision process. Ende *et al.* (1989) found an association between education status and the desire for autonomy. However, in our review, we found that Brundage *et al.* (2000) reported the opposite association. Because of the relatively small sample size of the studies as well as the heterogeneous education status of the patients, these findings should be considered with caution and verified in future studies.

As we conducted the systematic literature review for colon cancer patients (see Damm *et al.* 2014), too, the results should be compared shortly. The side effects of colorectal cancer patients were similar to the side effects of lung cancer patients, although the results for colorectal cancer patients related on few not completely homogenous studies. The important side effects for them were diarrhoea, nausea, vomiting and incontinence. Colorectal cancer patients attached great importance to avoid a permanent stoma. The trade-off between LoL and HRQoL could not be defined precisely in the colorectal cancer review. However, the before-mentioned study of Meropol *et al.* (2008) compared these two aspects directly for lung and colorectal cancer patients: They indicated that both patients groups prefer LoL over HRQoL, but the majority rated both equally. Relevant subgroup results could be assumed for gender and age, which we found also for lung cancer patients in this review. With regard to decision-making involvement, the majority of colorectal cancer patients preferred a passive rather than an active role. Consequently, the two systematic literature reviews showed a similar picture for lung and colorectal cancer patients although disease specific problems occurred.

Overall, the majority of the quantitative studies included in this review had only small sample sizes that did not allow for extensive statistical analyses; only five of the 20 studies included more than 100 patients. Additionally, only three of the studies reported a prospective sample size calculation for the statistical analyses (Jensen *et al.* 2008; Tang *et al.* 2008; Bridges *et al.* 2012) and the majority of the surveys limited their analysis to simple rating tasks (Likert scale). Furthermore, the methodology used in the studies regarding treatment preferences (qualitative interviews, questionnaire techniques and experiments) varied, which could lead to heterogeneous results. This impression agrees with the findings of Blinman *et al.* (2010), who also reported difficulties when comparing study results due to the use of different methods. Cultural differences might also influence the results, since the included studies were conducted in Asia ($n = 4$), Australia ($n = 1$), Europe ($n = 8$), or North America ($n = 6$) or incorporated patients from various nations ($n = 1$).

There is an urgent need for future research. Since the majority of the identified studies included only small sample sizes and did not use extensive statistical analyses, there is a need for larger studies that are suitable for subgroup analyses.

Studies examining treatment preferences should increasingly include more than just a single attribute and avert 'single-choice' or Likert scale designs. Herein, multi-attributive measurement techniques such as discrete choice methods are recommended. This kind of study designs measure the importance of different attributes compared to others. Hereby, e.g. the relevance of alopecia as one possible side effect can be measured compared to others, showing the relative importance of this treatment attribute. The preference for different forms of administration (oral or intravenous) can be measured as well as the relevance of the administration form compared to other treatment aspects. The trade-off between HRQoL and LoL can be examined as well by these techniques.

The above identified side effects that are important to avoid are also covered by lung cancer-specific questionnaires used to measure HRQoL in clinical studies (Damm *et al.* 2013). However, they are not based on multi-attributive or trade-off preference measurement techniques like discrete choice or gamble methods, and therefore do represent medical parameters, only. By using the mentioned techniques it is possible to refine the instruments and even derive reference-based single indices (Rowen *et al.* 2011).

Our review shows the heterogeneous picture of patients' preferences regarding lung cancer therapy. This is due not least to the fact that authors do not use all available options and adequate methods to measure the preferences. This should be considered in future research efforts.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

FUNDING

The project was funded by the Federal Ministry of Education and Research.

REFERENCES

Bernard M., Brignone M., Adehossi A., Pefoura S., Briquet C., Chouaid C. & Tilleul P. (2011) Perception of alopecia by patients requiring chemotherapy for non-small-cell lung cancer: a willingness to pay study. *Lung Cancer* **72**, 114–118.

Blinman P., Alam M., Duric V., McLachlan S.A. & Stockler M.R. (2010) Patients' preferences for chemotherapy in non-small-cell lung cancer: a systematic review. *Lung Cancer* **69**, 141–147.

Blinman P., McLachlan S.A., Nowak A.K., Duric V.M., Brown C., Wright G., Millward M., Fong K. & Stockler M.R. (2011) Lung cancer clinicians' preferences for adjuvant chemotherapy in non-small-cell lung cancer: what makes it worthwhile? *Lung Cancer* **72**, 213–218.

Bridges J.F., Mohamed A.F., Finnern H.W., Woehl A. & Hauber A.B. (2012) Patients' preferences for treatment outcomes for advanced non-small cell lung cancer: a conjoint analysis. *Lung Cancer* **77**, 224–231.

Brundage M.D., Feldman-Stewart D., Dixon P., Gregg R., Youssef Y., Davies D. & MacKillop W.J. (2000) A treatment trade-off based decision aid for patients with locally advanced non-small cell lung cancer. *Health Expectations* **3**, 55–68.

Chu D.T., Kim S.W., Kuo H.P., Ozacar R., Salajka F., Krishnamurthy S., Damyanov D., Altug S., Reece W.H. & Wang L. (2007) Patient attitudes towards chemotherapy as assessed by patient versus physician: a prospective observational study in advanced non-small cell lung cancer. *Lung Cancer* **56**, 433–443.

Coates A., Abraham S., Kaye S.B., Sowerbutts T., Frewin C., Fox R.M. & Tattersall M.H. (1983) On the receiving end–patient perception of the side-effects of cancer chemotherapy. *European Journal of Cancer and Clinical Oncology* **19**, 203–208.

Damm K., Roeske N. & Jacob C. (2013) Health-related quality of life questionnaires in lung cancer trials – a systematic literature review. *Health Economics Review* **3**, 15.

Damm K., Vogel A. & Prenzler A. (2014) Preferences of colorectal cancer patients for treatment and decision-making: a systematic literature review. *European Journal of Cancer Care* **23**, 762–772.

Davidson J.R., Brundage M.D. & Feldman-Stewart D. (1999) Lung cancer treatment decisions: patients' desires for participation and information. *Psychooncology* **8**, 511–520.

Davidson P.M., Jiwa M., Goldsmith A.J., McGrath S.J., Digiacomo M., Phillips J.L., Agar M., Newton P.J. & Currow D.C. (2011) Decisions for lung cancer chemotherapy: the influence of physician and patient factors. *Supportive Care in Cancer* **19**, 1261–1266.

Degner L.F. & Sloan J.A. (1992) Decision making during serious illness: what role do patients really want to play? *Journal of Clinical Epidemiology* **45**, 941–950.

Degner L.F., Sloan J.A. & Venkatesh P. (1997) The control preferences scale. *Canadian Journal of Nursing Research* **29**, 21–43.

Dorman S., Hayes J. & Pease N. (2009) What do patients with brain metastases from non-small cell lung cancer want from their treatment? *Palliative Medicine* **23**, 594–600.

Dubey S., Brown R.L., Esmond S.L., Bowers B.J., Healy J.M. & Schiller J.H. (2005) Patient preferences in choosing chemotherapy regimens for advanced non-small cell lung cancer. *Journal of Supportive Oncology* **3**, 149–154.

Ende J., Kazis L., Ash A. & Moskowitz M.A. (1989) Measuring patients' desire for autonomy: decision making and information-seeking preferences among medical patients. *Journal of General Internal Medicine* **4**, 23–30.

Federal Joint Committee (G-BA) (ed.) (2012) Verfahrensordnung: Bewertung des Nutzens von Arzneimitteln nach § 35a SGB. [Procedure: benefit assessment of medicines in accordance with § 35a SGB] Available at: http://www.g-ba.de/downloads/62-492-667/VerfO_2012-12-06.pdf (accessed April 4, 2013).

Ferlay J., Shin H.R., Bray F., Forman D., Mathers C. & Parkin D.M. (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *International Journal of Cancer* **127**, 2893–2917.

Food and Drug Administration (FDA) (ed.) (2009) Guidance for industry: patient-reported outcome measures used in medical product development to support labeling claims, draft guidance. Available at: www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf (accessed April 4, 2013).

Gerber D.E., Hamann H.A., Rasco D.W., Woodruff S. & Lee S.J. (2012) Patient comprehension and attitudes toward maintenance chemotherapy for lung cancer. *Patient Education and Counseling* **89**, 102–108.

Gironés R., Torregrosa D., Gómez-Codina J., Maestu I., Tenias J.M. & Rosell R. (2012) Lung cancer chemotherapy decisions in older patients: the role of patient preference and interactions with physicians. *Clinical and Translational Oncology* **14**, 183–189.

Griffin A.M., Butow P.N., Coates A.S., Childs A.M., Ellis P.M., Dunn S.M. & Tattersall M.H. (1996) On the receiving end. V: Patient perceptions of the side effects of cancer chemotherapy in 1993. *Annals of Oncology* **7**, 189–195.

Hickok J.T., Roscoe J.A., Morrow G.R., King D.K., Atkins J.N. & Fitch T.R. (2003) Nausea and emesis remain significant problems of chemotherapy despite prophylaxis with 5-hydroxytryptamine-3 antiemetics: a University of Rochester James P. Wilmot Cancer Center Community Clinical Oncology Program Study of 360 cancer patients treated in the community. *Cancer* **97**, 2880–2886.

Hirose T., Horichi N., Ohmori T., Kusumoto S., Sugiyama T., Shirai T., Ozawa T., Ohnishi T. & Adachi M. (2005) Patients preferences in chemotherapy for advanced non-small-cell lung cancer. *Internal Medicine* **44**, 107–113.

Hirose T., Yamaoka T., Ohnishi T., Sugiyama T., Kusumoto S., Shirai T., Okuda K., Ohmori T. & Adachi M. (2009) Patient willingness to undergo chemotherapy and thoracic radiotherapy for locally advanced non-small cell lung cancer. *Psychooncology* **18**, 483–489.

Hotta K., Kiura K., Takigawa N., Yoshioka H., Hayashi H., Fukuyama H., Nishiyama A., Yokoyama T., Kuyama S., Umemura S., Kato Y., Nogami N., Segawa Y., Yasugi M., Tabata M. & Tanimoto M. (2010) Desire for information and involvement in treatment decisions: lung cancer patients' preferences and their

physicians' perceptions: results from Okayama Lung Cancer Study Group Trial 0705. *Journal of Thoracic Oncology* **5**, 1668–1672.

Institute for Quality and Efficiency in Health Care (IQWiG) (ed.) (2013) General Methods. Available at: https://www.iqwig.de/download/General_Methods_4-0.pdf (accessed April 4, 2013).

Jensen L.H., Osterlind K. & Rytter C. (2008) Randomized cross-over study of patient preference for oral or intravenous vinorelbine in combination with carboplatin in the treatment of advanced NSCLC. *Lung Cancer* **62**, 85–91.

Lang H.C. (2010) Willingness to pay for lung cancer treatment. *Value Health* **13**, 743–749.

Leighl N.B., Tsao W.S., Zawisza D.L., Nematollahi M. & Shepherd F.A. (2006) A willingness-to-pay study of oral epidermal growth factor tyrosine kinase inhibitors in advanced non-small cell lung cancer. *Lung Cancer* **51**, 115–121.

Meropol N.J., Egleston B.L., Buzaglo J.S., Benson A.B. 3rd, Cegala D.J., Diefenbach M.A., Fleisher L., Miller S.M., Sulmasy D.P. & Weinfurt K.P.; CONNECT Study Research Group (2008) Cancer patient preferences for quality and length of life. *Cancer* **113**, 3459–3466.

Mühlbacher A.C., Bethge S. & Tockhorn A. (2009) Entscheidungen auf Basis von Effizienzgrenzen: Berücksichtigung von Patientenpräferenzen. *Public Health Forum* **17**, 25.e1-3.

Oliver A. & Greenberg C.C. (2009) Measuring outcomes in oncology treatment: the importance of patient-centered outcomes. *Surgical Clinics of North America* **89**, 17–25.

Osoba D., Hsu M.A., Copley-Merriman C., Coombs J., Johnson F.R., Hauber B., Manjunath R. & Pyles A. (2006) Stated preferences of patients with cancer for health-related quality-of-life (HRQOL) domains during treatment. *Quality of Life Research* **15**, 273–283.

Pardon K., Deschepper R., Stichele R.V., Bernheim J.L., Mortier F. & Deliens L., EOLIC-Consortium (2009) Preferences of advanced lung cancer patients for patient-centred information and decision-making: a prospective multicentre study in 13 hospitals in Belgium. *Patient Education and Counseling* **77**, 421–429.

Pardon K., Deschepper R., Stichele R.V., Bernheim J.L., Mortier F., Bossuyt N., Schallier D., Germonpré P., Galdermans D., Van Kerckhoven W. & Deliens L.; EOLIC-Consortium (2012) Changing preferences for information and participation in the last phase of life: a longitudinal study among newly diagnosed advanced lung cancer patients. *Supportive Care in Cancer* **20**, 2473–2482.

Reck M. (2009) Quality of life as an endpoint of clinical trials. *Lung Cancer* **64**(Suppl. 1), 22–23.

Rowen D., Brazier J., Young T., Gaugris S., Craig B.M., King M.T. & Velikova G. (2011) Deriving a preference-based measure for cancer using the EORTC QLQ-C30. *Value Health* **14**, 721–731.

Shelke A.R., Mustian K.M. & Morrow G.R. (2004) The pathophysiology of treatment-related nausea and vomiting in cancer patients: current models. *Indian Journal of Physiology and Pharmacology* **48**, 256–268.

Tang J.I., Shakespeare T.P., Lu J.J., Chan Y.H., Lee K.M., Wong L.C., Mukherjee R.K. & Back M.F. (2008) Patients' preference for radiotherapy fractionation schedule in the palliation of symptomatic unresectable lung cancer. *Journal of Medical Imaging and Radiation Oncology* **52**, 497–502.

Tariman J.D., Berry D.L., Cochrane B., Doorenbos A. & Schepp K. (2010) Preferred and actual participation roles during health care decision making in persons with cancer: a systematic review. *Annals of Oncology* **21**, 1145–1151.

Thornton M., Parry M., Gill P., Mead D. & Macbeth F. (2011) Hard choices: a qualitative study of influences on the treatment decisions made by advanced lung cancer patients. *International Journal of Palliative Nursing* **17**, 68–74.

Modul 2

Reha-Komponenten im Patientenfokus – Sekundärdatenauswertung der Befragungen von COPD- und Asthma-Rehabilitanden

Lingner, Heidrun*

Schmidt, Katharina*

Aumann-Suslin, Ines

Wittmann, Michael

Schuler, Michael

Schultz, Konrad

Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen 2018;

Online first, DOI: 10.1016/j.zefq.2018.03.008.

Verfügbar unter:

https://zefq-journal.com/article/S1865-9217(18)30061-8/fulltext

* *geteilte Erstautorenschaft*

Modul 3

Einflussfaktoren auf die Standortwahl von hausärztlichen Land- und Stadtärzten in Niedersachsen

Schmidt, Katharina

Marten, Ole

Kühne, Christian

Zeidler, Jan

Frank, Martin

Gesundheitsökonomie & Qualitätsmanagement 2017;

Online first. DOI: 10.1055/s-0043-103092.

Verfügbar unter:

https://www.thieme-connect.de/products/ejournals/abstract/10.1055/s-0043-103092

Modul 4

Therapy Preferences of Patients with Lung and Colon Cancer: A Discrete Choice Experiment

Schmidt, Katharina

Damm, Kathrin

Vogel, Arndt

Golpon, Heiko

Manns, Michael P.

Welte, Tobias

von der Schulenburg, J.-Matthias

Patient Preference and Adherence 2017;

11, S. 1647–1656

Patient Preference and Adherence

Dovepress
open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Therapy preferences of patients with lung and colon cancer: a discrete choice experiment

Katharina Schmidt[1]
Kathrin Damm[1]
Arndt Vogel[2]
Heiko Golpon[3,4]
Michael P Manns[2]
Tobias Welte[3,4]
J-Matthias Graf von der Schulenburg[1,4]

[1]Leibniz University of Hannover, Center for Health Economics Research (CHERH), Hannover, Germany; [2]Department of Gastroenterology, Hepatology and Endocrinology, Hannover Medical School, Hannover, Germany; [3]Department of Pneumology, Hannover Medical School, Hannover, Germany; [4]Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), Member of the German Center for Lung Research (DZL), Hannover, Germany

Objectives: There is increasing interest in studies that examine patient preferences to measure health-related outcomes. Understanding patients' preferences can improve the treatment process and is particularly relevant for oncology. In this study, we aimed to identify the subgroup-specific treatment preferences of German patients with lung cancer (LC) or colorectal cancer (CRC).

Methods: Six discrete choice experiment (DCE) attributes were established on the basis of a systematic literature review and qualitative interviews. The DCE analyses comprised generalized linear mixed-effects model and latent class mixed logit model.

Results: The study cohort comprised 310 patients (194 with LC, 108 with CRC, 8 with both types of cancer) with a median age of 63 (SD =10.66) years. The generalized linear mixed-effects model showed a significant ($P<0.05$) degree of association for all of the tested attributes. "Strongly increased life expectancy" was the attribute given the greatest weight by all patient groups. Using latent class mixed logit model analysis, we identified three classes of patients. Patients who were better informed tended to prefer a more balanced relationship between length and health-related quality of life (HRQoL) than those who were less informed. Class 2 (LC patients with low HRQoL who had undergone surgery) gave a very strong weighting to increased length of life. We deduced from Class 3 patients that those with a relatively good life expectancy (CRC compared with LC) gave a greater weight to moderate effects on HRQoL than to a longer life.

Conclusion: Overall survival was the most important attribute of therapy for patients with LC or CRC. Differences in treatment preferences between subgroups should be considered in regard to treatment and development of guidelines. Patients' preferences were not affected by sex or age, but were affected by the cancer type, HRQoL, surgery status, and the main source of information on the disease.

Keywords: patient preferences, lung cancer, colorectal cancer, Germany, latent class model, multi-criteria decision making

Introduction

In 2012, lung cancer (LC) and colorectal cancer (CRC) were two of the most common cancers worldwide.[1] In developed countries, the 5-year survival rates of patients with CRC improved significantly between 1995 and 2009,[2] whereas those of patients with LC showed only minor improvement.[2] The aim of the World Health Organization 2013–2020 Global Action Plan is to reduce the rate of cancer mortality by improving service delivery through early diagnosis and enhanced screening programs.[3] The prevalence of cancer will rise as a result of earlier detection of the disease; thus, the therapeutic options available will gain more attention.

Patients are often confronted with different therapeutic options, which may implicate severe adverse effects and uncertain outcomes. Typically, the patients evaluate therapeutic options in terms of their ability to prolong survival versus their

Correspondence: Katharina Schmidt
Leibniz Universität Hannover, Center for Health Economics Research Hannover (CHERH), Otto-Brenner-Str 1, 30159 Hannover, Germany
Tel +49 511 762 17346
Fax +49 511 762 5081
Email ks@cherh.de

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/PPA.S138863

expected effects on health-related quality of life (HRQoL). Consequently, it is important to know what patients prefer and what is most important to them during decision making. Previous studies have shown a discrepancy between the personal preferences of patients and the subjective assessments made by their physicians.[4–9] However, we performed a systematic literature review which showed that, in general, patients do not wish to decide on their therapy personally and would prefer their physician to make the decision.[10,11] Here, a conflict can occur, because the therapy that is provided to patients should be adjusted to meet their preferences regarding HRQoL and adverse effects. Previous studies on other diseases have shown that satisfaction with therapy can have a significant effect on disease outcomes and further treatment decisions.[12–14] On the basis of these findings, patient preferences should be examined and integrated during decision making regarding cancer therapy.

Furthermore, patient preferences might influence political decisions regarding reimbursement for pharmaceuticals. In Germany, there is growing interest in preference measurement, following the introduction of the Act on the Reform of the Market for Medical Products (Arzneimittelneuordnungsgesetz [AMNOG]) in 2011. Although, according to the AMNOG, patient-reported outcomes should be taken into account during early evaluation of the benefits of new pharmaceuticals,[15] patient preferences do not play an important role in Germany at present. The Federal Joint Committee (Gemeinsamer Bundesausschuss) has criticized the lack of quality in the scientific evaluations (dossiers) of pharmaceutical companies and overruled some patient-reported outcomes.[16] Hence, our findings might influence the ongoing debate about the evidence provided by studies of patient preferences and be relevant to the German health care system.

The aim of the study was to examine the therapy preferences of German patients with LC and CRC. These two types of cancer show high prevalence rates in Germany and worldwide. However, their divergent overall survival rates and disease-related adverse effects might lead to different therapy preferences among patients with LC and CRC. In addition, we wanted to identify subgroups of patients that shared similar preferences, irrespective of the cancer type. Members of these homogenous subgroups might share same sex, age, or educational level. Comparison of the two different cancer types and the resulting patient (subgroup-specific) preferences represents the added value of our study. Consequently, our aim is to confirm the importance of patient preference studies and their need for implementation in health care. These data could also help physicians to make clinical decisions by differentiating among the preferences of various subgroups of patients and might enable improvement of therapy guidelines.

Patients and methods

Derivation of attributes and discrete choice tasks

In a discrete choice experiment (DCE), two (or more) alternative scenarios are presented. Each alternative (profile) is described by several attributes.[17] The participant must choose which of the profiles they prefer.[18]

The whole study process is illustrated in Supplementary material. First, we conducted a systematic literature review to identify the key topics related to cancer therapy for use in subsequent qualitative interviews.[10,11] The systematic literature reviews identified 15 relevant studies of preferences with respect to therapy for CRC and 17 relevant studies of preferences with respect to therapy for LC. The most important concerns for patients with CRC were diarrhea, nausea, pain, requirement of a stoma, role functioning, emotional functioning, toxicity of chemotherapy, life expectancy/overall survival, and taking medication at home.[10,11] For patients with LC, the most important concerns identified by the literature review were: fatigue, diarrhea, nausea, pain, role functioning, intensity of treatment, overall survival, and HRQoL versus length of life.[10,11] Second, we conducted guided qualitative interviews that were based on the results of our systematic literature review. We interviewed 18 patients with LC and 17 patients with CRC, and then conducted content analyses (Aumann et al[19] for interviews with LC patients and [Damm et al: Supplementary material] for interviews with CRC patients). We used the inductive and deductive categories from the content analysis to identify the main topics: adverse effects, social quality of life, emotional quality of life, and organization. Further subcategories (10–23) were established for each main topic. We sorted the identified categories on the basis of the frequency with which they were mentioned, separated by patients with LC and patients with CRC (Supplementary material). Subsequently, we chose the most frequently mentioned categories and determined whether they overlapped with respect to meaning. We aimed to cover a large spectrum of categories, while simultaneously ensuring minimal overlap or correlation between the attributes. To this end, we aggregated the categories into topics that could serve as attributes. Another restriction was the required total number of attributes (five to nine) to prevent overstraining of the interviewee.[20]

To generate the questionnaire, we identified five attributes from the most important categories that did not overlap in their meanings and added the attribute of overall survival from the literature review. We realized that the resulting attributes were similar for patients with CRC and those with LC. Therefore, we decided to use the same attributes for both groups of cancer patients. The first attribute that we examined was the efficacy of therapy, measured as additional life expectancy after diagnosis. Given that the time of survival can vary considerably between patients with CRC and those with LC, we decided to examine the objective values rather than specific time periods. Adverse effects were separated into three attributes of "physical capacity", "appearance", and "food intake and digestion". Given this separation, we expected no overlaps between the attributes. The different possible levels assigned to the attributes were derived from the experiences of the interviewees and were divided into "minor", "medium", and "strong" effects. In the interviews, "physical capacity" was described as tiredness, decreased physical ability, and overall physical exhaustion. We carried these descriptions over into the questionnaire. The symptoms that were associated most commonly with "appearance" were hair loss, weight loss, and eczema. The fifth attribute identified was "waiting time in the clinic or therapy-associated practice". It corresponded to the time that patients had to spend waiting during therapy, for example, waiting time between blood tests and the start of chemotherapy. The final attribute referred to the provision of a "guide" who was independent and would provide information on the services and assistance associated with treatment for LC or CRC. During the interviews, the patients were highly critical of the treatment process and its organization. One of the more frustrating factors for the patients was the lack of information, rather than the waiting time itself, and more specifically, the strain that resulted from the lack of information on disease-associated proposals and paperwork. Some patients also mentioned that they had to coordinate communication between their doctors and their health insurance providers. Therefore, for the purpose of the questionnaire, we introduced the concept of a guide who would provide support for the patients either personally or over the telephone. This "guide" was defined as a free-of-charge service to reduce the effect of any monetary concerns that the patients might have. There were only two possible levels for this attribute: "yes" or "no". The final attributes and levels, including a description of the study participants, are presented in Table 1.

We used the Statistical Analysis Software % ChoicEff macro to construct choice sets.[21] We used two versions of the questionnaire because blocking certain choice sets was found to reduce the burden on patients' decision making. The first choice set enabled us to test patients' understanding of the DCE method because it included a dominant profile. In total, we provided 10 choice sets of DCE tasks to each participant (for an example of choice set, see Supplementary material).

Table 1 Descriptions of attributes used in the questionnaire

Attribute		Levels
Life expectancy	Life expectancy at the time of diagnosis with regard to mean survival in patients with lung or colon cancer (average of all cancer stages)	– Not increased – Slightly increased – Strongly increased
Physical capacity	Decrease in physical capacity that influences everyday life, for example, being out of breath quickly, being tired, sitting down often, or sleeping during the day	– Normal – Moderately decreased – Strongly decreased
Appearance	Changes in appearance caused by the disease itself or the treatment (adverse effects). Possible changes include hair loss, eczema, or weight loss	– Unchanged (no visible changes) – Slightly changed – Significantly changed
Food intake and digestion	Problems with food intake or digestion, such as loss of appetite, nausea, emesis, or diarrhea	– No problems/normal – Minor problems – Severe problems
Waiting time (in the clinic)	The time spent waiting in the clinic or practice for your therapy. This could be, for example, the waiting time between blood tests and the start of chemotherapy	– No waiting time – Moderate waiting time – Long waiting time

Ethical standards

The patients provided written informed consent to participate. Approval for this study was obtained from the ethics committee of the Hannover Medical School (Nr 1518–2012) and the Medical Association of Lower Saxony, the University of Goettingen, and the University Hospital Tuebingen.

Development of the questionnaire

We conducted a pretest to ensure that the final questionnaire could be understood easily by the patients. The pretest showed that most patients could not answer questions about their disease state or therapy goals (palliative, adjuvant, maintenance). Therefore, this question was excluded from

the questionnaire because we were not allowed to access medical records.

The final questionnaire consisted of a section on patient information, a form regarding informed consent, a definition of the attributes, Likert-scale questions about the therapy attributes (from 1, "very unimportant" to 5, "very important"), 10 DCE sets, sociodemographic questions, and the cancer-specific HRQoL questionnaire developed by the European Organization for Research and Treatment of Cancer (EORTC), termed the EORTC QLQ-C30[22,23] (for an overview of the variables, see Supplementary material).

Study population

Patients attending specialized ambulatory practices or the departments of pneumology or gastroenterology at eight hospitals in Germany were invited to participate in our study. The cooperating institutions were (for further information, see Supplementary material):

- Hannover Medical School, Department of Pneumology and Department of Gastroenterology, Hepatology and Endocrinology, Hannover;
- Johannes Wesling Medical Center, Department of Hematology, Oncology, Hemostaseology, and Palliative Care UKRUB, University of Bochum, Minden;
- Lung Cancer Center, Hospital Region Hannover;
- Clinic for Visceral, General, and Transplant Surgery, Surgical Study Center, University Hospital Tuebingen;
- Ambulatory Oncological Center, Hannover;
- Group Practice for Internal Medicine and Pulmonology, Celle;
- Interdisciplinary Short-term Oncology, Department of Hematology and Medical Oncology, Goettingen; and
- Group Practice for Hematology and Oncology, Hannover.

The participating clinics administered the questionnaire to patients with LC and CRC of all disease stages who were aged ≥18 years and had finished at least one cycle of chemotherapy (including in the past). Both modes of chemotherapy administration (tablet and infusion) were eligible for inclusion in the study.

In addition, we initiated an online survey with the same inclusion criteria. The link to the survey was distributed via the Facebook page of the German self-help organization ILCO, the Felix Burda Colon Cancer Website and Facebook page, the Center for Health Economics Research Hannover Facebook page, and the mailing lists of regional self-help groups for patients with CRC and LC.

The recruitment period was from September 2014 to October 2016. Neither patients nor physicians received any incentives for participating in the study. All participants provided informed consent. The minimum required sample size was 196, which was calculated in accordance with the study by de Bekker-Grob et al.[24]

Approval for the study was obtained from the ethics committees of the Hannover Medical School (reference number: 1518–2012), Medical Association of Lower Saxony, University of Goettingen, and University Hospital Tuebingen.

Data analyses

Following completion of the survey, we cleansed the data set (testing for impossible values, systematic missings, import errors, and so on) and calculated descriptive statistics for the variables (median, SD, percentages). The HRQoL was calculated using symptom scales, functional scales, and the global health score from the EORTC QLQ-C30 questionnaire.[25] We applied logistic regression analyses to determine factors (independent variables) that influenced the choices made between the profiles of each choice set (dependent variables). The utility of each profile was calculated using Formula 1 in Supplementary material.

We used a generalized linear mixed-effects model (GLMM) to examine the effects of multiple answers for each individual choice set (serial_no). We calculated the GLMM for patients with CRC and LC separately, so that any differences between the two patient groups could be identified based on Hauber et al[26] and McCulloch et al[27] (Formula 2 in Supplementary material).

Finally, we used the latent class mixed logit model (LCMLM) with a different number of classes to identify, strictly on the basis of the data, possible sample subgroups with specific characteristics (eg, sociodemographic status, disease-specific parameters). These subgroup characteristics were presented in the so-called class-membership effects model. An overview of the variables tested for all models is provided in Supplementary material. The final model is shown in Formula 3 in Supplementary material.

The β-coefficients from the GLMM and LCMLM represent the weights of the utility for choosing the profile. β-coefficients >0 indicated that an attribute level was preferred, whereas coefficients <0 indicated that it was disfavored. Alternatively, coefficients <0 suggested that an attribute level was accepted in order to gain advantages in other attributes. The results for the β-coefficients were assumed to be significant at a P-value ≤ 0.05.

The models were tested with different independent variables and, finally, lean models were targeted. Akaike and Bayesian information criteria were used to identify the model with the best fit for the data. All analyses were conducted

with R statistics 3.1.2 (The R Foundation for Statistical Computing, Vienna, Austria), using the packages "lme4" (for GLMM) and "lcmm" (for LCMLM).

Results

Descriptive statistics

In total, 369 patients participated initially in the study, but this number decreased to 310 participants after data cleansing. The distribution of mean age and sex did not differ significantly between the included and excluded groups of participants. Table 2 shows the characteristics of the patients. Given that only eight patients had both types of cancer, we did not assess their preferences separately.

The cohort was younger than the average ages of patients with LC and CRC in Germany.[28,29] However, the general sex distribution of patients with LC and CRC in Germany was similar to that evident in our sample.

Multivariate models

Generalized linear mixed-effects model

Figure 1 shows the results of the three GLMMs (LC, CRC, full sample). A strong increase in life expectancy was the attribute level that was given the most weight by all three groups ($\beta_{LC,OS2}$ =2.56, $\beta_{CRC,OS2}$ =1.77, $\beta_{ful,OS2}$ =2.17; all $P<0.001$). For patients with LC, the level of "normal physical capacity" was given greater weight than a "moderate" or "strong decrease" in physical capacity ($\beta_{LC,PC0}$ =0.79, $\beta_{LC,PC1}$ =0.34, $\beta_{LC,PC2}$ =–1.13; $P<0.001$). However, both the patients with CRC and the full sample rated "normal physical capacity" more highly than "moderately decreased capacity", although this was not statistically significant. With regard to "changes in appearance", all patient groups gave a greater weight to a "slightly changed appearance" than to an "unchanged appearance". "No problems" or "minor problems" with food intake and digestion were rated slightly higher by patients with LC than those with CRC ($\beta_{LC,FI0}$ =0.83, $P<0.001$; $\beta_{CRC,FI0}$ =0.49, $P<0.001$; $\beta_{LC,FI1}$ =0.18, P=0.05; $\beta_{CRC,FI1}$ =–0.14, P=0.15). "No waiting time" (reference category) was given slightly less weight by the full sample than by patients with CRC ($\beta_{full,WT0}$ =0.25, $\beta_{CRC,WT0}$ =0.35; $P<0.001$). In general, the preferences of the three groups were very similar (see also Supplementary material).

Latent class mixed logit model

The LCMLM identified three different classes of patients with specific class-membership effects (Table 3; for a graphical presentation, see Supplementary material). The first class showed a strong preference for "clearly longer survival" ($\beta_{cl1,OS2}$=1.56, $P<0.001$). In contrast, this class disfavored "slightly longer survival" ($\beta_{cl1,OS1}$=–0.2, $P<0.001$). Patients in Class 1 accepted a "moderately decreased physical capacity" compared with a "normal physical capacity". In addition,

Table 2 Sample characteristics of included participants

Characteristic	CRC	LC	Both	Total
Sample size	108	194	8	310
Sex	49.6% men	69.80% men	40% men	62.16% men
Median age (SD) in years	59.5 (12.66)	63 (10.58)	48.5 (8.90)	63 (10.66)
Cancer type				
CRC	100%	0%	0%	35.04%
LC	0%	100%	0%	63.03%
Both	0%	0%	100%	1.93%
Median disease duration (SD) in years	2 (5.92)	1 (2.14)	7.5 (7.20)	1 (4.16)
Marital status				
Single	8.6%	10.0%	0%	9.3%
Married	69.6%	70.3%	80.0%	70.2%
Divorced	13.2%	13.9%	0.2%	13.8%
Widowed	8.6%	5.7%	0%	6.6%
School-leaving qualifications				
None	2.3%	1.6%	40.0%	2.5%
Primary school	33.0%	48.2%	2.0%	42.4%
Secondary school	34.1%	30.0%	40.0%	31.6%
High school	30.7%	20.1%	0%	23.4%
Median global health status (SD)				
Scale from 0 (worst) to 100 (best)	66.7 (22.69)	58.3 (20.44)	58.3 (20.56)	66.7 (21.56)
Median HRQoL (SD)				
Scale from 1 (very bad) to 7 (excellent)	5 (1.48)	5 (1.27)	5 (0.87)	5 (1.35)

Abbreviations: CRC, colorectal cancer; LC, lung cancer; HRQoL, health-related quality of life.

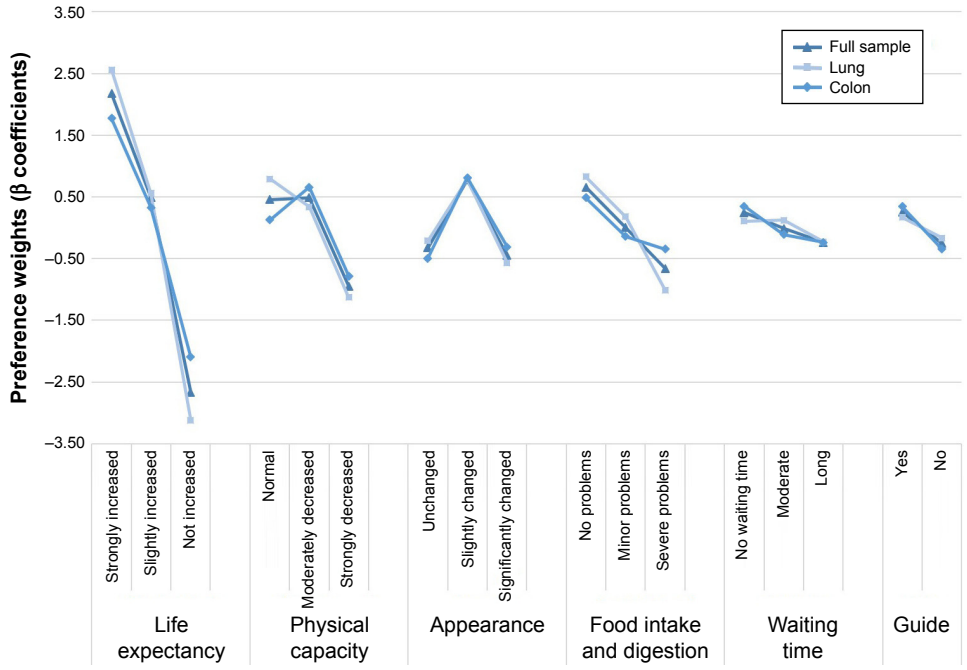


Figure 1 Results of mixed logit models.
Notes: Triangular shape, full sample; rectangular shape, lung cancer; diamond shape, colon cancer; random intercept: serial_no.

this class was willing to accept "moderate" and "long waiting times". Patients in Class 2 showed a preference for both "clearly longer" and "slightly longer survival" ($\beta_{cl2,OS2}$=0.64, $\beta_{cl2,OS1}$ =0.36; $P<0.001$). However, "physical capacity", "appearance", and "food intake and digestion" were also important attributes for this class. In this model, a decrease from "clearly longer" to "slightly longer survival" (β_{OS2}: 0.64−β_{OS1}: 0.36=0.28) could be compensated for by a change from a "strong decrease" to a "moderate decrease" in physical capacity (β_{PC2}: −0.66−β_{PC1}: 0.34=−1). Consequently, patients in Class 2 were willing to trade prolonged survival for smaller decreases in physical capacity. We cannot interpret the preferences of Class 3 in regard to "overall survival" because the results were not statistically significant ($P>0.05$). This group

Table 3 Latent class mixed logit model results – attribute preferences

Attribute	Level	Class 1		Class 2		Class 3	
		β_{cl1}	*P*-value	β_{cl2}	*P*-value	β_{cl3}	*P*-value
Intercept		ne		0.07		0.05	
Overall survival	Clearly longer	1.56	0.00	0.64	0.00	0.07	0.27
	Slightly longer	−0.20	0.00	0.36	0.00	0.07	0.33
	Not longer (ref)	−1.36		−1.00		−0.14	
Physical capacity	Normal	0.08	0.02	0.32	0.00	−0.49	0.00
	Moderate decrease	0.28	0.00	0.34	0.00	0.57	0.00
	Strong decrease (ref)	−0.35		−0.66		−0.08	
Appearance	Unchanged	−0.05	0.16	0.14	0.00	−0.58	0.00
	Slightly changed	−0.01	0.85	0.37	0.00	0.34	0.00
	Significantly changed (ref)	0.06		−0.51		0.25	
Food intake and digestion	No problems	0.17	0.00	0.42	0.00	−0.14	0.03
	Minor problems	−0.01	0.89	0.21	0.00	−0.52	0.00
	Strong problems (ref)	−0.17		−0.63		0.66	
Waiting time	None	−0.18	0.00	−0.04	0.25	0.04	0.52
	Moderate	0.02	0.51	0.29	0.00	−0.20	0.00
	Long (ref)	0.16		−0.26		0.16	
Guide	Yes	−0.03	0.21	0.22	0.00	0.07	0.14
	No (ref)	0.03		−0.22	0.04	−0.07	0.31

Notes: Age (standardized), sex, cancer type, HRQoL, disease duration (centered by mean), radiation therapy, and change of appearance are used as class membership effects.
Abbreviations: cl, class; HRQoL, health-related quality of life; ne, not estimated; ref, reference.

Table 4 Class-membership effects of latent class mixed logit models (reference: Class 3)

Fixed-effects class-membership model	Class 1			Class 2		
	Coefficient	Standard error	*P*-value	Coefficient	Standard error	*P*-value
Intercept	2.609	1.380	0.059	1.875	1.368	0.171
Age (mean centered)	0.021	0.029	0.485	0.035	0.029	0.226
Sex (ref = male)	−1.003	0.728	0.168	−0.497	0.714	0.486
CRC (ref = LC)	**−2.214**	**0.786**	**0.005**	**−1.686**	**0.757**	**0.026**
Both cancers (ref = LC)	−2.074	2.008	0.301	−1.790	2.021	0.376
HRQoL (mean centered)	**−0.537**	**0.267**	**0.044**	−0.403	0.263	0.125
Disease duration (mean centered)	0.000	0.069	0.998	−0.037	0.072	0.609
Surgery (ref = no)	**1.997**	**0.796**	**0.012**	**1.701**	**0.778**	**0.029**
Radiation (ref = no)	−1.069	0.691	0.122	−0.796	0.678	0.241
Changes in appearance (ref = no)	−0.322	0.336	0.338	−0.372	0.332	0.263
Information: physician	**2.575**	**0.784**	**0.001**	**2.714**	**0.738**	**0.000**

Note: Significant values are shown in bold.
Abbreviations: CRC, colorectal cancer; HRQoL, health-related quality of life; LC, lung cancer; ref, reference.

disfavored the most favorable levels of the attributes "physical capacity", "appearance", and "food intake problems" ($\beta_{cl3,PC0}$ =–0.49, $\beta_{cl3,AP0}$ =–0.58, $\beta_{cl3,FI0}$ =–0.14, $\beta_{cl3,FI1}$ =–0.52; $P<0.05$). However, they gave a greater weight to the middle levels for "physical capacity" and "appearance" than to the other levels.

Next, we investigated the class-membership effects for the three classes. Of all the patients, 42.13% were assigned to Class 1, 47.24% to Class 2, and 10.63% to Class 3. The differences between classes 1 and 2 (referenced against Class 3) are presented in Table 4. Patients in classes 1 and 2 did not differ significantly from patients in Class 3 in terms of age, sex, or duration of disease. Classes 1 and 2 had a lower proportion of patients with CRC than Class 3 ($\beta_{cl1,CRC}$ =–2.21, $\beta_{cl2,CRC}$ =–1.69; $P<0.05$). The classes also differed in terms of their therapy experiences: patients in Class 1 were more likely to have undergone surgery than those in classes 2 and 3 ($\beta_{cl1,treat\_1}$ =1.997, $\beta_{cl2,treat\_1}$ =1.7; $P<0.05$). We also observed a difference between the classes with regard to the main source of information on their disease. Patients in classes 1 and 2 were more likely to obtain relevant information from their physician than patients in Class 3 ($\beta_{cl1,info\_1}$ =2.58, $\beta_{cl2,info\_1}$ =2.71; $P<0.05$). Other sources of information (other patients, books, the Internet, self-help groups) were shown to have no significant influence on the model. In addition, patients in Class 1 showed significantly worse HRQoL outcomes (β_{cl1,LQ_30_s} =–0.54, P=0.04) than patients in the other classes.

Discussion

We systematically investigated the differences in the therapy preferences of patients with two divergent types of cancer. Whereas previous studies have examined the therapy preferences of patients with different disease states of the same cancer type, we compared the preferences of patients with CRC and LC. In our first model (mixed logit model), we found that patients with LC and CRC had almost the same preferences for therapy attributes and differed only slightly in their preferences. In the strictly data-driven LCMLM, we found that cancer type, current HRQoL status, and the source of information were important for the therapy preferences.

Subsequently, we will compare our findings in detail with the current knowledge. In accordance with other studies that examined the therapy preferences of patients with LC, "life expectancy" was the most important attribute.[11,30–32] This might be due to the shorter life expectancy of patients with LC compared with that of patients with CRC. Another important attribute identified in previous studies was "tumor-associated symptoms".[30] However, previous studies are quite inconsistent in terms of what they consider to be the chief adverse effect of cancer treatment. For example, one study identified fatigue and tiredness as the two attributes of most consequence, whereas another found that the most consequential attributes were nausea and vomiting.[30,31] Both assessed the preferences of patients with (advanced) non-small cell LC, which might have strongly influenced the overall results. In our study, we found that "slightly changed appearance" and "no problems in food intake and digestion" were the attribute levels related to adverse effects that were given the greatest weight by patients with CRC and LC, respectively. This might be explained by the fact that patients with CRC expect to experience disturbances of food intake and digestion.

Few studies have found that sociodemographic characteristics, such as sex or age, do not influence preferences for cancer therapy.[9,30,33] Other studies have reported that sociodemographic characteristics do influence the preferences of patients, but they did not include associations

between therapy preferences and the actual health status of patients.[34,35] Two previous studies did not find a difference in preference based on patients' proximity to death.[32,34,35] However, we observed that having undergone surgery had a noticeable influence on patients' preferences. Therapeutic guidelines recommend surgery at an early disease stage in patients with comorbidities (when the tumor is operable). Therefore, we can assume that patients with a poor prognosis due to LC and a low HRQoL would prefer to increase their length of life when the disease is detected early and surgery is an option.

In summary, our study yielded several novel findings. Patients who were better informed tended to prefer a more balanced relationship between length and quality of life, as compared with less-informed patients. The physicians involved in our study confirmed that they emphasized not only length of life, but also HRQoL as important considerations in their consultations with patients. The influence of physicians on the preferences of patients should be examined in further research. The second finding was that another subgroup (patients with LC and a low HRQoL who had undergone surgery) gave a great weight to increased length of life. Finally, we deduced from patients in Class 3 that those with a relatively good life expectancy (CRC compared with LC) gave a greater weight to moderate effects on HRQoL (physical capacity, appearance) than to a longer life.

However, our study was limited in terms of the unbalanced distribution of patients between the subgroups, which resulted in a small number of patients in Class 3, even though the recruitment period was extended. Furthermore, the results suggest that the online survey was inappropriate for some patients with CRC and LC, particularly patients of advanced age. Alternatively, inappropriate online distributors were used for this process of recruiting older patients. Overall, it appears that older patients were less willing to participate in our study than younger ones. Recruiting patients with LC and CRC at clinics or hospitals might also have biased the study sample, because patients who were not undergoing therapy were excluded. Given that patients were often unaware of their current disease stage or type of chemotherapy (palliative, adjuvant, maintenance), we were unable to include questions concerning this information. It might be possible to estimate disease stage on the basis of self-assessed health and surgery status, although the results can be incomplete or misleading.[36–40] In addition, surgery can be initiated at different disease stages, such as after diagnosis or in the case of disease progression. This means that the "treatment" variable should not be interpreted without further information. Consequently, future studies should obtain patient records to identify any possible associations between stage, therapy goals, and therapy preferences. Although we defined each attribute at the beginning of the questionnaire, we could not control for how patients interpreted the attributes and levels in their own way and as a result of their own disease experiences. However, we would have detected other results if other or further attributes had been included in the DCE tasks. This disadvantage of the DCE method is also discussed in other methodologic publications.[18,41]

The classes identified by LCMLM cannot be accounted for by typical sociodemographic aspects. Therapy should be adjusted to accommodate these three classes. Some class-specific preferences might be accommodated easily (the provision of a guide or shorter waiting times) and might compensate for some of the disadvantages of chemotherapy. Consequently, differences among the classes should be recognized in individual treatment options. This implies that physicians need time to explain and discuss the therapy alternatives with patients. Our findings can be used to develop treatment guidelines and to assess the benefits of pharmaceuticals. However, in accordance with previous studies, the ability to prolong their survival was the most important therapy attribute of a given therapy for patients, irrespective of the cancer type.

Acknowledgments

The authors would like to thank all of the patients who participated in the survey. They would also like to thank the physicians and (study) nurses for recruiting the patients, especially: PD Dr med. Heiko Golpon, Annette Nocon (Hannover Medical School: Department of Pneumology), Prof Dr med. Arndt Vogel and team (Hannover Medical School: Department of Gastroenterology, Hepatology and Endocrinology, Hannover, Germany), Dr med. Parvis Sadjadian (Johannes Wesling Medical Center, Department of Hematology, Oncology, Hemostaseology, and Palliative Care UKRUB, University of Bochum, Minden), Dr med. Monika Heilmann and Mrs J Brandes (Lung Cancer Center at the Hospital Region Hannover), Dr rer. nat. Silvia Wagner and Dr med. Silke Templin (Clinic for Visceral, General and Transplant Surgery Surgical Study Center, University Hospital Tuebingen), Dr med. Michael Gärtner and his team (Ambulatory Oncological Center Hannover), Dr med. Raack and Gabriele Germann (Group Practice for Internal Medicine and Pulmonology, Celle), PD Dr med. Annalen Bleckmann (Interdisciplinary Short-term Oncology, Department of

Hematology and Medical Oncology, Goettingen), Dr med. Detlev Schröder and the study nurse Annette Gieseke (Group Practice for Hematology and Oncology, Hannover).

This study was supported by the Federal Ministry of Education and Research (grant number 01EH1201A).

Author contributions

All authors contributed toward data analysis, drafting and revising the paper and agree to be accountable for all aspects of the work.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*. 2015;136(5):E359–E386.
2. Allemani C, Weir HK, Carreira H, et al; CONCORD Working Group. Global surveillance of cancer survival 1995–2009: analysis of individual data for 25 676 887 patients from 279 population-based registries in 67 countries (CONCORD-2). *Lancet*. 2015;385(9972):977–1010.
3. World Health Organization. Global Action Plan for the Prevention and Control of NCDs 2013–2020. Available from: http://apps.who.int/iris/bitstream/10665/94384/1/9789241506236_eng.pdf?ua=1. Accessed May 26, 2016.
4. Pieterse AH, Baas-Thijssen MCM, Marijnen CAM, Stiggelbout AM. Clinician and cancer patient views on patient participation in treatment decision-making: a quantitative and qualitative exploration. *Br J Cancer*. 2008;99(6):875–882.
5. Pieterse AH, Stiggelbout AM, Baas-Thijssen MC, van de Velde CJ, Marijnen CA. Benefit from preoperative radiotherapy in rectal cancer treatment: disease-free patients' and oncologists' preferences. *Br J Cancer*. 2007;97(6):717–724.
6. Harrison JD, Solomon MJ, Young JM, et al. Patient and physician preferences for surgical and adjuvant treatment options for rectal cancer. *Arch Surg*. 2008;143(4):389–394.
7. Chu DT, Kim SW, Kuo HP, et al. Patient attitudes towards chemotherapy as assessed by patient versus physician: a prospective observational study in advanced non-small cell lung cancer. *Lung Cancer*. 2007;56(3):433–443.
8. Solomon MJ, Pager CK, Keshava A, et al. What do patients want? Patient preferences and surrogate decision making in the treatment of colorectal cancer. *Dis Colon Rectum*. 2003;46(10):1351–1357.
9. Mühlbacher AC, Nübling M. Analysis of physicians' perspectives versus patients' preferences: direct assessment and discrete choice experiments in the therapy of multiple myeloma. *Eur J Health Econo*. 2011;12(3):193–203.
10. Damm K, Vogel A, Prenzler A. Preferences of colorectal cancer patients for treatment and decision-making: a systematic literature review. *Eur J Cancer Care*. 2014;23(6):762–772.
11. Schmidt K, Damm K, Prenzler A, Golpon H, Welte T. Preferences of lung cancer patients for treatment and decision-making: a systematic literature review. *Eur J Cancer Care*. 2015;25(4):580–591.
12. Chue P. The relationship between patient satisfaction and treatment outcomes in schizophrenia. *J Psychopharmacol*. 2006;20(Suppl 6):38–56.
13. Casado JL, Marín A, Romero V, et al. The influence of patient beliefs and treatment satisfaction on the discontinuation of current first-line antiretroviral regimens. *HIV Med*. 2016;17(1):46–55.
14. Hann D, Allen S, Ciambrone D, Shah A. Use of complementary therapies during chemotherapy: influence of patients' satisfaction with treatment decision making and the treating oncologist. *Integr Cancer Ther*. 2006;5(3):224–231.
15. Federal Ministry of Health. The act on the reform of the market for medicinal products (Gesetz zur Neuordnung des Arzneimittelmarktes – AMNOG); 2011. Available from: https://www.bgbl.de/xaver/bgbl/start.xav?startbk=Bundesanzeiger_BGBl&bk=Bundesanzeiger_BGBl&start=//*[@attr_id=%27bgbl110s2262.pdf%27]. Accessed August 6, 2017.
16. Staeck F. Lebensqualität. Forscher und GBA wollen Methoden-Sicherheit. Available from: http://www.aerztezeitung.de/politik_gesellschaft/arzneimittelpolitik/nutzenbewertung/article/925499/lebensqualitaet-forscher-gba-wollen-methoden-sicherheit.html. Accessed July 05, 2017.
17. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmaco Economics*. 2008;26(8):661–677.
18. Mühlbacher A, Johnson FR. Choice experiments to quantify preferences for health and healthcare: state of the practice. *Appl Health Econ Health Policy*. 2016;14(3):253–266.
19. Aumann I, Kreis K, Damm K, Golpon H, Welte T, Graf von der Schulenburg JM. Treatment-related experiences and preferences of patients with lung cancer: a qualitative analysis. *Health Expect*. 2015;19(6):1226–1236.
20. Auspurg K, Liebe U. Choice-Experimente und die Messung von Handlungsentscheidungen in der Soziologie [Choice-experiments and the measurement of behavioral decisions in sociology]. *Köln Z Soziol*. 2011;63(2):301–314. German.
21. Kuhfeld WF. Marketing research methods in SAS. Experimental design, choice, conjoint, and graphical techniques. Available from: http://support.sas.com/techsup/technote/mr2010.pdf. Accessed March 31, 2014.
22. Aaronson NK, Ahmedzai S, Bergman B, et al. The European Organization for Research and Treatment of Cancer QLQ-C30: a quality-of-life instrument for use in international clinical trials in oncology. *J Natl Cancer Inst*. 1993;85(5):365–376.
23. Waldmann A, Schubert D, Katalinic A, Janda M. Normative data of the EORTC QLQ-C30 for the German Population: a population-based survey. *PLoS One*. 2013;8(9):e74149.
24. de Bekker-Grob, Esther W, Donkers B, Jonker MF, Stolk EA. Sample size requirements for discrete-choice experiments in healthcare: a practical guide. *Patient*. 2015;8(5):373–384.
25. European Organisation for Research and Treatment of Cancer (EORTC). EORTC QLQ-C30 Scoring Manual. Available from: http://www.eortc.be/qol/files/SCManualQLQ-C30.pdf. Accessed January 11, 2017.
26. Hauber AB, Gonzalez JM, Groothuis-Oudshoorn CGM, et al. Statistical methods for the analysis of discrete choice experiments: a report of the ISPOR conjoint analysis good research practices task force. *Value Health*. 2016;19(4):300–315.
27. McCulloch CE, Lin H, Slate EH, Turnbull BW. Discovering subpopulation structure with latent class mixed models. *Stat Med*. 2002;21(3):417–429.
28. Zentrum für Krebsregisterdaten, Robert Koch Institut. Darmkrebs [Colon and rectum]. Available from: http://www.krebsdaten.de/Krebs/EN/Content/Publications/Cancer_in_Germany/cancer_chapters_2011_2012/cancer_c18-21.pdf?__blob=publicationFile. Accessed July 26, 2016. German.
29. Zentrum für Krebsregisterdaten, Robert Koch Institut. Lungenkrebs [Lung]. Available from: http://www.krebsdaten.de/Krebs/EN/Content/Publications/Cancer_in_Germany/cancer_chapters_2011_2012/cancer_c33-34.pdf?__blob=publicationFile. Accessed July 26, 2016. German.
30. Mühlbacher AC, Bethge S. Patients' preferences: a discrete-choice experiment for treatment of non-small-cell lung cancer. *Eur J Health Econ*. 2015;16(6):657–670.

31. Bridges JF, Mohamed AF, Finnern HW, Woehl A, Hauber AB. Patients' preferences for treatment outcomes for advanced non-small cell lung cancer: a conjoint analysis. *Lung Cancer*. 2012;77(1):224–231.
32. Kool M, van der Sijp, Joost RM, Kroep JR, et al. Importance of patient reported outcome measures versus clinical outcomes for breast cancer patients evaluation on quality of care. *Breast*. 2016;27:62–68.
33. Laryionava K, Sklenarova H, Heußner P, et al. Cancer patients' preferences for quantity or quality of life: German Translation and validation of the quality and quantity questionnaire. *Oncol Res Treat*. 2014;37(9):472–478.
34. Wright AA, Mack JW, Kritek PA, et al. Influence of patients' preferences and treatment site on cancer patients' end-of-life care. *Cancer*. 2010;116(19):4656–4663.
35. Stiggelbout AM, de Haes, J C, Kiebert GM, Kievit J, Leer JW. Tradeoffs between quality and quantity of life: development of the QQ Questionnaire for Cancer Patient Attitudes. *Med Decis Making*. 1996;16(2):184–192.
36. Goeckenjan G, Sitter H, Thomas M, et al. Prevention, diagnosis, therapy, and follow-up of lung cancer: interdisciplinary guideline of the German Respiratory Society and the German Cancer Society. *Pneumologie*. 2010;64(Suppl 2):e1–e164.
37. German Guideline Program in Oncology. German Cancer Society, German Cancer Aid editors. *Evidenced-Based Guideline for Colorectal Cancer*. Available from http://www.awmf.org/fileadmin/user_upload/Leitlinien/021_D_Ges_fuer_Verdauungs-_und_Stoffwechselkrankheiten/021-007_S3_Colorectal_Cancer_2015_03-extended.pdf. Accessed August 8, 2017.
38. Polanski J, Jankowska-Polanska B, Rosinczuk J, Chabowski M, Szymanska-Chabowska A. Quality of life of patients with lung cancer. *Onco Targets Ther*. 2016;9:1023–1028.
39. Adams SV, Ceballos R, Newcomb PA. Quality of life and mortality of long-term colorectal cancer survivors in the seattle colorectal cancer family registry. *PLoS One*. 2016;11(6):e0156534.
40. Marventano S, Forjaz M, Grosso G, et al. Health related quality of life in colorectal cancer patients: state of the art. *BMC Surgery*. 2013;13(Suppl 2):S15.
41. Coast J, Al-Janabi H, Sutton EJ, et al. Using qualitative methods for attribute development for discrete choice experiments: issues and recommendations. *Health Econ*. 2012;21(6):730–741.

Modul 5

Which attributes of whole genome sequencing tests are most important to the general population? Results from a German preference study

Plöthner, Marika

Schmidt, Katharina

Schips, Clarissa

Damm, Kathrin

Pharmacogenomics and Personalized Medicine 2018;

11, S. 7–21.

Pharmacogenomics and Personalized Medicine

Dovepress

open access to scientific and medical research

Open Access Full Text Article

ORIGINAL RESEARCH

Which attributes of whole genome sequencing tests are most important to the general population? Results from a German preference study

Marika Plöthner
Katharina Schmidt
Clarissa Schips
Kathrin Damm

Leibniz University of Hannover, Center for Health Economics Research Hannover (CHERH), Germany

Objective: The aim of this study was to identify the preferences for whole genome sequencing (WGS) tests without genetic counseling.

Methods: A discrete choice experiment was conducted where participants chose between two hypothetical alternatives consisting of the following attributes: test accuracy, test costs, identified diseases, probability of disease occurrence, and data access. People from the general German population aged ≥18 years were eligible to participate in the survey. We estimated generalized linear mixed effects models, latent class mixed-logit models, and the marginal willingness to pay.

Results: Three hundred and one participants were included in the final analysis. Overall, the most favored WGS testing attributes were 95% test accuracy, report of severe hereditary diseases and 40% probability of disease development, test costs of €1,000, and access to test results for researchers. Subgroup analysis, however, showed differences in these preferences between males and females. For example, males preferred reporting of results at a 10% probability of disease development and females preferred reporting of results at a 40% probability. The test cost, participant's educational level, and access to data influenced the willingness to participate in WGS testing in reality.

Conclusion: The German general population was aware of the importance of genetic research and preferred to provide their own genetic data for researchers. However, among others, the reporting of results with a comparatively relatively low probability of disease development at a level of 40%, and the test accuracy of 95% had a high preference. This shows that the results and consequences of WGS testing without genetic counseling are hard to assess for individuals. Therefore, WGS testing should be supported by qualified genetic counseling, where the attributes and consequences are explained.

Keywords: whole genome sequencing, discrete choice experiment, genetic testing, preferences, willingness to pay, latent class model

Introduction

In the past 10 years, significant progress has been achieved in the fields of genomics and genetics.[1] The usage of genetic information has steadily increased in medical research, diagnosis, and therapy. Essential drivers for this development are as follows: 1) technological progress such as next-generation sequencing (NGS) technologies, 2) the reduction in costs of sequencing,[2] 3) growth in population and clinical-based biobanks,[3] and 4) the increasing knowledge of genotype–phenotype correlations based on genome-wide association studies (GWAS).[4]

Correspondence: Marika Plöthner
Otto-Brenner-Straße 1, 30159 Hannover, Germany
Tel +49 511 7621 4240
Fax +49 511 762 5081
Email mp@cherh.de

submit your manuscript | www.dovepress.com
Dovepress
http://dx.doi.org/10.2147/PGPM.S149803

Genetic information is essential for personalized medicine. This knowledge enables preventive health care management as well as the administration of personalized and targeted therapies based on an individual's genetic characterization.[5] The scope of analysis (gene, panel, exome, or whole genome sequencing [WGS]) and the amount of genetic data vary with the aim of the investigation. WGS provides an opportunity to identify almost all disease-causing variants.[6] For this reason, WGS seems to be the most appropriate method for comprehensive predictive analysis.

In recent years, the suitability of WGS as a screening tool has been discussed, especially in newborn[7] or population-based screening.[8] Notwithstanding the economic (eg, clinical utility),[9] ethical, and legal debates (eg, information of self-determination),[10] the detection of rare and/or highly penetrant diseases before the onset of disease may have considerable advantages. For example, previous surveys indicated that early diagnosis of cystic fibrosis[11] or Lynch syndrome[12] is beneficial for treatment, and the knowledge of predispositions to oncological and cardiovascular diseases can be useful for prevention. Knowledge of a BRCA I/BRCA II mutation allows the development of a prevention strategy including regular checkups and mastectomy.[13]

Several studies showed that people are interested in genetic testing.[14–16] They want to take a proactive role in preventive health care management for themselves as well as for their family members.[17] However, WGS testing aimed at primary prevention without a suspected disease is generally not covered by health insurance plans (eg, in Germany). Genetic analysis distributed via the Internet is a less expensive alternative than the conventional market.[18] Such offers often lack qualified genetic counseling,[19] which is essential for an informed decision regarding WGS testing. Qualified genetic counseling supports complex decision-making with regard to the following questions: Do the results affect my family members? Who has access to my genetic information? What is the potential for genetic discrimination (eg, in terms of insurability)? Am I willing to pay for the testing out-off-pocket? Do I want to know the probability of developing all diseases or only the probability of developing treatable diseases? How sensitive is the test?

For the purpose of identifying relevant attributes of online WGS testing, we conducted a discrete choice experiment (DCE) to evaluate the preferences of the general population. We investigated the people's preference estimates without prior qualified genetic counseling. We analyzed 1) the preferences of our study population and subgroup effects (eg, sociographic characteristics, genetic predisposition, and desire for children), 2) the willingness to pay of these subgroups, and 3) factors influencing the willingness to take part in WGS tests.

Methods

DCE

We conducted a DCE to measure the preferences for WGS testing. A DCE is a de-compositional approach to the measurement of stated preferences. Participants have to choose between hypothetical alternatives. One alternative consists of several attributes with varying levels.[20] The attributes are characteristics of the alternatives that are specified by their levels for each alternative.

Attributes and levels

First, we conducted a literature search to achieve a comprehensive overview of the available attributes of WGS. However, no literature focusing on preferences for WGS attributes could be identified. Hence, we adopted relevant attributes from actual discussions and literature focused on genetic analysis. The final relevant attributes for the DCE were "test accuracy",[21] "test cost",[22] report of results[23–25] (divided into "identified diseases" and "probability of occurrence"), and "access to data".[26] The range of levels was also determined by specific discussion points or based on the literature on the subject. Finally, attributes and levels were discussed with experts. To improve the validity and reliability of each item, a pretest of the questionnaire was conducted with 11 people. Table 1 illustrates the attributes and their corresponding levels. The attributes and levels are explained using colloquial language and icons, and they were adjusted after the pretest.

Data collection and recruitment

People from the German general population aged ≥18 years were eligible to participate in the survey. It was an online survey via Facebook and Xing that was conducted from June to August 2016, as well as by direct (and random) approach of passersby with a paper–pencil questionnaire at the main railway station in the city of Hannover (northwestern Germany). We used a simple random sampling strategy and did not select participants according to age and sociodemographic or economic status. We obtained study approval from the ethics committee of Hannover Medical School (Re No 3325-20016) prior to the start of the survey. To take part in the study, participants had to give written informed consent.

Table 1 Overview of attributes with the corresponding levels

Attribute	Description in the questionnaire	Level 1	Level 2	Level 3	Level 4
Accuracy (sensitivity)	Test accuracy describes the proportion of persons with an identified genetic mutation that actually have this mutation For example, a level of 90% means that 90 of the 100 people really have the risk to develop a certain disease. In contrast, in 10 of the 100 people, a disease risk is identified because of inaccuracy of the test, although they do not have this risk You can choose between different tests with different accuracy values	90% 	95% 	99% 	
Identified diseases	You can choose about the test results you want to be informed You can choose the test results that you want to be informed about. You have the choice between reporting of all test results, only treatable diseases (preventive and therapeutic treatments), and serious hereditary diseases In case of serious hereditary diseases, it is assumed that these are inherited with a high probability and are characterized by a serious disease progression	All diseases	Treatable disease	Serious hereditary disease	
Test costs	A WGS is an innovative, diagnostic instrument and currently associated with high execution costs. You should decide how much money you are willing to pay for this comprehensive genetic analysis	€500 	€1,000 	€1,500 	
Probability of occurrence	The results of a WGS determine the risk of being affected by a specific disease. A genetic mutation enables statements about the probability of developing different diseases. You can decide which probability of developing a disease you want to be informed	10%	40%	70%	
Access to data	WGS is associated with a large amount of personal data. You can decide who can get access to your test results in addition to you and your treating physician For example, you can make your genetic data accessible to researchers and thus contribute to medical research	No one else 	Insurer 	Researcher 	Insurer and researcher 

Abbreviation: WGS, whole genome sequencing.

Questionnaire

The final questionnaire consisted of three sections. The first part was the DCE choice sets. In total, the attributes and levels resulted in $3^4 \times 4^1 = 324$ possible combinations (four attributes with three levels and one attribute with four levels).[20] To generate feasible choice sets of the DCE, a *D*-efficient fractional factorial design (reduced design) was created using the R statistical program. The best *D*-efficiency occurred for 18 choice sets. To avoid overstraining of the participants, we divided the 18 choice sets into two questionnaires (blocking). Therefore, participants answered nine DCE decisions with two alternatives (called Test 1 and Test 2) each. Additionally, we asked whether the participant would carry out the chosen test in reality (refer the example of the choice in Figure 1). The second part focused on sociodemographic questions, such as sex, age, education, occupation, monthly net income, and insurance company (statutory or private). The third part included questions about overall health status, prevention behavior, hereditary diseases, and desire for children.

Data analysis

Following survey completion, we cleaned the data set and determined descriptive statistics for the variables (median, standard deviation [SD], and percentages). We tested the potential independent variables for multicollinearity to reduce the bias of the results. In the multivariate analyses, we applied generalized linear mixed-effects models (GLMMs) and latent class mixed logit models (LCMLMs) to identify systematic or group differences for the participants' WGS preferences. The choice of an alternative between two hypothetical WGS

	Test 1	Test 2
Test accuracy How many people are to be identified who actually have the disease risk?	 95%	 99%
Identified diseases Which test results you want to be informed?	Treatable diseases	Serious hereditary diseases
Test costs How much money you are willing to pay for this comprehensive genetic analysis.	 € 1,500	 € 500
Probability of occurrence Which probability of developing potential diseases you want to be informed?	10%	70%
Access to data Who can get access to your test results in addition to you and your treating physician?	 Insurer	 No one else

Which test would you choose?

☐ Test 1

☐ Test 2

Would you carrry out the chosen test under the given condition also in reality?

☐ Yes

☐ No

Figure 1 Example of a choice set.

Notes: Explanation for the example choice set: The participant could choose between test 1 and test 2. Test 1 is characterized by a lower test accuracy (95%), with the reporting of treatable results at a 10% probability of disease occurrence as well as higher cost (€1,500), and the access for insurer. Test 2 is designed with a higher accuracy (99%), with the reporting of serious hereditary diseases at a higher probability of disease occurrence (70%) and at lower cost (€500). Furthermore, in test 2, no one else had access to the test results. The participant has to trade-off between a test accuracy of 95 and 99%, the costs of €1,500 and €500, and so on.

tests (choice) was used as the dependent variable, whereas the attributes and levels were the independent variables in all models. In addition, personal characteristics of the participants were used as independent variables, mixed effects (taking into account that personal characteristics influence the response behavior and therefore including subgroup specific "baseline" values [random intercept] or slope adjustments [random slope] for some of the independent variables in addition to the fixed effects), or class-membership effects (for LCMLM). We calculated the average marginal willingness to pay (mWTP) for each attribute by dividing the coefficients for the other attributes by the coefficient of the cost attribute (test costs). Therefore, we used the attributes as metric independent variables in conditional logit models and conducted the mWTP analysis separately for the different classes from the LCMLM analyses. Coefficients of attributes above zero were favored, and negative coefficients were disfavored. The 95% confidence intervals (CIs) are based on the Krinsky and Robb[27] method.

We calculated the GLMM for participants willing to participate in reality (potential users) and the full sample separately, so that any differences between these two groups could be identified. In the GLMM, we used the set ID (identification number of the choice set) as a mixed effect to

inform the model about which of the alternatives formed a set. Finally, we investigated the factors influencing the willingness to participate in genetic testing in reality. Therefore, we applied another GLMM based on the variable "real" as a dependent variable. The random effect used in this model was the person identifier (PersonID) to enable us to investigate influencing participants' characteristics and test characteristics based on the decision. An overview of used variables is provided in Table S1.

We tested different independent variables and mixed effects in the models (Table S2) and chose the model with the best fit for data based on Akaike and Bayesian information criteria. All analyses were conducted with R statistics 3.1.2 and the packages "lme4" (for GLMM), "lcmm" (for LCMLM), and "support.CEs" (for mWTP analyses).

Results

Descriptive statistics

In total, 323 people participated in the study and 301 people could be included in the DCE analyses. All sample characteristics are provided in Table 2. Twenty-two participants had to be excluded because of missing data for all DCE tasks or an age of <18 years. The sample consisted of 69% women, and the median age was 28 years. The educational level was higher compared to that of the general population of Germany,[28] but the average amount of income was similar.[29] Both facts indicated that the proportion of students was higher compared to the general population. The majority (56%) of the participants were in good health.

In a second step, we prepared the data for the multivariate analyses. We found strong correlations between age and employment status, having children and employment status, and age and desire to have a child (refer correlation plot in Figure S1). Therefore, we adapted the models for these correlations due to not using both correlating variables in one model or due to including interaction effects between the correlating variables.

Table 2 Sample description

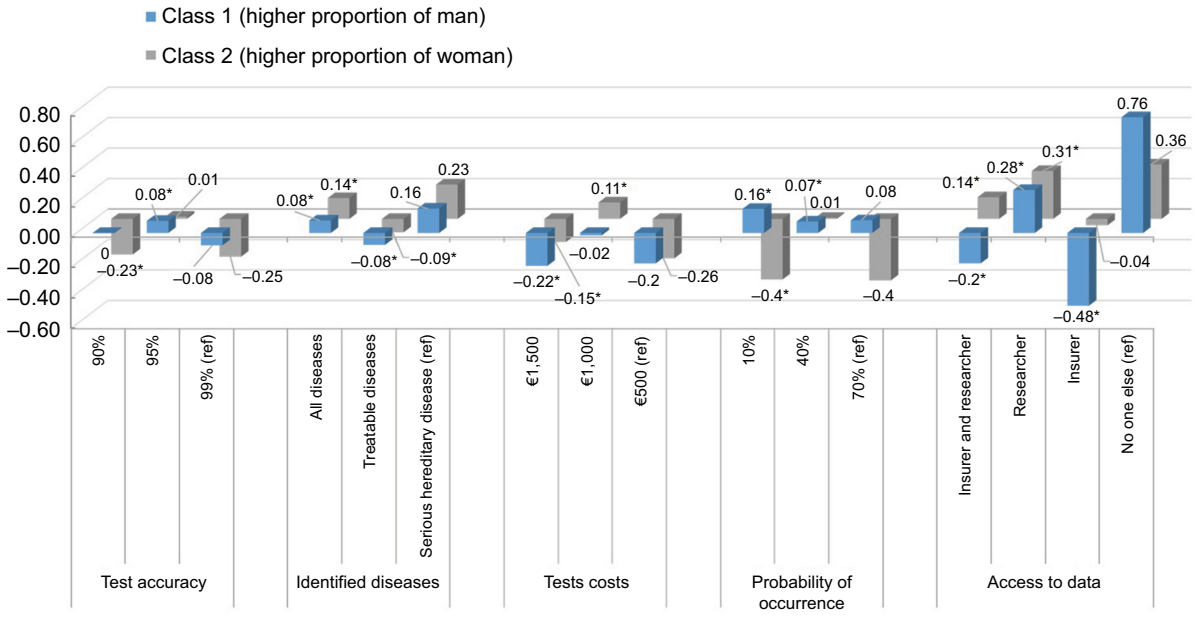
Variable	Occurrence in the sample
Participants (number)	323
With at least one valid DCE task	301
Sex (% women)	69
Age in years (median, SD)	28 (13.86)
Own children (% having at least one child)	41
Desire to have children (%)	
Yes	50
No	39
Unsure	11
Highest level of education (%)	
No graduation	1
Primary school	6
Secondary school	34
High school	24
University	34
Income (%)	
No own income (€)	16
<1,000	27
1,000–<2,000	29
2,000–<3,000	17
3,000–<4,000	6
≥4,000	4
Participation in screening program (%)	
Never	51
Every 10 years	3
Every 5 years	9
Every 2 years	21
1–2 times a year	15
Subjective health status (%)	
Very bad	0
Bad	4
Medium	24
Good	56
Very good	16
Hereditary diseases in the family (% yes)	20
Afraid of hereditary diseases (% yes)	21

Note: Median: average.
Abbreviations: DCE, discrete choice experiment; SD, standard deviation.

Subgroup-specific preferences for WGS tests

In the LCMLM, we identified two classes that differed in regard to their preferences for genetic testing (Figure 2 and Table S3). Class 1 comprised 46.13% (n=137) of the sample. The only significant differentiator between the people in the two classes was their sex. The proportion of women was significantly lower in class 1 than in class 2 (refer the table in Figure 2). The educational level, health status, and income are relevant for the class membership but did not show significant differences between the classes.

In class 1, a higher proportion of men compared to the other classes strongly preferred the restricted "access to data only for themselves" ($\beta_{\text{class 1,access no}}$=0.76, reference level) and disfavored the "access to data for insurer" the most ($\beta_{\text{class 1,access ins}}$=–0.48, $P < 0.001$). They also disfavored any "test costs" where €1,000 had a utility weight of ~0 but was not significant. Class 1 preferred "serious hereditary diseases identified" and a "10% probability of occurrence" ($\beta_{\text{class 1, ser.d}}$=0.16, $\beta_{\text{class 1, 10% occ}}$=0.16, P<0.001) (Figure 2). In contrast, class 2 disfavored "10% and 70% probability of occurrence" but also preferred "serious hereditary diseases



Graph adjusted for further effects: mixture = –Att_TA + Att_DIS + Att_TC + Att_ACC, random = ~seti, subject= "personID", classmb= ~sex + EDL + INCn + HSn

Class-membership effects	Class 1 (ref = class 2)		
	Coefficient	Standard error	*P*-value
Intercept	1.93	0.94	0.04
Sex (ref = male)	–0.64	0.29	0.03
Educational level	0.09	0.14	0.54
Health status	–0.28	0.19	0.14
Income	–0.09	0.11	0.40

Figure 2 LCMLM for preferences concerning genetic testing – attribute effects.
Note: *Significant values ($P<0.05$).
Abbreviations: EDL, educational level; HSn, health status (numeric); INCn, income (numeric); LCMLM, latent class mixed logit model.

identified". Indeed, the highest preferences occurred for access to data only for themselves and "for researchers" ($\beta_{\text{class 2, access no}}$=0.36, reference level; $\beta_{\text{class 2, access res}}$=0.31, $P<0.001$). Class 2 also preferred "access to data only for insurer and researcher". Class 2 disfavored "90% and 99% test accuracy" and showed a significant positive utility for "€1,000 test costs".

To conclude, men emphasized the importance of access to data only for themselves and favored a test with 95% accuracy also for diseases with a low probability of occurrence. The class with a higher proportion of women favored instead a test that identifies serious hereditary diseases, where test costs on the intermediate level arise, and that enables data access for themselves or researchers.

In addition, we calculated the mWTP for each attribute, separated for class 1 and class 2 from the LCMLM (Table 3). The mWTP showed different starting points for class 1 and class 2 models (intercept$_{\text{class 1}}$: €786.3 and intercept$_{\text{class 2}}$: €–1,931.3). From this, it can be concluded that people in class 2 were willing to pay less money for genetic testing than those in class 1. Furthermore, class 2 was willing to pay on average €740 for an increase of one unit (90%–95% or 95%–99%) in test accuracy (CI: €489.5; €1,218.2) and on average €1,500 (€1,071.5; €2,435.5) for diseases with higher probability of occurrence. In contrast, the mWTP was negative for the identified diseases (€–303.7 [€–560.2; €–127.1]) and the access to data (€–383.8 [€–645.3; €–228.7]). Therefore, people were willing to receive monetary compensation for identifying only treatable and hereditary diseases. Class 1 was willing to pay on average less for a higher test accuracy, although the monetary value was still positive (intercept €786–128=€658 for a change from 90% to 95%). In addition, this class showed negatively associated mWTP for identified diseases (€–164.6 [€–289.7; €–45.1]) and the probability of occurrence (€–502.3 [€–707.4; €–356.8]). In contrast, class 1 was willing to pay ~€723 [€561.2; €967.9] more for less access to data.

Table 3 Marginal willingness of classes to pay for test attributes

Attribute	Levels	Class 1: mWTP in € (95% CI)	Class 2: mWTP in € (95% CI)
Intercept		786.3 (308.5; 1,233.9)	–1,931.3 (–3,935.2; –905.2)
Test accuracy	90%–99%	–127.6 (–258.7; –17.9)	737.8 (489.5; 1,218.2)
Identified diseases	All, treatable, hereditary	–164.6 (–289.7; –45.1)	–303.7 (–560.2; –127.1)
Probability of occurrence	10%–70%	–502.3 (–707.4; –356.8)	1,514.5 (1,071.5; 2,435.5)
Access to data	Insurer, researcher and insurer, researcher, no one else	722.9 (561.2; 967.9)	–383.8 (–645.3; –228.7)

Note: Class 1: higher proportion of men; Class 2: higher proportion of women.
Abbreviations: CI, confidence interval; mWTP, marginal willingness to pay.

Analysis of participation in genetic testing

We estimated GLMMs (full sample, potential users) to identify the preferences for genetic testing. The most important attribute level for genetic testing for both subgroups was the "identification of severe hereditary diseases" (Table S4). Therefore, this attribute level is more important for potential users ($\beta_{user,ser.dis.}$=0.88) than for the full sample ($\beta_{full,ser.dis.}$=0.49). However, the most disfavored attribute level for both subgroups was access to data for insurer ($\beta_{full,insur}$=–0.81, $\beta_{user,insur.}$=–0.64, both P<0.001). It is striking that for test accuracy, identified diseases, test costs, and probability of occurrence, the intermediate level gained the highest utility weight in both subgroups. Although the preferences were similar between the subgroups, the full sample preferred "95% test accuracy", €1,000 test costs, and "access to data for researchers" more strongly than the potential user subgroup.

In the last step, we investigated the factors that influenced the willingness of respondents to participate in genetic testing in reality or if they just preferred the chosen alternative hypothetically. The GLMM showed that from the attributes, only test accuracy and access to data were relevant for the decision (Table 4). All costs reduced the willingness to participate in genetic testing; however, €500 was the least disfavored level ($\beta€_{500}$=–0.024). In addition, people were more willing to participate when the access to data would be denied to insurers and researchers. In contrast to previous models, the decision to participate in reality was positively influenced by access to data for researchers and not "only for themselves". Educational level showed a negative association to the participation in genetic testing. In addition, people who would participate in screenings if the social or private health insurance (SHI) subsidized it were more willing to participate in genetic testing ($\beta_{scr\ subs\ SHI}$=1.86, P<0.001). "Employment status", "income", and "fear of genetic diseases" did not show significant results, although the direction of the coefficients was as expected.

Table 4 GLMM fixed-effects results for participation in genetic testing

Variables	Levels	Coefficient	SE	*P*-value
Test costs	€1,500	–0.261	0.100	0.009
	€1,000	–0.237	0.090	0.009
	€500 (ref)	–0.024		
Probability of occurrence	10%	–0.089	0.101	0.375
	40%	–0.012	0.094	0.897
	70% (ref)	–0.077		
Access to data	Insurer and researcher	–0.275	0.118	0.019
	Researcher	0.097	0.106	0.358
	Insurer	–0.349	0.134	0.009
	No one else (ref)	–0.024		
Educational level		–0.693	0.263	0.008
Employment status		–0.858	0.541	0.113
Income		0.338	0.226	0.134
Screening utilization: subsidy by SHI		1.857	0.465	0.000
Afraid of genetic diseases		0.975	0.564	0.084

Notes: Intercept coefficient 1.409; SE 1.231; *P* 0.252 and random intercept PersonID variance 9.765; standard deviation 3.125.
Abbreviations: GLMM, generalized linear mixed-effects model; SE, standard error; SHI, social or private health insurance.

Main findings

The most preferred test for the overall sample was characterized by the following aspects: 1) the test accuracy of 95%, 2) report of severe hereditary diseases, 3) the test cost of €1,000, 4) report of results for diseases with a probability of occurrence from 40%, and (5) access to genome data for researcher but not for insurers (Table S4). Except for "access to genome data", all intermediate levels achieved the highest utility weights in both the full sample and the sample of potential users (Table S3).

Discussion

In this study, the preferences for WGS testing without qualified genetic counseling were assessed.

The test accuracy of 95%, especially sensitivity in this case, was the most favored level of this attribute. This may show that the participants did not understand (or only partly understood) the underlying concept of test sensitivity and

false-positive results. We expected that the most preferred level would be 99% test accuracy. False-positive findings lead to anxiety and uncertainty for the tested person as well as for their families.[30] This in turn may require an additional diagnostic clarification or leads to an increased treatment demand (eg, psychological counseling). Finally, false-positive results could cause an unnecessary rising cost for the statutory health insurance. Otherwise, the participants may understand the underlying concept but accept the uncertainties to receive other advantages, eg, lower test costs.

The amount of reported results was also an important aspect for the decision regarding WGS tests. This aspect is represented by the probability of occurrence (in this experiment 10%, 40%, or 70%) as well as by the kinds of reported diseases (all disease dispositions, only treatable [potential] disorders, or only severe hereditary diseases). The majority of the participants preferred the reporting of serious hereditary diseases. "All disease dispositions" were not attributed with the highest utility score; this may be in accordance with the aspects of efficiency and evidence. Technological progress and genetic research enables the detection of a majority of diverse gene variants. However, many identified genetic variations are not assigned to phenotypes, or the interaction of the specific gene variants is actually unknown.[31] This may change in the future because of further genomic research, especially through GWAS. So far, there are no therapy options for most of the identified gene variants and diseases. However, the participants preferred 40% "probability of disease occurrence". This may indicate that the general population cannot assess the absolute risks for developing a disease without counseling or the influence on disease development caused by lifestyle changes (e.g., sports, nutrition), or that prevention measures may be assessed as a more important and changeable factor. These preferences could occur because of unawareness about genetic risk factors of the participants, due to lack of qualified counseling, or because of their risk aversion. Another limiting factor could be the three given levels of the probabilities. Since the participants were forced to prefer one of the given levels, the range of the outcomes could also be limited. However, the first explanation is emphasized by the negative effect of educational level on the willingness to participate (Table 4).

Cost reduced the willingness to participate in the WGS testing in reality (Table 4). Accordingly, subsidies by SHI for WGS testing showed a positive effect on the willingness to participate in testing. However, €1,000 received the highest approval in the LCMLM. This may be due to the association between the rising costs and the quality or the knowledge of the "$1,000 genome", which means the often discussed cost reduction of a WGS to $1,000 in recent years.[32] Otherwise, health care systems with little or no out-of-pocket payments for prevention measures could influence the importance of cost attributes for the participants' decisions. However, the participants' income did not influence the class membership and preferences. In the mWTP analyses, we found that the willingness to pay in class 2 (higher proportion of women) was highest for the attribute of probability of disease occurrence, whereas the highest mWTP occurred for access to data in class 1 (higher proportion of men). Furthermore, the direction of mWTP for several attributes was different for these two classes. Thus, the mWTP seemed highly dependent on the examined subgroup. The formation of class 1 (higher proportion of women) and class 2 (higher proportion of men) highlights the differences between males and females. While males preferred restricted access to data only for themselves, females wanted to make their genetic data accessible to research. Secrecy of personal data is seemingly very important to men, while women may want to contribute to genetic research. Further differences arose in reporting of results. Females and males preferred a reporting of results at a 40% and 10% probability of disease occurrence, respectively. Fear of a variety of predictive findings (women) or the desire to know almost all dispositions (men) may be possible explanations for this finding.

In the future, cost reductions will be expected because of the focus on genetic analyses of specific variants. Currently, for example, in the case of presumed heredity of breast cancer, the first-degree-relative risk patients are often tested only for the specific variant (eg, BRCA I and BRCA II).[33] Further improvements in WGS testing could contribute to it becoming the favorable alternative compared to panel or single gene sequencing.

Potential users as well as the full sample rejected the access of test results to insurance agencies. Fear of genetic discrimination, eg, in terms of insurability or direct and/or indirect risk selection, seems to be particularly substantial.[34] However, due to a ban on discrimination and the obligation to contact, this risk is excluded in the statutory health insurance in Germany. In other insurance areas (private health insurance, life insurance, and occupational disability insurance), these data could have a stronger influence on insurability and insurance premium, which may lead to uncertainty and anxiety. Despite the strong regulations, anxiety and fear of data misuse seem to be the sensitive issues. Further research is needed in these areas. However, the DCE results suggested that potential users preferred to give researchers access to

genetic data. Genetic research is a dynamic field, and comprehensive genetic databases are the prerequisite for research. The fear of disease as well as the interest in research and further medical developments may be essential drivers for the preferences in this study. Thus, people have the opportunity to contribute to medical research. With regard to large genome sequencing projects, such as the 100,000 Genomes Project (UK),[35] the Saudi Human Genome Program (Saudi Arabia),[36] and the GoNL (the Netherlands),[37] the German population also showed interest. The reporting of test results could be restricted or completely rejected in qualified WGS testing, eg, to findings of the ACMG-positive list (Recommendations for Reporting of Incidental Findings in Clinical Exome and Genome Sequencing).[38] Basically, the decision for or against a WGS test in reality depended on the specific design (characteristics level) in 53.26% of the cases. While 26% of the participants rejected a WGS test independent of specific levels, 20.74% of the participants would execute a WGS test independent of the test characteristics in reality.

The possibilities for using genetic testing results in diagnosis and therapy have steadily increased. Therefore, the WGS offers an opportunity to detect a majority of disorders, especially using a predictive approach. However, in Germany, the costs of genetic analyses for patients at risk (eg, first-degree relatives of breast cancer patients) are covered by a variety of health insurance plans, whereas predictive genetic testing for nonpredisposed people is an out-of-pocket expense. Therefore, comprehensive genetic direct-to-consumer (DTC) analysis via the Internet seems to be a less expensive alternative,[18] although DTC options often lack qualified genetic counseling.[19] As we can see from our survey, not all stated preferences are consistent with the qualified recommendations. Therefore, our study results emphasize the importance of genetic counseling. In Germany, human genetic counseling for predictive analysis is obligatory in accordance with the § 10 German Act of Gene Diagnostics (GenDG). Two main results underline the claim for genetic counseling: 1) the chosen test accuracy of 95% and the associated higher risk of false-positive results (in contrast to a test accuracy of 99%) and 2) the selected probability of disease occurrence at a level of 40% for the reporting of results. For a majority of disease dispositions, there are no treatment options at the moment. Therefore, people may be confronted with information on a large number of potential diseases, which will lead to anxiety. Genetic counseling may help to understand what penetrance really means and which consequences of a finding with a probability of 40% occurrence will arise. However, a possible explanation for these preferences might be that people assume that their doctors will receive the WGS test results and help them to understand and interpret their results. The attribute access to data is characterized by the possibility of access to the genetic information by the treating physician. Due to medical secrecy, we excluded the risk and the anxiety of data misuse. A person can decide if they want to share these genetic results with the treating physician, which would be beneficial for understanding. Prior genetic consultations may have an influence on the general decision for the execution and the scope of reporting of the results. However, in the present study, we excluded such a prior consultation to explore the preferences without a qualified genetic counseling (which is partially lacking in a genetic DTC analysis).

One limitation of this experiment is the hypothetical character. The revealed preferences may lead to another distribution of utility weights. Furthermore, the importance of test specificity was neglected. The difference between sensitivity and specificity is difficult for the general population to understand, and therefore, we focused on test sensitivity in the DCE. The representativeness of the sample is also limited. The sample of a primarily online acquisition is mainly characterized by younger and Internet-savvy people. However, we assumed that the topic is most relevant for this group. In the direct approach, we only recruited a small number of participants (n<10), so we could exclude a selection bias. Although we included the relevant test attributes and important sociodemographic characteristics of the study population, further factors (eg, risk aversion) could influence the preferences. The calculations of mWTP should be considered with caution. We treated the level differences as linear, although this is not intuitive. For example, we assumed that the difference from 90% test accuracy to 95% had the same effect as a change from 95% to 99% in mWTP. However, we needed to assume linear effects for calculating the average willingness to pay and show differences between the classes. At the time of our study, there was a lack of literature describing the levels used for the attributes. Therefore, we considered the available literature and current discussion to derive the characteristics of the attributes. These data were discussed and approved by experts. Having a published qualitative study available would have led to a higher objectification of attribute and level selection. However, due to the short duration of the study, we had to forgo this possibility. In order to assess the relevance of the test conditions for nontest-savvy participants, an integration of an opt-out option was omitted. The study can be considered a feasibility study based on the number of participants. To extrapolate the results to the

whole country, the number of participants needs to be larger and nationally representative.

This study reports on the interest and preferences for WGS testing among Germans. Our study sample from the general population of Germany was aware of the importance of WGS results, and they preferred to make their data accessible for researchers but not for insurers because of possible discrimination. A positive attitude toward population-wide screening projects could therefore be assumed if data privacy is assured and the costs do not exceed €1,000. In general, the decision for or against a WGS is complex and could have far-reaching consequences. Hence, this decision should be a result of an informed consent process, where the attributes and consequences of a WGS are clarified.

Acknowledgments

The publication of this article was funded by the Open Access Fund of the Leibniz Universität Hannover. We thank all study participants and thank Editage for assistance in language editing.

Disclosure

The authors report no conflicts of interest in this work.

References

1. Heard E, Tishkoff S, Todd JA, et al. Ten years of genetics and genomics: what have we achieved and where are we heading? *Nat Rev Genet*. 2010;11(10):723–733.
2. Bentley DR, Balasubramania S, Swerdlow HP, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008;456(7218):53–59.
3. Vora T, Thacker N. Impacts of a biobank: bridging the gap in translational cancer medicine. *Indian J Med Paediatr Oncol*. 2015;36(1):17–23.
4. Rockman MV. Reverse engineering the genotype-phenotype map with natural genetic variation. *Nature*. 2008;456(7223):738–744.
5. Verma M. Personalized medicine and cancer. *J Pers Med*. 2012;2(1):1–14.
6. Meienberg J, Bruggmann R, Oexle K, Matyas G. Clinical sequencing: is WGS the better WES? *Hum Genet*. 2016;135(3):359–362.
7. Bodian DL, Klein E, Iyer RK, et al. Utility of whole-genome sequencing for detection of newborn screening disorders in a population cohort of 1,696 neonates. *Genet Med*. 2016;18(3):221–230.
8. Cho MK. Preventive genomic sequencing in the general population: do PGS fly? *Am J Bioeth*. 2016;15(7):1–2.
9. Ellingford JM, Barton S, Bhaskar S, et al. Whole genome sequencing increases molecular diagnostic yield compared with current diagnostic testing for inherited retinal disease. *Ophthalmology*. 2016;123(5):1143–1150.
10. EURAT Group. Cornerstones for an ethically and legally informed practice of Whole Genome Sequencing: Code of Conduct and Patient Consent Models: Project EURAT "Ethical and Legal Aspects of Whole Human Genome Sequencing". Available from: http://www.uni-heidelberg.de/md/totalsequenzierung/informationen/mk_eurat_position_paper.pdf. Accessed March 15, 2017.
11. Farrell PM, Kosorok MR, Rock MJ, et al. Early diagnosis of cystic fibrosis through neonatal screening prevents severe malnutrition and improves long-term growth. Wisconsin Cystic Fibrosis Neonatal Screening Study Group. *Pediatrics*. 2001;107(1):1–13.
12. Strafford JC. Genetic testing for Lynch Syndrome, an inherited cancer of the bowel, endometrium, and ovary. *Rev Obstet Gynecol*. 2012;5(1):42–49.
13. Petrucelli N, Daly MB, Pal T. BRCA1- and BRCA2-associated hereditary breast and ovarian cancer. In: Pagon RA, Adam MP, Ardinger HH, et al., editors. *GeneReviews(R)*. Seattle, WA: University of Washington, Seattle; 1993.
14. Vermeulen E, Henneman L, van El CG, Cornel MC. Public attitudes towards preventive genomics and personal interest in genetic testing to prevent disease: a survey study. *Eur J Public Health*. 2014;24(5):768–775.
15. Sanderson SC, Wardle J, Jarvis MJ, Humphries SE. Public interest in genetic testing for susceptibility to heart disease and cancer: a population-based survey in the UK. *Prev Med*. 2004;39(3):458–464.
16. Wilde A, Meiser B, Mitchell PB, Hadzi-Pavlovic D, Schofield PR. Community interest in predictive genetic testing for susceptibility to major depressive disorder in a large national sample. *Psychol Med*. 2011;41(8):1605–1613.
17. Berg S. The well-informed patient: a new breed of health care consumer. *Asthma Mag*. 2005;10(4):28–30.
18. Altman RB. Direct-to-consumer genetic testing: failure is not an option. *Clin Pharmacol Ther*. 2009;86(1):15–17.
19. Gollust SE, Wilfond BS, Hull SC. Direct-to-consumer sales of genetic services on the Internet. *Genet Med*. 2003;5(4):332–337.
20. Lancsar E, Louviere J. Conducting discrete choice experiments to inform healthcare decision making: a user's guide. *Pharmacoeconomics*. 2008;26(8):661–677.
21. Goldfeder RL, Priest JR, Zook JM, et al. Medical implications of technical accuracy in genome sequencing. *Genome Med*. 2016;8(1):24.
22. National Human Genome Research Institute. The Cost of Sequencing a Human Genome. Available from: https://www.genome.gov/sequencingcosts/. Accessed July 6, 2016; March 13, 2017.
23. Green ED, Guyer MS, National Human Genome Research Institute. Charting a course for genomic medicine from base pairs to bedside. *Nature*. 2011;470(7333):204–213.
24. Johnson KJ, Gehlert S. Return of results from genomic sequencing: a policy discussion of secondary findings for cancer predisposition. *J Cancer Policy*. 2014;2(3):75–80.
25. McLaughlin HM, Ceyhan-Birsoy O, Christensen KD, et al; MedSeq Project. A systematic approach to the reporting of medically relevant findings from whole genome sequencing. *BMC Med Genet*. 2014;15:134.
26. van Hoyweghen I, Horstman K. European practices of genetic information and insurance: lessons for the Genetic Information Nondiscrimination Act. *JAMA*. 2008;300(3):326–327.
27. Krinsky I, Robb AL. On approximating the statistical properties of elasticities. *Rev Econ Stat*. 1986;68(4):715–719.
28. German Federal Statistical Office. Bildungsstand der Bevölkerung. Available from: https://www.destatis.de/DE/Publikationen/Thematisch/BildungForschungKultur/Bildungsstand/BildungsstandBevoelkerung5210002167004.pdf?__blob=publicationFile. Accessed March 15, 2017.
29. German Federal Statistical Office [webpage on the Internet]. Einkommen und Einnahmen sowie Ausgaben privater Haushalte (Laufende Wirtschaftsrechnungen): Deutschland, Jahre, Haushaltsgröße. Available from: https://www-genesis.destatis.de/genesis/online;jsessionid=D950CB1AF32F174301847A1DC5FADC47.tomcat_GO_2_1?operation=previous&levelindex=2&levelid=1484292140539&step=2. Accessed March 15, 2017.
30. Institute of Medicine (US) Committee on Assessing Genetic Risks. *Assessing Genetic Risks: Implications for Health and Social Policy*. Washington, DC: National Academies Press (US); 1994.
31. Petersen BS, Fredrich B, Hoeppner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet*. 2017;18(1):14.
32. Hayden EC. Technology: the $1,000 genome. *Nature*. 2014;507(7492):294–295.
33. National Cancer Institute [webpage on the Internet]. BRCA1 and BRCA2: Cancer Risk and Genetic Testing. Available from: https://www.cancer.gov/about-cancer/causes-prevention/genetics/brca-fact-sheet. Accessed March 16, 2017.

34. Lemke T. "A slap in the face". An exploratory study of genetic discrimination in Germany. *Genomics Soc Policy*. 2009;5:22–39.
35. Department of Health [webpage on the Internet]. The 100,000 Genomes Project. Available from: https://www.genomicsengland.co.uk/the-100000-genomes-project/. Accessed March 15, 2017.
36. Project Team SG. The Saudi Human Genome Program: an oasis in the desert of Arab medicine is providing clues to genetic disease. *IEEE Pulse*. 2015;6(6):22–26.
37. Boomsma DI, Wijmenga C, Slagboom EP, et al. The genome of the Netherlands: design, and project goals. *Eur J Hum Genet*. 2014;22(2):221–227.
38. Green RC, Berg JS, Grody WW, et al. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*. 2013;15(7):565–574.

Supplementary materials

Table S1 Overview of used variables

Topics	Variable	Meaning	Explanation	Characteristics	Type
DCE-specific variables	Questionnaire				
	Set				
	Seti		Questionnaire combined with set		
	Alternative			1	
				2	
	Choice			0: no	
				1: yes	
	Realn	Real decision (numeric)	Would you also choose the chosen alternative in reality?	0: no 1: yes	Numeric
Attributes	Att_TA	Test accuracy	Test accuracy	1: 90% 2: 95% 3: 99%	
	Att_DIS	Identified diseases	Test results	3: all 2: treatable diseases 1: serious hereditary disease	
	Att_TC	Test costs	Test costs	3: €1,500 2: €1,000 1: €500	
	Att_PROB	Probability of occurrence	Probability of occurrence of disease	1: 10% 2: 40% 3: 70%	
	Att_ACC	Access to data	Access to data	4: insurer and researcher 3: researcher 2: insurer 1: no one else	
Sociodemographic aspects	PersonID	Person identifier			
	Sex	Sex		1: male 2: female	Binary
	Age	Age			Numeric
	EDL	EDL	Highest level of education	0: no graduation 1: primary school 2: secondary school 3: high school 4: university	Numeric
	ES	ES		0: nonemployed 1: in training/student 2: employed/self-employed	Numeric
	INCn	INCn		0: no own income 1: <€1,000 2: €1,000–<€2,000 3: €2,000–<€3,000 4: €3,000–<€4,000 5: ³€4,000	Numeric
Health insurance and utilization of screening	SHI	Insurance		1: statutory 2: private	Binary
	PSC	PSC program		1: 1–2 times the year 2: every 2 years 3: every 5 years 4: every 10 years 5: never	Numeric
	PSChin	PSC program at full-cost coverage by health insurance		0: no 1: yes	Numeric
	PSCshare_r	PSC if health insurance pays a share	Recoded variable if Kostzu =1 or Kostal =1 then Kostzu_r =1	0: no 1: yes	Binary

(*Continued*)

Table S1 (*Continued*)

Topics	Variable	Meaning	Explanation	Characteristics	Type
	PSCsharen	PSC if health insurance pays a share (numeric)		0: no 1: yes	Numeric
	PSCpocketn	PSC on own payment (numeric)		0: no 1: yes	Numeric
Health status and diseases	HSn	Subjective HSn		1: very bad 2: bad 3: medium 4: good 5: very good	Numeric
	FHD	Known FHD		0: no 1: yes	Binary
	FHDfree	Open questions to hereditary diseases in the family		Free text	Free text
	CHIn	CHIn		0: no 1: yes	Binary
	DCHIn	DCHIn		0: no 1: I do not know 2: yes	Numeric
	AFHD	AFHD		0: no 1: yes	Numeric
	AFHDfree	Fear of which hereditary disease		Free text	Free text

Abbreviations: AFHD, afraid of hereditary disease; CHIn, children (numeric); DCHIn, desire to have children (numeric); FHD, family hereditary disease; EDL, educational level; ES, employment status; HSn, health status (numeric); INCn, income (numeric); PSC, participation in screening; SHI, social or private health insurance.

Table S2 Overview of included independent variables used in GLMM and LCMLM

Model	Dependent variable	Independent variables tested	Mixed effects	Lean model
GLMM (for both participants and full-sample)	Choice	Att_TA + Att_DIS + Att_TC + Att_PROB + Att_ACC, ES × EDL, KF, AFHD, CHI, DCHI, SE, HSn, PSC	PersonID, serial, Set, Seti, age, sex, EDL, ES	Wahl ~ Att_TA + Att_DIS + Att_TC + Att_PROB + Att_ACC + ES × EDL + (1|Seti)
LCMLM	Choice	Att_TA + Att_DIS + Att_TC + Att_PROB + Att_ACC	PersonID, Att_TA + Att_DIS + Att_TC + Att_PROB + Att_ACC, classmb: age, sex, SHI, ES, EDL, INCn, HSn, PSC, KF, AFHD, CHI, DCHI, Kostzu_r, EDL × HSn	Wahl ~ Att_TA + Att_DIS + Att_TC + Att_PROB + Att_ACC, random = ~ Seti, subject = "PersonID", mixture = ~ Att_TA + Att_DIS + Att_TC + Att_PROB + Att_ACC, classmb = ~ sex + EDL + INCn + HSn, ng =2, data = Daten, link = "linear"
GLMM real	Real	Datentn$Att_TA + Datentn$Att_DIS + Datentn$Att_TC + Datentn$Att_PROB + Datentn$Att_ACC	PersonID Datentn$sex + Datentn$age, +PSCpocketn + SHI, EDL+ES + INCn + PSC + Kostzu_r + Khf + CHIn + HSn + DCHIn + PSC, AFHD	Real ~ Att_TC + Att_PROB + Att_ACC + EDL + ES + INCn + Kostzu_r + AFHD (1|PersonID)

Abbreviations: AFHD, afraid of hereditary disease; CHI, children; CHIn, CHI (numeric); DCHIn, desire to have children; DCHIn, DCHI (numeric); EDL, educational level; ES, employment status; GLMM, generalized linear mixed-effects model; HSn, health status (numeric); INCn, income (numeric); KL, known familar hereditary diseases; LCMLM, latent class mixed logit model; PSC, participation in screening; SHI, social or private health insurance.

Table S3 Latent class mixed logit model results – attribute effects

Attributes and levels	Class 1 (higher proportion of men)			Class 2 (higher proportion of woman)		
	β coefficient	SE	*P*-value	*β* coefficient	SE	*P*-value
Test accuracy						
90%	–0.002	0.04244	0.962	–0.234	0.03229	0.000
95%	0.079	0.03596	0.027	0.015	0.03102	0.634
99% (ref)	–0.081			–0.248		
Identified diseases						
All diseases	0.082	0.0405	0.043	0.137	0.03581	0.000
Treatable diseases	–0.078	0.03621	0.030	–0.088	0.03373	0.009
Serious hereditary disease (ref)	0.160			0.225		
Test costs						
€1,500	–0.216	0.03467	0.000	–0.151	0.03073	0.000
€1,000	–0.016	0.03283	0.620	0.108	0.03043	0.000
€500 (ref)	–0.200			–0.259		
Probability of occurrence						
10%	0.158	0.03623	0.000	–0.398	0.0341	0.000
40%	0.075	0.03431	0.029	0.007	0.03158	0.834
70% (ref)	0.083			–0.404		
Access to data						
Insurer and researcher	–0.200	0.04125	0.000	0.142	0.03933	0.000
Researcher	0.282	0.03912	0.000	0.314	0.03644	0.000
Insurer	–0.478	0.04563	0.000	–0.043	0.03765	0.258
No one else (ref)	0.760			0.357		
Intercept						
	0	NA	NA	–0.01679	0.0276	0.54311

Notes: Adjusted for class-membership effects, sex, educational level, and income; subject, "PersonID".
Abbreviations: SE, standard error; NA, not applicable.

Table S4 Results from the generalized linear mixed-effects model

Topics	Variables	Levels	Full sample			Potential users		
			β coefficient	SE	*P*-value	*β* coefficient	SE	*P*-value
Attributes	Test accuracy	90%	–0.330	0.050	0.000	–0.251	0.072	0.000
		95%	0.120	0.051	0.020	0.028	0.075	0.709
		99% (ref)	–0.450			–0.279		
	Identified diseases	All diseases	0.228	0.049	0.000	0.496	0.071	0.000
		Treatable diseases	–0.259	0.050	0.000	–0.386	0.073	0.000
		Serious hereditary disease (ref)	0.487			0.882		
	Test costs	€1,500	–0.515	0.051	0.000	–0.497	0.073	0.000
		€1,000	0.067	0.046	0.148	–0.013	0.067	0.842
		€500 (ref)	–0.582			–0.483		
	Probability of occurrence	10%	–0.411	0.051	0.000	–0.373	0.073	0.000
		40%	0.100	0.050	0.043	0.092	0.072	0.199
		70% (ref)	–0.511			–0.466		
	Access to data	Insurer and researcher	–0.011	0.062	0.860	–0.033	0.089	0.709
		Researcher	0.755	0.065	0.000	0.554	0.092	0.000
		Insurer	–0.812	0.067	0.000	–0.636	0.102	0.000
		No one else (ref)	0.046			0.049		
Person-specific data	Employment		0.000	0.131	1.000	–0.007	0.342	0.983
	Educational level		0.000	0.076	1.000	–0.006	0.194	0.975
	Employment × educational level		0.000	0.045	1.000		0.106	0.981
	Intercept		0.007	0.258	0.978	0.020	0.654	0.975

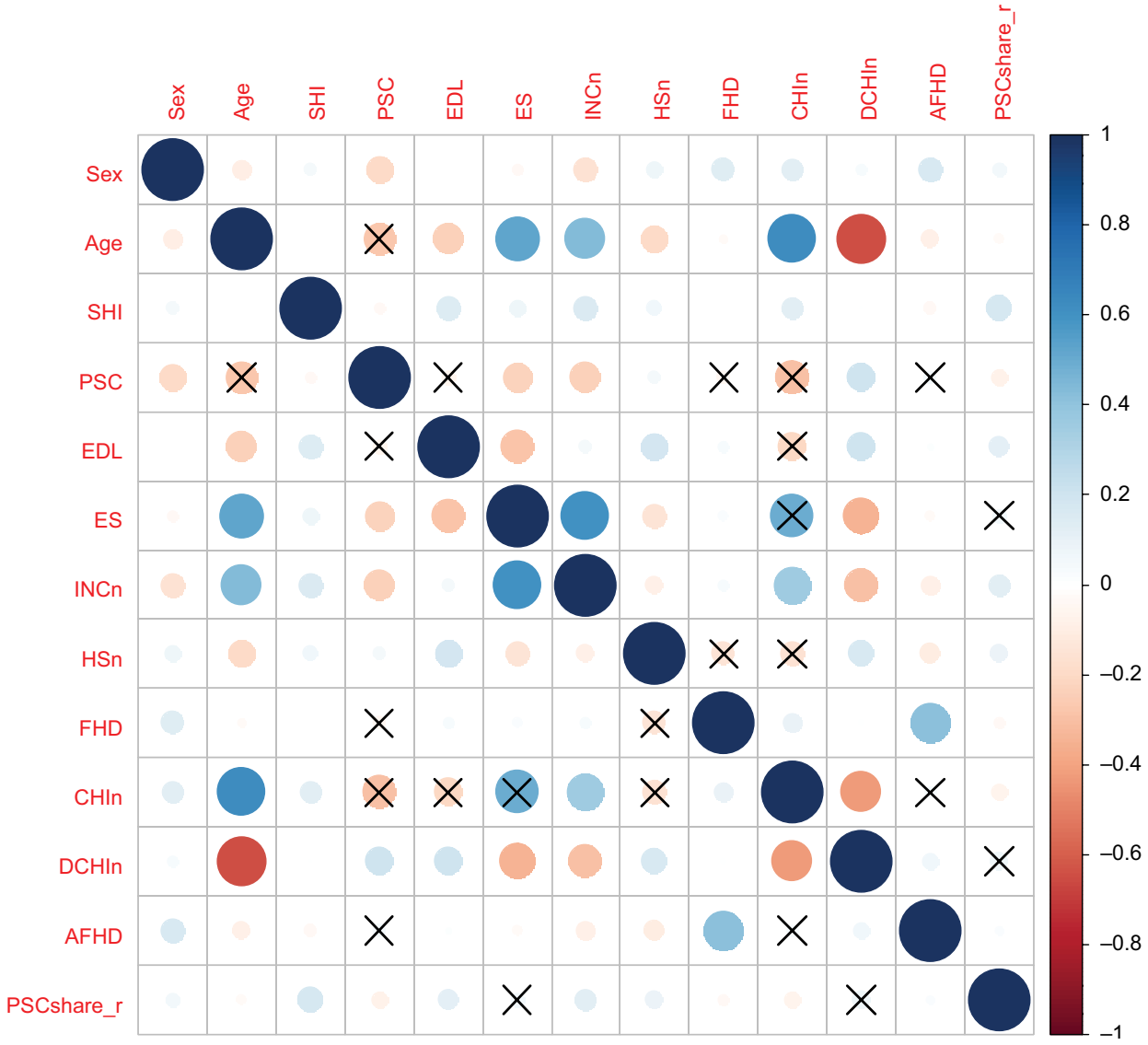


Figure S1 Correlation plot of independent variables.
Notes: The significance level was a *P*-value of 0.05. X: not significant correlations. Dark blue indicates highly positive correlations. Dark red indicates highly negative correlations. Larger circles indicate higher correlations. PSCshare_r, PSC if health insurance pays a share.
Abbreviations: AFHD, afraid of hereditary disease; CHIn, children (numeric); DCHIn, desire to have children (numeric); EDL, educational level; ES, employment status; FHD, family hereditary disease; HSn, health status (numeric); INCn, income (numeric); PSC, participation in screening; SHI, social or private health insurance.

Modul 6

Applying the Analytic Hierarchy Process in healthcare research: A systematic literature review and evaluation of reporting

Schmidt, Katharina

Aumann, Ines

Hollander, Ines

Damm, Kathrin

von der Schulenburg, J.-Matthias

BMC Med Inform Decis Mak 2015;

15: 112

Schmidt et al. BMC Medical Informatics and Decision Making (2015) 15:112
DOI 10.1186/s12911-015-0234-7

BMC Medical Informatics and Decision Making

RESEARCH ARTICLE Open Access

Applying the Analytic Hierarchy Process in healthcare research: A systematic literature review and evaluation of reporting

Katharina Schmidt[1*], Ines Aumann[1], Ines Hollander[2], Kathrin Damm[1] and J.-Matthias Graf von der Schulenburg[1]

Abstract

Background: The Analytic Hierarchy Process (AHP), developed by Saaty in the late 1970s, is one of the methods for multi-criteria decision making. The AHP disaggregates a complex decision problem into different hierarchical levels. The weight for each criterion and alternative are judged in pairwise comparisons and priorities are calculated by the Eigenvector method. The slowly increasing application of the AHP was the motivation for this study to explore the current state of its methodology in the healthcare context.

Methods: A systematic literature review was conducted by searching the Pubmed and Web of Science databases for articles with the following keywords in their titles or abstracts: "Analytic Hierarchy Process," "Analytical Hierarchy Process," "multi-criteria decision analysis," "multiple criteria decision," "stated preference," and "pairwise comparison." In addition, we developed reporting criteria to indicate whether the authors reported important aspects and evaluated the resulting studies' reporting.

Results: The systematic review resulted in 121 articles. The number of studies applying AHP has increased since 2005. Most studies were from Asia (almost 30 %), followed by the US (25.6 %). On average, the studies used 19.64 criteria throughout their hierarchical levels. Furthermore, we restricted a detailed analysis to those articles published within the last 5 years ($n = 69$). The mean of participants in these studies were 109, whereas we identified major differences in how the surveys were conducted. The evaluation of reporting showed that the mean of reported elements was about 6.75 out of 10. Thus, 12 out of 69 studies reported less than half of the criteria.

Conclusion: The AHP has been applied inconsistently in healthcare research. A minority of studies described all the relevant aspects. Thus, the statements in this review may be biased, as they are restricted to the information available in the papers. Hence, further research is required to discover who should be interviewed and how, how inconsistent answers should be dealt with, and how the outcome and stability of the results should be presented. In addition, we need new insights to determine which target group can best handle the challenges of the AHP.

Keywords: Multi-criteria decision making, Priorities, Analytic Hierarchy Process, Methodological standards, Systematic literature review

* Correspondence: ks@cherh.de
[1]Center for Health Economics Research Hannover (CHERH), Leibniz University of Hanover, Otto-Brenner-Str. 1, 30159 Hannover, Germany
Full list of author information is available at the end of the article

Background

The resources in health care systems are limited. Exacerbating this issue is the problem that many developed countries face, that is, the rising proportion of older, multimorbid patients, who serve to raise the cost of health care. Furthermore, innovations in medical care, such as equipment, pharmaceuticals, and treatment methods, are also driving up costs. German politicians have adopted new laws to manage the costs of pharmaceuticals, e.g. the Act on the Reform of the Market for Medicinal Products in 2011 (in German: AMNOG [1]). In this context, patient-relevant outcomes have drawn greater attention because the added benefit for patients determines the reimbursement price. But also, other countries are interested in reliable methods to measure benefits for patients, for example, to support Health Technology Assessments by patient preferences [2, 3]. Therefore, while it is now important to measure the benefits and to prioritize the needs of patients, it will be even more so in the future. However, several studies have found a divergence in patients' and physicians' preferences or priorities regarding prevention and therapy (e.g. [4–6]). Thus, one mean of evaluating these preferences and bringing them into accord is to take the required perspective for the situation. In order to find appropriate methods for measuring the benefits and for prioritizing them, beside the established methods, new approaches of decision making tools are transferred from other fields of research, like the marketing sector. For all of these methods it is essential to measure the trade-off between attributes in multi-criteria decision situations for each participant or the group, and as such, adequate and understandable methods are essential.

Several methods are known for multi-criteria decision making in the field of health care, including value based methods, strategy based methods, and conjoint analyses [7]. In Germany, the Institute for Quality and Efficiency in Health Care (IQWiG) suggested two methods for multi-attribute decision making: Conjoint Analysis (CA) and the Analytic Hierarchy Process (AHP) [8]. Although they concluded that both methods are applicable for decision making, they were also confronted with methodological limitations. As the advantages and disadvantages of established methods like the CA have been discussed in a number of publications (e.g. [9–11]), the AHP method has received less attention. Therefore, we wanted to figure out whether the AHP method could become a good alternative in multi-criteria decision making.

Relevance and objective of the study

The Analytic Hierarchy Process (AHP) was developed by Saaty in the late 1970s and originally was applied to the marketing sector [12, 13]. Dolan et al. were the first to apply this method to health economics research in 1989 [14, 15]; since then, it has been accepted slowly as a method in the field of multi-criteria decision making in healthcare. Liberatore and Nydick described the importance of applying the AHP as follows: "Health care and medical decision making has been an early and on-going application area for the AHP" [16]. The AHP method was applied to different contexts, for example, the development of clinical guidelines [17, 18] or biomedical innovations and technology development [19, 20].

The increasing application of the AHP has been the motivation for this study to explore the current state of its methodology. The method is the basis for assessing the best instrument for each decision situation and reflecting each participant's opinion correctly. A review provides an overview of published papers in this field. In line with De Bekker-Grob et al. [21], we provide a systematic review of the AHP. Therefore, an overview is given of the year of publication, country, and number of criteria used in the AHP (Section 3). In addition, Hummel and Ijzerman [22] analyzed the thematic field in which AHP is used. They identified the different areas of application (e.g., shared decision making, clinical guidelines, and healthcare management), number of criteria and alternatives, individual or group decisions, participants, and rating method. We focus on the methodological applications in the second step. In addition, the analyzed time horizon (2010–2015) should provide an update on Hummel and Ijzerman's study and allow us to provide details of the most recent developments in the subject area. As in Mühlbacher's overview [23], the field of application and the sample are inspected, although our focus remains on the current state of the research (the last 5 years) and the reporting of methodological aspects in the papers. In addition, the evaluation of studies' reporting allows deeper insights. Therefore, we develop criteria for reporting the AHP method and determine to what extent the studies fulfill the criteria. We conclude by proposing recommended situations in which the AHP can be used.

AHP – a short introduction

As a short introduction into the method of AHP, we report the most important aspects here. We refer to detailed papers to provide deeper insights into specific methodological aspects.

The AHP disaggregates a complex decision problem into different hierarchical levels (see Saaty's axioms for the AHP [24]). The application of an AHP is structured into six steps (see also Fig. 1), suggested by Dolan et al. [25] and Dolan [7], as follows: 1. define the decision goal, criteria, and alternatives, 2. rate the criteria in pairwise comparisons, 3. calculate the relative priority weights for the (sub-)criteria, 4. calculate the criteria's global priority weights and combine the alternatives' priorities, 5. control for inconsistency, and 6. perform sensitivity analysis.

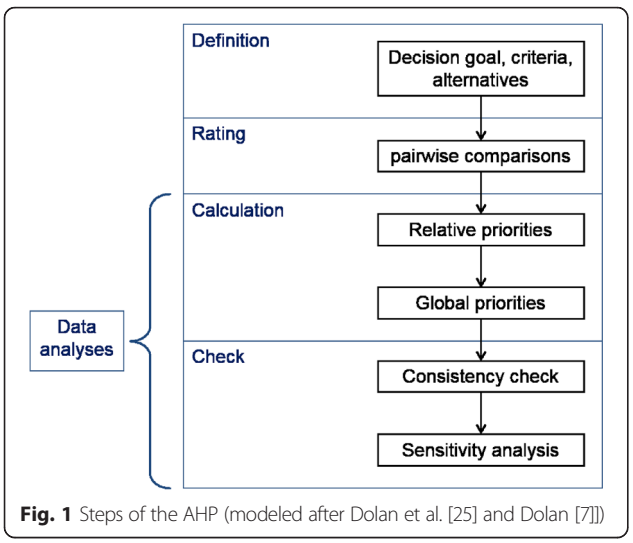


Fig. 1 Steps of the AHP (modeled after Dolan et al. [25] and Dolan [7])

At the first hierarchical level, the aim of the study is defined followed by the main criteria, which can be divided further at lower levels into sub-criteria. If necessary, alternatives that contain specific combinations of characteristics can be arranged at the lowest level of the hierarchy. Although the AHP was introduced for group decisions, it may also be applied to single person decisions [26]. Pairwise comparisons at each hierarchical level present the judgments and they must be evaluated according to a scale developed by Saaty, which ranges from 9 to 1 to 9. If the alternatives consisted of subjective combinations of the criteria, the alternatives would be judged also with regard to each criterion. Saaty provided a detailed description of his scale and its intensities [12].

In order to analyze the interviews, the pairwise comparisons of (sub-)criteria at each level are displayed in ordered schemes (matrixes). An example is seen in Saaty ([24], p. 164). Only half of the matrix has to be filled in, as the other half is obtained from the reciprocal weights. The Eigenvector method (EV) is the most common means of calculating the priority vector, although other methods, such as additive normalization, weighted least-squares, logarithmic least-squares, logarithmic goal programming, and fuzzy preference programming methods, yield comparable results [27]. The EV relies on the matrix's principle eigenvalue, which results from a process of repeated squaring and normalization (for more information, see Srdjevic [27] or Saaty [12]). The resulting local weights describe the relative priorities in relation to their parent criterion. The local weights form the global weights for the criteria through multiplication with the local weights from their parent criteria [24]. Thereby, global weights for criteria show the importance of each criterion in the overall context of the hierarchy. The priorities for the alternatives of the AHP are calculated by the sum of the particular local and global weights for each alternative [23]. For detailed information and examples concerning the calculations, see Saaty [28].

The aggregation of the individual judgments or priorities is fundamental to the outcome of the study. The first option is to have the group of participants vote by finding consensus. Another alternative is to aggregate the individual judgments. Still further, the literature suggests finding the geometric mean [29] or arithmetic mean [30]. In addition, the timing of calculating the average affects the results [30], specifically, the average of participants' judgments or the average of participants' global weights. Yet another option is to give special weight to one participant's decision on the basis of that participant being an expert in the field or holding an exceptional position within the group [30]. The consistency ratio (CR) measures the uniformity of a respondent's answers to the AHP questions. Saaty [24] describes the calculation of the CR in detail. The CR can also be calculated for a group of respondents.

Although the AHP has been applied to a variety of topics within the healthcare field, the sensitivity analyses on hierarchical decision making has received little investigation [31]. It should be noted that there are two distinct types of sensitivity analysis, that of judgments and that of priorities [32]. The former has been explained and tested by Arbel [33], Moreno-Jimenez and Vargas [34], and Sugihara and Tanaka [35]. They determined the judgments' upper and lower bounds and articulated the preferences through preference structures. Other approaches originate from Moreno-Jimenez and Vargas [34], Triantaphyllou and Sánchez [36], Sowlati et al. [37], Masuda [38], and Huang [39]. Erkut and Tarimcilar [40] provided "a collection of practical tools for a potentially powerful sensitivity analysis in the AHP". In addition, Altuzarra et al. [41] proposed a method for determining the stability of group decisions. If the AHP includes alternatives, the sensitivity analysis could show the effect of varying weights on the alternatives' rankings [23]. Therefore, potential rank reversal of alternatives can be simulated. Rank reversal occurs when adding or deleting an (irrelevant) alternative leads to a shift in the previous alternatives' ranking order [42].

Methods

This chapter is divided into two parts to introduce the methods used in this paper. The first part describes the method of the systematic review, which includes the key words and a flow chart. Further, in chapter 2.2, we describe our evaluation of reporting quality for the included studies.

Systematic literature review

The basis of this review is a systematic literature research on the Pubmed and Web of Science databases (date of research: 10/27/2015). As we focused our research question

on healthcare, we did not include further databases in the other scientific fields. We searched both databases for articles with the following keywords in their titles or abstracts: "Analytic Hierarchy Process," "Analytical Hierarchy Process," "multi-criteria decision analysis," "multiple criteria decision," "stated preference," and "pairwise comparison." We provided the search strategy in Appendix: Table 1. It was technically not possible to search Web of Science for keywords in the abstracts. We refined the search by including only articles written in German or English and those associated with healthcare. Two independent reviewers evaluated the titles and abstracts of the resulting studies. Therefore, the criterion for inclusion was that the article is the primary source and the study used the AHP method within the healthcare setting. Additionally, we conducted a manual search to find further articles not included in the aforementioned databases. Thereafter, the two reviewers screened the full texts of the remaining articles and discussed whether to include them in the review. After reaching consensus, the important information was summarized in a table (not shown). Apart from common information, like the author, title, publication year, country, and journal, we extracted additional information regarding the study's aim, source of criteria identification, hierarchy design, form of implementation, and analytical steps in order to conduct our analysis. The results are described in Section 3 for the entire period and in detail for the last 5 years in Subsection 3.1. The first step should give a short overview of all studies that were conducted with AHP in health care. In the second step, we reported the current state of research in more detail.

Evaluation of reporting quality

The papers identified from the last 5 years resulting from the systematic review were evaluated with regard to their reporting quality. Because there was no set standard by which to judge the AHP's methodological issues, the evaluation of the studies' quality was quite challenging. The before mentioned studies by De Bekker-Grob et al. [21], Hummel and Ijzerman [22], and Mühlbacher et al. [23] did not report quality criteria. However, the Consolidated Standards of Reporting Trials (CONSORT) Statement for randomized controlled trials [43] and the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) Statement [44] may provide some direction by providing checklists for transparent and complete reporting. The reason why authors should report specific aspects is the traceability of the study. Some criteria from the CONSORT Statement could be transferred to AHP studies: sample size, participants (eligibility criteria), trial designs, and statistical methods. In the case of the AHP method, the latter criterion consists of the CR, the method used to calculate the weights, the statistical software, and sensitivity analyses. Another checklist item is the description of the intervention. Transferred to the AHP method, authors should provide information about the interview process. Besides, another guideline for good research practices is published by Bridges et al. [9]. They provide a detailed checklist that is specific for conducting conjoint analyses. Since it suggests quality aspects only for those kinds of studies, the checklist cannot be used directly for our evaluation. However, we summarized the recommendations from the different fields and we obtained a simplified measurement of reporting by counting the elements that were included in the studies. Therefore, we evaluated whether the authors mentioned aspects for the following elements in their papers:

- Decision goal, criteria (and if necessary alternatives)
- Number of participants
- Type of participants (patients, potential consumers, or experts)
- Decision making by group or individually
- Scale for pairwise comparisons
- Interview process (face to face, email, questionnaire, judgments based on literature)
- Software
- CR
- Calculation of weights
- Sensitivity analysis

The last criterion was valid only for studies including alternatives. Thus, for the other papers without alternatives, we could determine only whether descriptive statistics (e.g., standard deviation, SD and confidence intervals, CI) were reported for the judgments or weights. We calculated the sum of all reported aspects for each study and present the results in Appendix: Table 2 and we show charts in Subsection 3.2. Nevertheless, we could not evaluate the content of each of the abovementioned criteria but only whether the criteria were reported in the study.

Results

The search in Pubmed yielded to 1,956 articles and the search in Web of Science yielded to 4,829 articles, as Fig. 2 shows. Furthermore, 44 additional records were found via manual search. By screening titles and abstracts, we limited the sample to 246 articles (we excluded a total of 6,485 articles based on language or irrelevance to healthcare and we found 54 duplicates). Thereafter, we examined the full articles in order to determine whether they apply AHP to the field of healthcare. An additional 125 papers were excluded because they were not original studies or they used other stated preference methods (e.g., discrete choice experiment). In total, this process yielded to 121 relevant studies; the Appendix: Table 3 provides a complete list. We provide a brief overview of these studies to show how many studies have been published in this

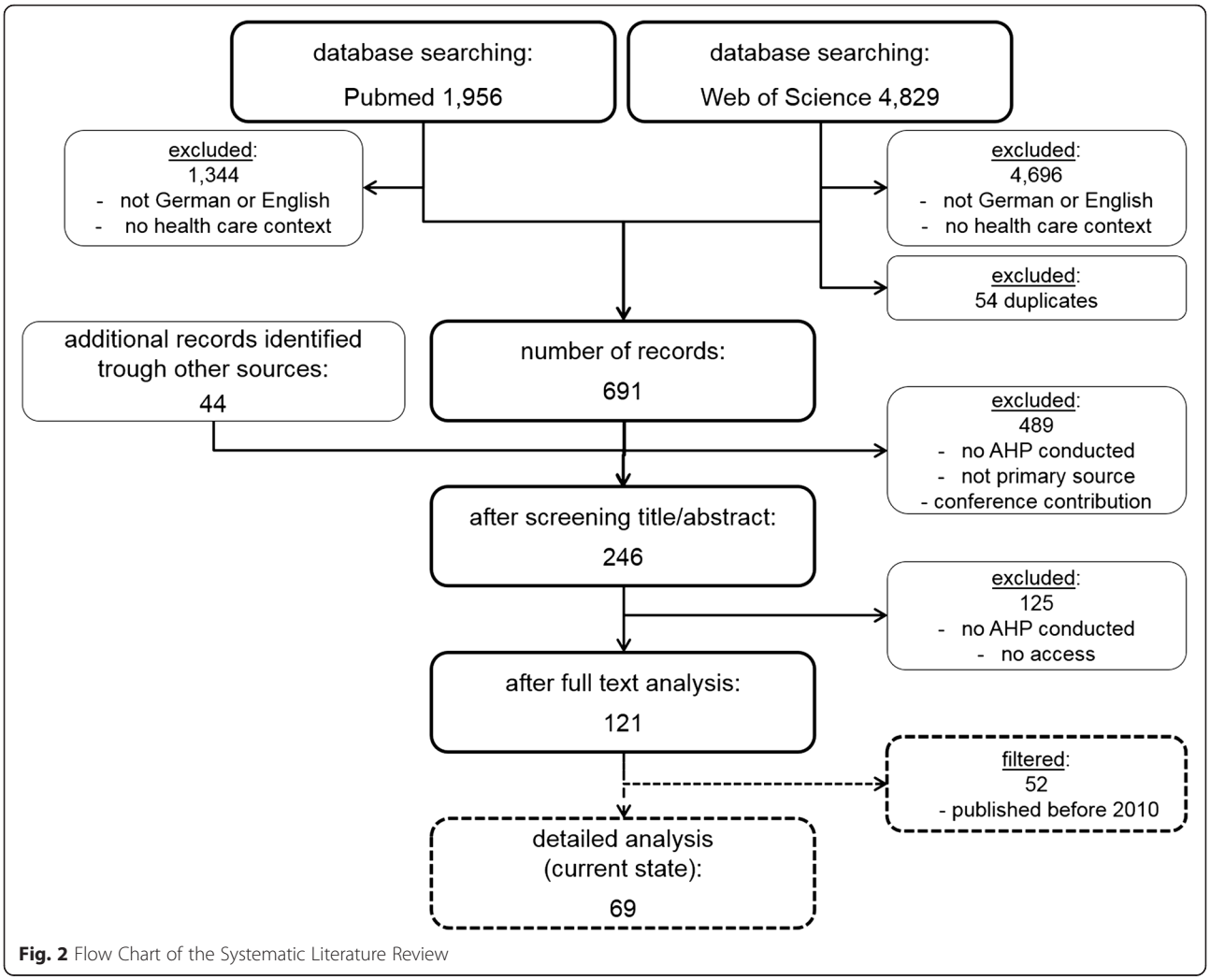


Fig. 2 Flow Chart of the Systematic Literature Review

field and in which context the authors used the AHP. In addition, the overview presents the development and the relevance to the AHP method. In order to explore the current state of the literature, we limited the body of our research to articles published within the last 5 years. This restriction reduced the number of studies to 69. The detailed analysis of these studies' methodologies made it necessary to reduce the number of articles.

For a first overview, we briefly summarized the key factors of all of the relevant articles ($n = 121$), such as their publication year, country, number of attributes, and levels.

The earliest study to use the AHP was published in 1981, but the AHP has become increasingly popular since 2005 (see also Fig. 3). The 2 years with the greatest number of studies published on the subject were 2011 and 2012 with nine each. However, it should be noted that our evaluation period contains only the first 10 months of 2015, in which as many as 20 studies were published. On average, there were 2.5 studies per year between 1981 and 2013. During the 1990s, there was an average of 1.7 publications on the AHP per year, which increased to 4.6 per year between 2000 and 2013. In 2014 and 2015 the average increased to the peak of 18.5 studies, although the last two months of 2015 are not included.

Most studies were from Asia (29.75 %), followed by the US (25.62 %). Almost all studies published before 2000 were conducted in the US ($n = 15$). However, between 2000 and 2010, a larger proportion came from Asia ($n = 8$) and Europe ($n = 7$), although most were still from the US ($n = 8$). Since 2010, Asia ($n = 26$) and Europe ($n = 17$) have surpassed the number of publications in the US ($n = 8$).

Another important aspect of these studies is the number of hierarchical levels that they include. Therefore, the studies could include more than one hierarchy, so in some cases the number of studies did not sum up to 121. More than half of the studies (51 %) included three hierarchical levels, 23 % described their hierarchy with two levels, and 21 % used four levels. On average, the studies used 19.76 criteria throughout their hierarchal levels. At the second hierarchical level, 96 articles (78 %)

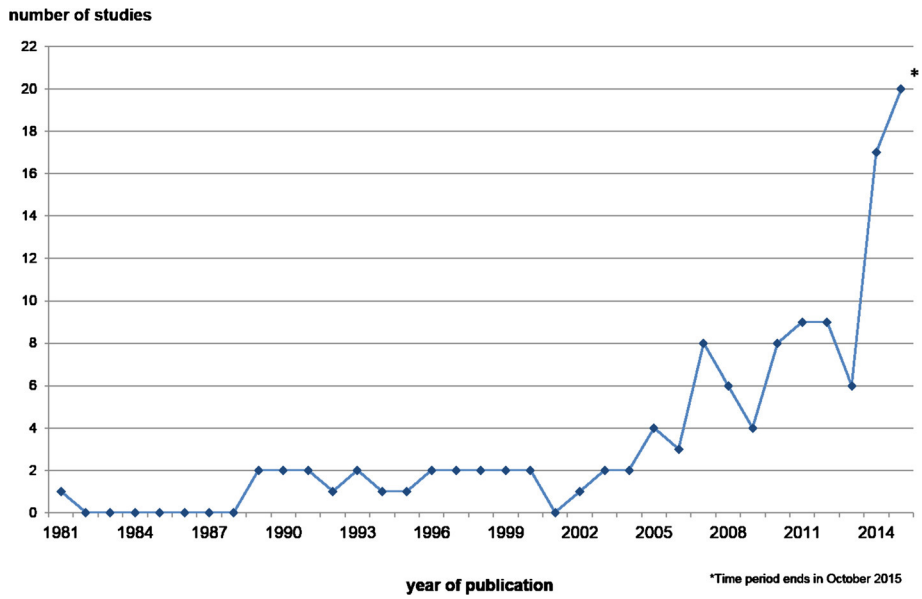


Fig. 3 Included Studies by Year of Publication

included between 1 and 5 criteria (Fig. 4). At the third and fourth levels, most studies ($n = 39$ and $n = 16$ or 45 and 47 %, respectively) used between 11 and 20 criteria. The number of studies with five hierarchical levels was quite small ($n = 3$). As expected, the number of criteria increases as the hierarchical level increases. The right bar in Fig. 4 shows the total number of criteria for all hierarchical levels per study.

Following the method set forth by Hummel and Ijzerman [22], we divided the studies into five categories: development of clinical guidelines, healthcare management, government policy, shared decision making, and biomedical

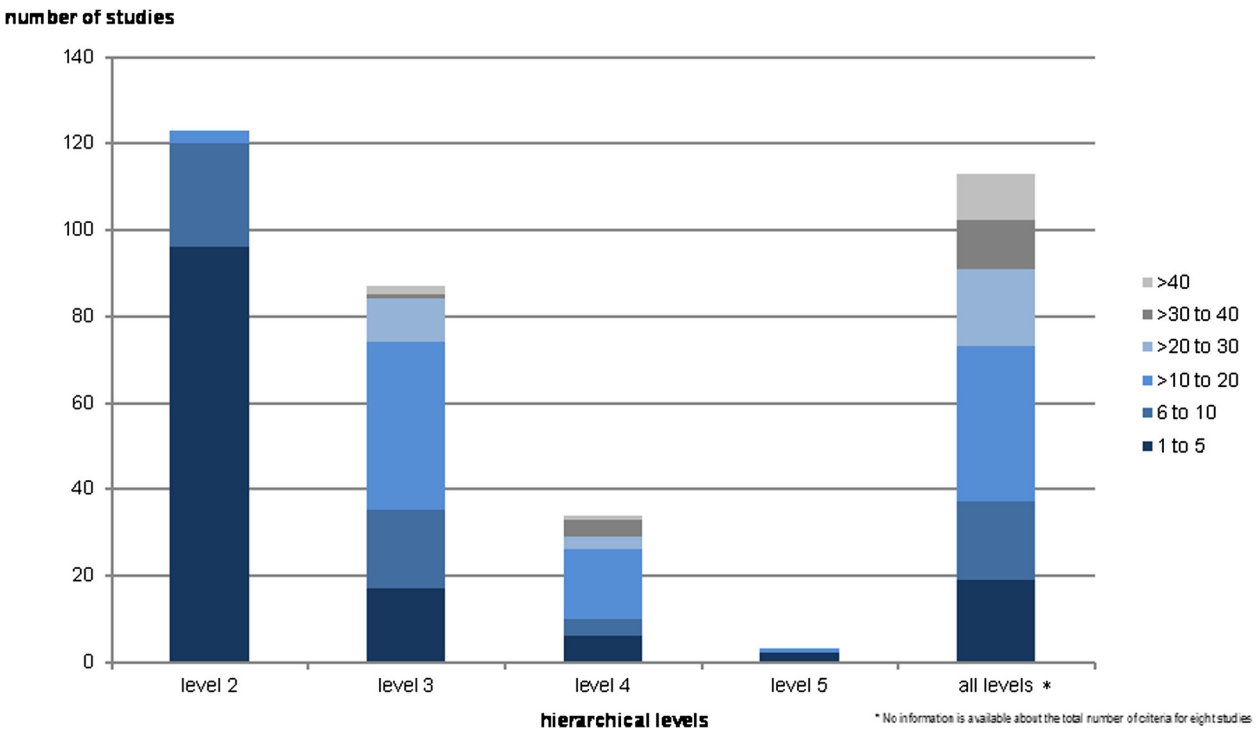


Fig. 4 Number of Criteria per Hierarchical Level

innovation. We classified 38 studies (31 %) as pertaining to the development of clinical guidelines or recommendations, 30 (25 %) to healthcare management, 26 (21 %) to government policy, 15 (12 %) to biomedical innovation, and12 (10 %) to shared decision making.

Detailed analysis of the current state of research

This subsection summarizes the results of our analyses of the articles published within the last 5 years (January 2010 to October 2015). We examine how the studies design their hierarchies and carry out their surveys, and which analytical steps they take. In doing so, we follow the steps for conducting an AHP shown in Fig. 1.

Definition of decision goal, criteria, and alternatives

The first step in conducting an AHP is to define the decision goal and criteria that describe the goal at a lower hierarchical level. In order to do this, many studies relied on literature research [20, 25, 26, 45–83]. In addition, many studies relied on expert interviews [20, 45–49, 51, 54, 56–58, 61, 66–71, 74, 75, 77, 78, 81–97] or focus groups [26, 51, 69, 82, 87, 98]. Almost all of the studies defined their criteria by analyzing more than one source of information, although five publications did not explain their process for this step [99–103]. Some authors defined the criteria according to standards or established guidelines [25, 50, 52, 59, 80, 84, 92, 93, 104–108] or even from previous study results [25, 47, 62, 68, 69, 71, 72, 81]. Still other authors relied on their own expertise [64, 73, 107, 109, 110].

Judgment through pairwise comparisons

The sample sizes varied between one author who judged the AHP for himself [73, 107–109] to 1,283 participants [55]. In total, 50 of the 69 articles reported the number of participants in their AHP studies. The mean number of participants in these studies was about 109. Depending on the studies' goal, the participants belonged to the following groups: hospital employees [49, 92], patients [25, 47, 55, 59, 60, 64, 69, 72, 75, 82, 95, 98], public/consumers [52, 70, 103], doctors or specialists [26, 71, 72, 74, 79, 81, 83, 93, 94, 96, 97, 99, 110], medical students [80] or teachers [77], biomedical engineers [94], technical experts [93], managers [93], administrators [20], and stakeholders [75]. Of the studies, 44 interviewed experts [20, 26, 45, 46, 48–51, 54, 56–58, 61, 62, 66–68, 71, 74, 76–79, 81, 83–94, 96, 97, 99, 104–107, 110], 11 studies surveyed consumers or patients [25, 47, 52, 55, 59, 60, 69, 70, 82, 98, 103], and four studies included both [64, 72, 75, 95]. However, six authors did not mention who answered the AHP questions [53, 63, 65, 100–102].

Next, we considered whether the AHP was applied at individual or group level. Most of the studies questioned their participants individually [20, 25, 26, 47, 55, 56, 59, 61, 62, 64, 66, 69–71, 74, 75, 77, 79–83, 87–90, 94, 97–99, 103, 104, 109–111]. On the other hand, only six articles mentioned group decisions [46, 49, 72, 84, 92, 96]. Five studies conducted individual judgments as well as group decisions [51, 60, 86, 93, 95]. The remaining 23 articles did not describe the judgment, or they had only one person who answered.

In addition, there were differences in the applied scales for the pairwise comparisons. As explained in Subsection 1.1, the original scale implemented by Saaty ranges from nine (or 1/9) to one to nine. This scale was adopted by 37 of the articles in our sample [25, 45, 46, 50–52, 54–57, 60–62, 66, 71–73, 75, 79, 80, 83, 84, 86–89, 91, 92, 94, 95, 97, 98, 102, 103, 107–109, 111]. Other studies used ranges between 1 and 4 [20, 59], 1 and 5 [67, 70, 106], 5 and 1 and 5 [26, 81, 90, 110], 6 and 1 and 6 [99],1 and 7 [47],1 and 9 [58, 77, 96], and 1 and 11 [74]. The remainder of the studies did not provide information about their scale [48, 49, 53, 63–65, 68, 69, 76, 78, 82, 85, 93, 104].

Furthermore, there were major differences in how the surveys were conducted. Once again, not all of the authors discussed their process in detail, but those that did so used online questionnaires [20, 47, 51, 55, 58, 70, 74, 75, 81–83, 111] (emailed) questionnaires [26, 59, 64, 66, 71, 77, 79, 80, 86, 91, 94, 95, 104, 110], face-to-face interviews [25, 45, 87, 90, 98], group discussions or workshops [49, 60, 64, 72, 84, 86, 92, 93, 96], or Delphi panel method [61].

Analysis and validation of results

Specific software can support the AHP design and further analyses. However, only 35 of the 69 studies (49.28 %) mentioned which software they used. The majority of the studies that reported software chose Expert Choice® (23.19 %), while others used such packages as Microsoft Excel [25, 77, 88, 90], or IBM SPSS Statistics [45, 53, 80, 99, 104]. In the last 5 years, a more diverse range of software packages has been in use; in addition to the aforementioned packages, researchers have chosen Super Decisions TM or Crystal Xcelsius [73, 107], or programmed their own software [20].

The detailed analysis showed that 22 out of the 69 studies did not state a CR. However, 31 studies used a CR of 0.1 [20, 26, 45, 46, 49–51, 56, 57, 60–62, 67, 71–74, 76, 77, 83, 87, 89, 91, 98–102, 107–109], five studies widened the range to a CR of 0.15 [25, 59, 64, 75, 111], and three studies accepted a CR of 0.2 or less [70, 81, 97]. The remaining studies did not establish a threshold prior to measuring average CRs [55, 80]. As a consequence of these consistency conditions, 14 of the studies reported the number of participants that must be excluded in order to meet their established threshold [47, 55, 59, 61, 63, 70–72, 75, 78, 81, 98, 99, 104]. However, only a small proportion of the studies actually

outlined a procedure for dealing with excessive inconsistency (i.e., a CR above the established threshold). Chen et al. [70] and Pecchia et al. [26] asked the participants to fill out their questionnaires again. Hummel et al. [94], Suner et al. [83], Velmurugan et al. [102], and Cancela et al. [51] asked the participants to check and revise their decisions. Chung et al. [71], Li et al. [77], and Pecchia et al. [81] excluded the inconsistent participants from their analyses. Hou et al. [67] wrote that, in this case, "the judgment matrix has to be modified and recalculated." Page et al. [80] ran simulations in which they assumed that the inconsistent answers were, in fact, consistent in the first place.

Furthermore, we examined group decision making. Danner et al. [72], Lin et al. [91], Papadopoulos et al. [56], Reddy et al. [86], Shojaei et al. [87], Jaberidoost et al. [66], and Hsu et al. [90] explored this topic by taking the geometric mean of the individual weights. Hilgerink et al. [93] and Hummel et al. [94] summarized the individual judgments with geometric means, and then, calculated the group weights. Conversely, other studies only averaged the group judgments [75, 95]. Olivieri et al. [79] presented two AHPs; in the first, they calculated geometric means for the ranks and in the second, they calculated the inter-participant, standardized, geometric means of the weights as well as the inter-participant means. Perseghin et al. [96], Uzoka et al. [97], and Kuruoglu et al. [98] aggregated the participants' judgments according to the median, and then, calculated the weights. By contrast, Taghipour et al. [49] constructed the group judgments by using weighted means. Unfortunately, 40 of the studies did not describe their weight calculations in detail [20, 45–48, 50–55, 57, 58, 61–65, 67–70, 73, 74, 77–79, 82, 85, 88, 89, 96, 99–101, 103, 104, 106, 107, 110]. However, 39 authors mentioned that they used the EV [25, 26, 45–47, 49, 50, 55–57, 59, 60, 62, 65, 66, 71, 72, 75, 76, 80, 81, 83, 86–95, 97, 100, 102, 104, 105, 108, 109].

Very few of the studies ($n = 14$) examined the robustness of the weights [46, 53, 56, 73, 76, 78, 80, 82, 86, 93, 100, 101, 105, 107]. Diaz-Ledezma et al. [107] and Diaz-Ledezma and Parvizi [73] referred to Erkut and Tarimcilar [40], who introduced sensitivity analysis for the AHP. Hilgerink et al. [93] factored in uncertainty regarding the included criteria by asking participants to rate the sensitivity and specificity of the pairwise judgments on a three-point scale; this yielded negative, average, and positive scenarios for the overall priorities. The other studies did not mention efforts to account for uncertainty. Further studies conducted their sensitivity analyses with the graphics provided in Expert Choice ® [100, 101].

This subsection presents the most relevant aspects of conducting AHP, and thereby, reveals a high proportion of missing information from the literature. However, we summarize these facts in Subsection 3.2 and evaluate the number of reported aspects.

Evaluation of reporting

In a final step, we evaluated the reporting of the studies (see Subsection 2.2). Therefore, we suggested ten criteria that the authors should address in their articles. Most of the aspects are described in Subsection 3.1, and so, we focus on the number of reported elements for evaluating the studies in this section. We evaluated the studies published between 2010 and 2015 (until the 27th of October) and the detailed table can be found in Appendix: Table 1. In addition, we summarized the most important aspects from the table in the following graphs.

Figure 5 shows that all of the studies ($n = 69$) reported their decision goal and their criteria in their publications. However, several studies did not describe their interview process and did not mention which software they used. Particularly, only 15 out of 69 studies reported that they conducted sensitivity analysis.

The minimum number of reported criteria is one, namely, the study of Hsu et al. [63]. They described the aim of the study (assessment of oral phosphodiesterase type 5 inhibitors for treatment decisions of erectile dysfunction) and the hierarchy for the AHP but said nothing about the methods or study process. The studies that reported the highest number of ten criteria were published by Page [80] and Maruthur et al. [111]. The mean of the reported elements is 6.75, whereas only 12 out of 69 studies (17.39 %) reported less than half of the criteria.

The next figure demonstrates the results from our evaluation of reporting quality (Fig. 6). This figure shows the results from our evaluation regarding the reporting quality of all publications between 2010 and 2015. The highest number of studies reached seven or eight points in the evaluation. Only a small number of studies ($n = 2$) reported one or two aspects required. However, two publications also reported all of the criteria. The mean of reported criteria is 6.75.

Furthermore, we divided the publications into two time periods because we wanted to examine whether the reporting quality has changed (not shown graphically). Therefore, we took the studies published between 2010 and 2013 and compared them with the recent state of research since 2014 (the peak of published studies seen in Fig. 3). In the last 2 years, five studies got nine points in comparison to only three studies in the early time period. Indeed, two publications from the last 2 years only reached one or two points compared to no publications between 2010 and 2013. As the mean of the reported criteria is 6.88 for the early period and 6.65 for the last 2 years. Apparently we do not see the expected increase of reporting quality.

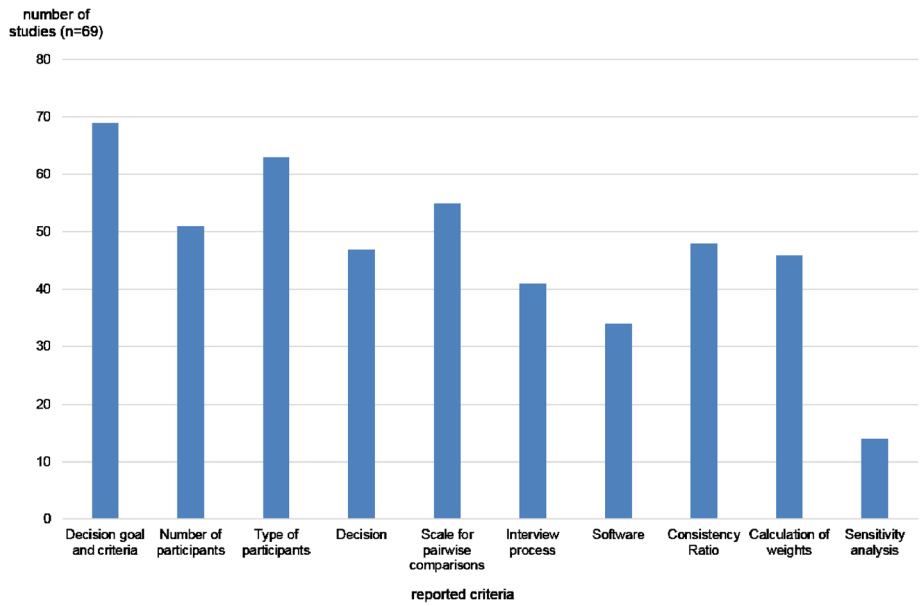


Fig. 5 Number of Studies by the Reported Criteria

Discussion

As seen from the review, in the last 10 years (and particularly in the last 2 years), there has been a clear upward trend in the number of publications that apply the AHP to healthcare. One reason for this could be the increasing acceptance and the discussion about integration of this method into policy decision processes. For example, the IQWiG in Germany suggests the AHP in decision making regarding reimbursement as one appropriate method [8]. Currently, the development of clinical guidelines is the most popular subject for AHP studies, followed by healthcare management decisions.

In the first step, the authors have to decompose their research question and set up a hierarchy for the AHP. Therefore, we have seen that most of the authors rely on literature research and expert opinions. This proceeding could carry the risk to not including further important criteria that have not been covered before but that are important for the overall problem and for the complete hierarchy. In particular, the perspective of the participants

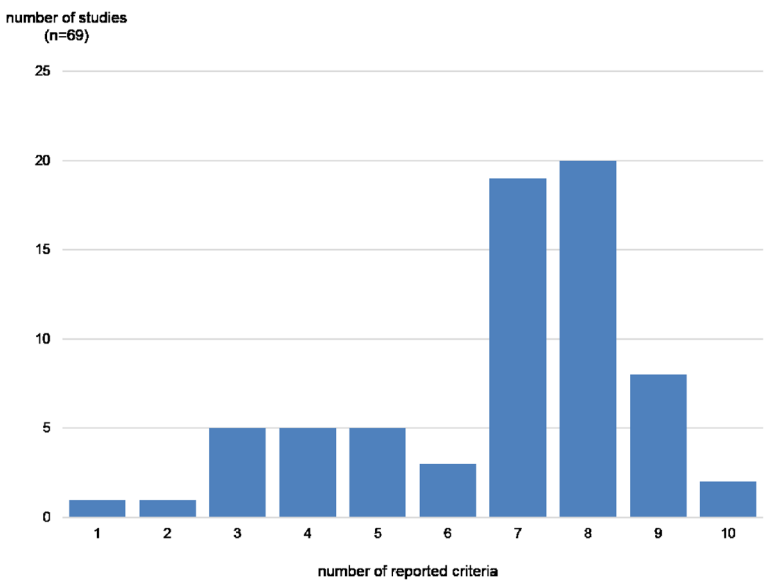


Fig. 6 Evaluation Results for Reporting Quality

(in contrast to previous research) could require new criteria for the AHP.

The review showed wide fields for choosing participants in the AHP studies, even though a large portion of papers described their samples as experts or potential consumers of goods or services in question. Sample size was an important factor in these studies, for while there is no precise rule, there is general consensus that the AHP does not require a particularly large sample [23]. Consequently, it should be noted that the results are not necessarily representative. The number of participants ranged from 1 (a single author who judged the AHP for himself) to almost 1,300 with the mean being about 109. This wide range could influence the studies' results. The evaluation of reporting in Subsection 3.2 examined satisfactory reporting of the participants in most of the papers. However, common rules for the process should be developed and several of its aspects improved upon. For instance, future research should develop a standardized method for calculating the sample size. Furthermore, the identification of the correct study sample is imperative in order to answer the studies' research question properly.

In some cases, the participants were invited to revise their answers in case of inconsistency, and thereby, participants could be unsettled and biased. However, inconsistent judging could also be an indicator of overstraining the participants. Furthermore, most of these studies carried out the AHP on an individual basis, whereas only four authors mentioned group decisions. This was an unexpected finding because the AHP was introduced initially to study group decisions. However, our evaluation of the studies' reporting showed that only six authors did not mention whether they had conducted group or individual decisions. Moreover, the aggregation of the AHP results from the individual level to a group did not present a uniform set of results. The advantage of group consensus is that it allows for the discussion of pairwise comparisons, which, in turn, improves participants' understanding of the problem and criteria, and thereby, participants answer less inconsistently. This is because, on the one hand, they discuss their decisions before they set their judgments, but on the other hand, it may be because of the consensus or average extreme judgments being compensated by the group. Thus, the quality of the decision, seen as consistency, is improved [112]. Otherwise, the composition of the group would be a highly influential factor in the process of reaching consensus. This is because individuals within the group could have opposite priorities or else could be unwilling to discuss their positions. In this case, it would not be possible to reach a unanimous vote. Thus, another alternative is to aggregate the individual judgments [113]. In order to do this, one may take the geometric mean or median of either the individual judgments or the individual weights. One prerequisite is that the reciprocal of the aggregated values must correspond to the individual reciprocal values [28]; this can be achieved only by taking the geometric mean [113]. Unfortunately, only 29 of the 69 studies describe their exact processes for calculating the weights, but 39 reported using the EV in some way.

Recently, researchers have paid some attention to whether the results of these studies are robust. Despite the fact that sensitivity analyses could offer more information on the problem of rank reversal as well as the interpretation of the outcome [23], only 14 out of the 69 studies that we examine reported conducting such tests [73, 76, 78, 82, 93, 107]. However, sensitivity analysis for AHP is relevant only when alternatives are included in the hierarchy. Consequently, 25 of 37 studies from our analysis missed reporting sensitivity analyses, as shown in Appendix: Table 2. One study without alternatives in the hierarchy suggested the use of standard deviations for weights [80]. The other sensitivity analysis presented in Subsection 1.1 requires a firm understanding of matrix algebra, does not yield fast or easy solutions, and is not supported by any software package. Although Expert Choice® provides the opportunity for sensitivity analysis, it offers only graphical simulation of one weight at the first hierarchical level [31]. Despite these challenges, sensitivity analyses remain vitally important as they allow researchers to assess the robustness of judgments, identify critical criteria or alternatives, find consensus through a range of judgments, and investigate different scenarios that support the decision [31]. Recently, Broekhuizen et al. have taken a further step concerning sensitivity analysis by providing an overview of dealing with uncertainty in multi-criteria decision making [114]. The results from sensitivity analysis can indicate potential rank reversal. The long-running dispute of rank reversal in AHP raised the question of "[…] the validity of AHP and the legitimacy of rank reversal" [42]. Wang et al. [42] argued that rank reversal is not only a phenomenon in the AHP but also in other decision making approaches. Saaty stated that the relative measurement of alternatives in the AHP implied by definition that all included alternatives were relevant, in contrast to utility theory that could face rank reversal problems [115]. Apart from these fundamental questions, several authors have suggested modifications to the AHP to overcome the problem of rank reversal [116].

Our evaluation of the reported criteria emphasizes the need to increase the number of given information in AHP studies. In general, authors should improve reporting on methodology, which is essential for comprehending and reproducing other authors' results. This would serve to facilitate other researchers' evaluations of study quality. In our opinion, two central explanations are possible for the current underreporting in the literature. First, the AHP,

being fairly new, has few precisely formulated methodological rules. Second, what rules there are do not hold in practice. The latter observation also encompasses cases in which the AHP was too difficult for participants, either because of the formulations of the criteria or because of the method itself. It can be concluded that further research, in particular, methodological research, is needed in this field.

Although this study is based on systematic literature research and transparent evaluation criteria, there are a number of limitations that bear mentioning. As we primarily conducted our research on the Pubmed and Web of Science databases, it is possible that we did not include all relevant articles from other databases, even though we conducted a manual research. In addition, not all studies reported their procedures and methodologies in detail; therefore, the resulting statements in this review and the evaluation of the studies' reporting could be biased, as we were restricted to available information. We are unable to make statements about the appropriateness of the evaluated content, like the sample size. By contrast, our evaluation criteria considered only whether a point was mentioned. Furthermore, the evaluation of reporting relied on the CONSORT and PRISMA Statements in order to develop criteria for the AHP. These statements suggest evaluation criteria for RCTs and systematic literature reviews, thus it could be criticized that we apply them to the subjective method of the AHP. The importance of each criterion can be criticized and our overall evaluation provides only an indication of the studies' reporting with respect to informational content—not the quality. Moreover, we summarized the articles' procedures but were unable to convey their results without some adaptions and generalizations; some aspects of the AHP must be adapted to suit the situation.

Conclusion

We found that there is a pressing need to develop methodological standards for the AHP; otherwise, discrepancies in methodology could bias studies' results. In particular, future research should establish a standard procedure for aggregating individual data, specifically, a standard for using the geometric mean versus the arithmetic mean and aggregating judgments or priorities. We should place special emphasis on finding practical sensitivity analysis to address the criticisms regarding rank reversal due to changed judgments. In addition, suggestions are necessary for reporting the robustness of weights for AHPs that do not include alternatives.

Besides the methodological aspects of the AHP, we should also think about the topic that is researched. We carved out that the AHP is based on the hierarchical structure and the criteria that are included. If the author uses improper assumptions, he will find biased results. Therefore, the AHP hierarchy should not only base on one source of information but also on a combination of different methods (e.g. literature research and expert interview). Hence, further research is required about how to determine the interviewees, what should be done with inconsistent answers, and how the outcomes and the stability of the results should be presented. In the future, we need new insights as to which target groups can best handle the challenges of the AHP. These challenges are mainly consistent answering, preventing overstraining by using adequate numbers of pairwise comparisons, and deciding between group and individual AHP. Therefore, researchers should investigate specific groups, like elderly people, healthy people, and patients with different diseases or disabilities.

In our study, we analyzed whether authors reported important aspects of the AHP in their studies. This could be a first step to evaluate the quality of studies applying AHP in healthcare. In addition, guidelines should be formulated as to which statistics should be reported and how to conduct high-quality AHPs. As mentioned before, Bridges et al. published a checklist that contains recommendations for conducting conjoint analyses on healthcare topics on behalf of the International Society for Pharmacoeconomics and Outcomes Research (ISPOR) group [9]. Besides aspects for study presentation, it suggests criteria for evaluating the choice of attributes and the appropriateness of the method for the research question. Still further, we should take the current criticisms of the AHP into consideration so that we can find solutions to address them.

This systematic literature review shows a heterogeneous picture for application of the AHP in health economics research. It is likely that interest in the AHP will rise in the future, particularly in its application to health economic evaluations, the weighing of therapy outcomes, and benefit assessments. In this context, the AHP method could support decision making regarding reimbursement of pharmaceuticals. This is largely owing to its ability to translate complex questions into stepwise comparisons at different hierarchical levels. In these hierarchies, both quantitative and qualitative criteria can be compared, which provides a more accurate representation of real-world healthcare issues. Therefore, it should be used for complex decision problems that can completely be decomposed into a hierarchical structure. Thus, patients could apply the AHP to clarify their priorities. The patients could also benefit from these structured decisions in conversations with their physicians. The second important point is to figure out by researches which are the appropriate participants that are able to judge this research problem reliably.

Appendix

Table 1 Key words for systematic literature review

	Search terms	Pubmed	Web of Science
Block A	Analytic Hierarchy Process	481	10,127
	Analytical Hierarchy Process	486	3,148
	multi-criteria decision analysis	236	2,821
	multiple criteria decision	2,135	8,291
	stated preference	977	32,773
	Expert Choice	2,676	5,601
	pairwise comparison	2,873	10,385
Block B	Health economics	283,801	10,684
	Health care	1,346,972	412,669
Combination Block A	Analytic Hierarchy Process OR Analytical Hierarchy Process OR multi-criteria decision analysis OR multiple criteria decision OR stated preference OR Expert Choice OR pairwise comparison	9,685	68,767
	(Analytic Hierarchy Process[Title/Abstract]) OR (Analytical Hierarchy Process[Title/Abstract]) OR (multi-criteria decision analysis[Title/Abstract]) OR (multiple criteria decision[Title/Abstract]) OR (stated preference[Title/Abstract]) OR (Expert Choice[Title/Abstract]) OR (pairwise comparison[Title/Abstract])	1,966	4,923
	(Analytic Hierarchy Process[Title/Abstract]) OR (Analytical Hierarchy Process[Title/Abstract]) OR (multi-criteria decision analysis[Title/Abstract]) OR (multiple criteria decision[Title/Abstract]) OR (stated preference[Title/Abstract]) OR (pairwise comparison[Title/Abstract])	1,956	4,829
Block A AND Block B	(Analytic Hierarchy Process OR Analytical Hierarchy Process OR multi-criteria decision analysis OR multiple criteria decision OR stated preference OR pairwise comparison) AND health care	306	137
	((Analytic Hierarchy Process[Title/Abstract]) OR (Analytical Hierarchy Process[Title/Abstract]) OR (multi-criteria decision analysis[Title/Abstract]) OR (multiple criteria decision[Title/Abstract]) OR (stated preference[Title/Abstract]) OR (Expert Choice[Title/Abstract]) OR (pairwise comparison[Title/Abstract])) AND health care	307	139
Final search	(Analytic Hierarchy Process[Title/Abstract]) OR (Analytical Hierarchy Process[Title/Abstract]) OR (multi-criteria decision analysis[Title/Abstract]) OR (multiple criteria decision[Title/Abstract]) OR (stated preference[Title/Abstract]) OR (pairwise comparison[Title/Abstract]) Filter: Language English, German	1,839	4,474

Table 2 Evaluation of reporting quality

Authors	Year	Decision goal, criteria (and alternatives)	Number of participants	Type of participants	Decision	Scale for pairwise comparisons	Interview process	Software	CR	Calculation of weights	Sensitivity analysis	Reported elements
Ajami S, Ketabi S [92]	2012	yes	3 hospitals	E	g	9–1–9	f2f	Expert Choice®	n/a	EV, GA	n/a (alt)	8
Bahadori M et al. [117]	2014	yes	48	E	g	9–1–9	nominal group technique	Expert Choice®	1	n/a	n/a (alt)	8
Basoglu N et al. [69]	2012	yes	14	P	ind	n/a	n/a	n/a	n/a	n/a	n/a (alt)	4
Bi Y, Lai D, Yan H [45]	2010	yes	n/a	E	n/a	1–9	f2f	SPSS	0.1	EV	n/a	6
Cabrera-Barona P et al. [50]	2015	yes	32	E	n/a	9–1–9	n/a	n/a	0.1	n/a	n/a	5
Cancela J, Fico G, Arredondo Waldmeyer MT [51]	2015	yes	16	E	ind + g	1–9	online	BPMSG	0.1	n/a, median	n/a	9
Chen L et al. [70]	2014	yes	102	C	ind	1–5	online	n/a	0.2	n/a	n/a (alt)	7
Chung KP et al. [71]	2013	yes	66	E	ind	9–1–9	email	n/a	0.1	EV	n/a (alt)	8
Danner M et al. [72]	2011	yes	19 (12P, 7E)	E + P	g	9–1–9	f2f (workshop)	Expert Choice®	<0.1	EV, GGM	n/a	9
Diaz-Ledezma C et al. [107]	2014	yes	1	A	n/a	9–1–9	n/a	SuperDecisionsTM	0.1	n/a	yes (alt)	7
Diaz-Ledezma C, Parvizi J [73]	2013	yes	1	A	n/a	9–1–9	lit	SuperDecisionsTM	0.1	n/a	yes (alt)	8
Dolan JG et al. [25]	2013	yes	484	P	ind	9–1–9	f2f	Excel, Crystal Xcelsius, Expert Choice®	0.15	EV	n/a (alt)	9
Dou L et al. [61]	2015	yes	40	E	ind	1/9–1–9	delphi method	Expert Choice	0.1	n/a	n/a	8
Fang LF, Tung HH [104]	2010	yes	65	E	ind	n/a	questionnaire	SPSS	n/a	EV, GA	n/a	7
Guariguata L, Whiting D [110]	2011	yes	10	E	ind	5–1–5	questionnaire	n/a	n/a	n/a, GA	n/a (alt)	7
Hilgerink MP et al. [93]	2011	yes	7	E	ind + g	n/a	f2f (discussion)	Expert Choice®	n/a	EV, GGM	yes (alt)	8
Hou D et al. [67]	2014	yes	n/a	E	n/a	n/a	lit	n/a	0.1	n/a	n/a	4
Hsu HC et al. [90]	2010	yes	n/a	E	ind	5–1–5	f2f	MS Excel	n/a	EV, GGM	n/a (alt)	7
Hsu JC, Tang DH, Lu CY [63]	2015	yes	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	1
Hsu JC, Hsieh, C-Y, Yang Y-HK, Lu CY [65]	2015	yes	n/a	n/a	n/a	n/a	n/a	n/a	n/a	EV	n/a (alt)	2
Hu H et al. [68]	2010	yes	n/a	E	n/a	n/a	n/a	n/a	n/a	n/a, GGM	n/a	3
Hummel JM et al. [94]	2012	yes	6	E	ind	9–1–9	questionnaire	n/a	n/a	EV, GGM	n/a (alt)	7
Ijzerman MJ et al. [95]	2012	yes	86	E + P	ind + g	9–1–9	ppq	Expert Choice®	n/a	EV	n/a (alt)	8

Table 2 Evaluation of reporting quality *(Continued)*

Jaberidoost M et al. [66]	2015	yes	n/a	E	ind	1–9	questionnaire	Expert Choice®	n/a	EV, GGM	n/a	7
Joshi V et al. [74]	2011	yes	58	E	ind	1–11	online	n/a	0.1	n/a	n/a	7
Joshi V et al. [20]	2014	yes	422	E	ind	1–4	online	own software	0.1	n/a	n/a	8
Kadohira M [64]	2015	yes	313	E + C	ind	n/a	workshop, email	ASHtools.xls	0.15	n/a, GA	n/a (alt)	8
Karagiannidis A et al. [46]	2010	yes	n/a	E	g	1–9	n/a	Expert Choice®	0.1	EV	yes (alt)	8
Kitamura Y [47]	2010	yes	31	P	ind	1–7	online	n/a	0.3	EV	n/a (alt)	7
Krishnamoorthy K, Mahalingam M [100]	2015	yes	n/a	n/a	n/a	1/9–1–9	n/a	Expert Choice®	0.1	EV	yes (alt)	6
Kunasekaran V, Krishnamoorthy K [101]	2014	yes	n/a	n/a	n/a	1/9–1–9	n/a	Expert Choice®	0.1	n/a	yes (alt)	5
Kuruoglu E et al. [98]	2015	yes	96	P	ind	1–9	f2f	Expert Choice®	0.1	n/a, median of judgments	n/a	9
Lambooij MS, Hummel MJ [75]	2013	yes	66	E + P	ind	9–1–9	online	n/a	0.15 (in group)	EV, GA	n/a (alt)	8
Lee CW, Kwak NK [76]	2011	yes	n/a	E	n/a	n/a	n/a	n/a	0.1	EV	yes (alt)	5
Lee WC et al. [52]	2015	yes	200	C	n/a	1–9	n/a	Matlab	n/a	n/a	n/a (alt)	5
Li A-T, Lin J-W [77]	2014	yes	25	E	ind	1–9	email	Excel	0.1	n/a	n/a	8
Li C, Yu C [78]	2013	yes	n/a	E	n/a	n/a	n/a	n/a	n/a	n/a	yes (alt)	3
Lin RH, Chuang CL [91]	2010	yes	5	E	n/a	1–9	questionnaire	Expert Choice®	0.1	EV, GGM	n/a	8
Lu L et al. [53]	2015	yes	n/a	n/a	n/a	n/a	n/a	SPSS	n/a	n/a	yes	3
Maruthur NM et al. [111]	2015	yes	9	E	ind	"usual AHP scale"	computer	Expert Choice®	0.15	EV, GGM	yes (alt)	10
Mok H-P et al. [85]	2014	yes	n/a	E	n/a	n/a	n/a	n/a	0.01	n/a	n/a	3
Moslehi S, Atefi Manesh P, Sarabi Asiabar A [54]	2015	yes	5	E	n/a	1–9	n/a	K-Goepel Version 9.5.2012	0.072	n/a	n/a	6
Mühlbacher AC et al. [55]	2015	yes	1283	P	ind	9–1–9	online	n/a	0.006, 0.005	EV	n/a	8
Mühlbacher AC, Juhnke C, Kaczynski A [60]	2015	yes	24	P	ind + g	9–1–(−9)	group discussion	n/a	0.1	EV, consensus	n/a	8
Munoz DA, Nembhard HB, Kraschnewski Jennifer L [109]	2014	yes	1	A	ind	1–9	n/a	n/a	0.1	EV	n/a	7
Olivieri A et al. [79]	2012	yes	7	E	ind	1/9–1–9	questionnaire	n/a	n/a	n/a, GGM	n/a (alt)	7
Page K [80]	2012	yes	94	C	ind	9–1–9	ppq	SPSS	average at 0.3	EV	SD	10
Papadopoulos A et al. [56]	2015	yes	7	E	ind	1–9	n/a	n/a	0.1	EV, GGM	yes (alt)	8

Table 2 Evaluation of reporting quality *(Continued)*

Pecchia L et al. [81]	2011	yes	63	E	ind	5–1–5	online	n/a	0.2	EV, WM	n/a	8
Pecchia L et al. [26]	2013	yes	5	E	ind	5–1–5	ppq	n/a	0.1	EV	n/a	8
Perseghin P et al. [96]	2014	yes	11	E	g	1–9	email	n/a	n/a	n/a, GA	n/a	7
Petit J et al. [108]	2012	yes	n/a	A	n/a	9–1–9	n/a	n/a	0.1	EV	n/a (alt)	5
Ramezanpour B et al. [57]	2015	yes	24	E	g	1–9	n/a	n/a	0.1	EV	n/a	7
Reddy BP et al. [86]	2014	yes	8	E	ind + g	1/9–1–9	workshop, email	n/a	"standard"	EV, GGM and consensus	yes (alt)	9
Riepe MW [99]	2015	yes	42	E	ind	6–1–6	workshop	SPSS, spreadsheet file	0.1	n/a	n/a	8
Sharma PS et al. [82]	2011	yes	96	P	ind	9–1–9	f2f, (computer)	n/a	n/a	n/a	one-way for hybrid (alt)	7
Shojaei P et al. [87]	2014	yes	30	E	ind	9–1–1/9	f2f	Expert Choice®	0.1	EV, GGM	n/a (alt)	9
Smith J, Cook A, Packer C [48]	2010	yes	4 experienced horizon analysts	E	n/a	n/a	n/a	n/a	n/a	n/a	n/a (alt)	3
Šoltés V, Gavurová B [88]	2014	yes	16	E	ind	1–9	n/a	MS Excel	0.1 (for CI)	EV	n/a (alt)	8
Suner A et al. [83]	2012	yes	5	E	ind	9–1–9	online	Expert Choice®	0.1	EV	n/a	9
Taghipour H et al. [49]	2014	yes	40 hospitals	E	g	n/a	n/a	Expert Choice®, MS Excel	0.1	EV, WM	n/a (alt)	7
Tu C et al. [89]	2014	yes	41	E	ind	1–9	n/a	n/a	0.1	EV, GA	n/a (alt)	7
Uzoka FM et al. [97]	2011	yes	6	E	ind	9–1–9	n/a	n/a	0.2	EV, GA	n/a	7
Velmurugan R et al. [102]	2011	yes	n/a	n/a	n/a	9–1–9	n/a	n/a	0.1	AN	n/a (alt)	4
Wollmann D et al. [103]	2012	yes	400	C	ind	9–1–9	n/a	n/a	procedure by Silvac	n/a, GGM	n/a (alt)	7
Xu X, Cao Y, Luan X [58]	2014	yes	n/a	E	n/a	1–9	mobile phone app	n/a	n/a	n/a	n/a	4
Xu Y et al. [59]	2015	yes	954	P	ind	1–4	email	SAS	0.15	EV, arithmetic mean	n/a (alt)	9
Zhang S et al. [106]	2015	yes	n/a	E	n/a	1–5	n/a	JMP10.0	n/a	n/a	n/a	4
Zhu Q et al. [62]	2014	yes	9	E	ind	1–9	n/a	n/a	0.1	EV, GA	n/a	7

P patients, *C* potential consumers, *E* Experts, *n/a* not applicable, *ind* individual, *g* group, *online* online or web-based questionnaire, *f2f* face-to-face interview, *lit* literature, *quest* questionnaire (not further defined), *ppq* paper-pencil questionnaire, *email* mailed questionnaire, *CR* accepted consistency ratio, *EV* Eigenvector method, *GA* group average, *GGM* group geometric mean, *WM* weighted means, *AN* additive normalization method, *alt* alternatives included in the study, *SD* standard deviation

Table 3 List of all included studies

Author	Year	Title	Journal	Volume	Issue	Page
Ajami S, Ketabi S	2012	Performance evaluation of medical records departments by analytical hierarchy process (AHP) approach in the selected hospitals in Isfahan	Journal of Medical Systems	36	3	1165–1171
Angelucci E et al.	2008	Italian Society of Hematology practice guidelines for the management of iron overload in thalassemia major and related disorders	Hematology journal	93	5	741–752
Bahadori M et al.	2014	Assessing the service quality of Iran military hospitals: Joint Commission International standards and Analytic Hierarchy Process (AHP) technique	Journal of education and health promotion	3		98
Balestra G et al.	2007	AHP for the acquisition of biomedical instrumentation	Engineering in Medicine and Biology Society-Conference proceedings: 29th Annual International Conference of the IEEE			3581–3584
Barosi G et al.	2007	A unified definition of clinical resistance-intolerance to hydroxyurea in essential thrombocythemia: results of a consensus process by an intl. working group	Leukemia	21	2	277–280
Basoglu N, Daim TU, Topacan U	2012	Determining patient preferences for remote monitoring	Journal of Medical Systems	36	3	1389–1401
Baykasoğlu A, Dereli T, Yılankırkan N	2009	Application of cost-benefit analysis for surgical gown and drape selection: a case study	American Journal of Infection Control	37	3	215–226
Bi Y, Lai D, Yan H	2010	Synthetic evaluation of the effect of health promotion: impact of a UNICEF project in 40 poor western counties of China	Public Health	124	7	376–391
Brent A C et al.	2007	Application of the analytical hierarchy process to establish health care waste management systems that minimise infection risks in developing countries	European Journal of Operational Research	181		403–424
Cabrera-Barona P et al.	2015	A multi-criteria spatial deprivation index to support health inequality analyses	International journal of health geographics	14		11
Cancela J et al.	2015	Using the Analytic Hierarchy Process (AHP) to understand the most important factors to design and evaluate a telehealth system for Parkinson's disease	BMC medical informatics and decision making	15	Suppl 3	S7
Carter KJ et al.	1999	Analysis of three decision-making methods: a breast cancer patient as a model	Medical Decision Making	19	1	49–57
Castro F et al.	1996	Sequential test selection in the analysis of abdominal pain	Medical Decision Making	16	2	178–183

Table 3 List of all included studies *(Continued)*

Chang PY et al.	2006	Factors influencing medical students' choice of specialty	Journal of the Formosan Medical Association = Taiwan yi zhi	105	6	489–496
Cheever MA et al.	2009	The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research	Clinical Cancer Research	15	17	5323–5337
Chen L et al.	2014	Development of a decision support engine to assist patients with hospital selection	Journal of medical systems	38	6	59
Cho KT, Kim SM	2003	Selecting medical devices and materials for development in Korea: the analytic hierarchy process approach	International Journal of Health Planning and Management	18	2	161–174
Chung KP et al.	2013	Application of the analytic hierarchy process in the performance measurement of colorectal cancer care for the design of a pay-for-performance program in Taiwan	International Journal for Quality in Health Care	25	1	81–91
Cook DR, Staschak S, Green WT	1990	Equitable allocation of livers for orthotopic transplantation: an application of the Analytic Hierarchy Process	European Journal of Operational Research	48	1	49–56
Czaja S et al.	2003	A methodology for describing and decomposing complex psychosocial and behavioral interventions	Psychology and Aging	18	3	385–395
da Rocha LS, Sloane EB, Bassani JWM	2005	Optimal Medical Equipment Maintenance Service Proposal Decision Support System combining Activity Based Costing (ABC) and the Analytic Hierarchy Process (AHP)	Conference proceedings: 27th Annual International Conference of the IEEE Engineering in Medicine and Biology Society	7	7	103–106
Danner M et al.	2011	Integrating patients' views into health technology assessment: AHP as a method to elict patient preferences	International Journal of Technology Assessment in Health Care	27	4	369–375
Dey P K, Hariharan S, Clegg B	2006	Measuring the operational performance of intensive care units using the analytic hierarchy process approach	International Journal of Operations& Production Management	26	8	849–865
Diaz-Ledezma C et al.	2014	Diagnosis of Periprosthetic Joint Infection in Medicare Patients: Multicriteria Decision Analysis	Clinical Orthopaedics and Related Research	URL: http://link.springer.com/article/10.1007%2Fs11999-014-3492-2 Accessed: 15 Feb 2014		
Diaz-Ledezma C et al.	2014	Diagnosis of Periprosthetic Joint Infection in Medicare Patients: Multicriteria Decision Analysis	Clinical Orthopaedics and Related Research	472	11	3275–3284
Diaz-Ledezma C, Parvizi J	2013	Surgical approaches for cam femoroacetabular impingement: the use of multicriteria decision analysis	Clinical Orthopaedics and Related Research	471	8	2509–2516
Dolan JG	2000	Involving patients in decisions regarding preventive health interventions using the analytic hierarchy process	Health Expectations	3	1	37–45

Table 3 List of all included studies *(Continued)*

Dolan JG	1995	Are patients capable of using the analytic hierarchy process and willing to use it to help make clinical decisions	Medical Decision Making	15	1	76–80
Dolan JG	1990	Can decision analysis adequately represent clinical problems?	Journal of Clinical Epidemiology	43	3	277–284
Dolan JG	1989	Medical decision making using the analytic hierarchy process: choice of initial antimicrobial therapy for acute pyelonephritis	Medical Decision Making	9	1	51–56
Dolan JG et al.	2013	Patients' Preferences and Priorities Regarding Colorectal Cancer Screening	Medical Decision Making	33	1	59–70
Dolan JG, Bordley DR	1994	Isoniazid prophylaxis-The Importance of Individual Values	Medical Decision Making	14	1	1–8
Dolan JG, Bordley DR	1993	Involving patients in complex decisions about their care: an approach using the analytic hierarchy process	Journal of General Internal Medicine	8	4	204–209
Dolan JG, Bordley DR, Miller H	1993	Diagnostic strategies in the management of acute upper gastrointestinal bleeding: patient and physician preferences	Journal of General Internal Medicine	8	10	525–529
Dolan JG, Frisina S	2002	Randomized controlled trial of a patient decision aid for colorectal cancer screening	Medical Decision Making	22	2	125–139
Dolan JG, Iadarola S	2008	Risk communication formats for low probability events— an exploratory study of patient preferences	BMC medical informatics and decision making [electronic resource]	URL: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2330036/ Accessed 31 Dec 2013.		
Dolan JG, Isselhardt BJ, Cappuccio JD	1989	The analyic hierarchy process in medical decision making: a tutorial	Medical Decision Making	9	1	40-50
Dou L et al.	2015	An evaluation system for financial compensation in traditional Chinese medicine services	Complementary therapies in medicine	23	5	637–643
Eden KB et al.	2009	Patients were more consistent in randomized trial at prioritizing childbirth preferences using graphic-numeric than verbal formats	Journal of Clinical Epidemiology	62	4	415–424
Fang LF, Tung HH	2010	Comparison of nurse practitioner job core competency expectations of nurse managers, nurse practitioners, and physicians in Taiwan	Journal of the American Academy of Nurse Practitioners	22	8	409–416
Guariguata L, Whiting D	2011	The International Diabetes Federation diabetes atlas methodology for estimating global and national prevalence of diabetes in adults	Diabetes Research and Clinical Practice	94	3	322–332
Hannan EL, O'Donnell J, Freedland T	1981	A priority assignment model for standards and conditions in a long term care survey	Socio-economic Planning Sciences	15	6	277–289

Table 3 List of all included studies *(Continued)*

Hariharan S et al.	2005	Application of analytic hierarchy process for measuring and comparing the global performance of intensive care units	Journal of Critical Care	20	2	117–125
Hariharan S et al.	2004	A new tool for measurement of process-based performance of multispecialty tertiary care hospitals	International Journal of Health Care Quality Assurance	17	6	302–312
Hilgerink MP et al.	2011	Assessment of the added value of the Twente Photoacoustic Mammoscope in breast cancer diagnosis	Medical Devices: Evidence and Research	4		107–115
Hou D et al.	2014	A real-time, dynamic early-warning model based on uncertainty analysis and risk assessment for sudden water pollution accidents	Environmental science and pollution research international	21	14	8878–8892
Hsu HC et al.	2010	Constructing area-level indicators of successful ageing in Taiwan	Health and Social Care in the Community	18	1	70–81
Hu H et al.	2010	Establishment and evaluation of a model of a community health service in an underdeveloped area of China	Public Health	124	4	206–217
Hummel JM et al.	2012	Predicting the Health Economic Performance of new non-fusion Surgery in Adolescent Idiopathic Scoliosis	Orthopaedic Research Society	30	9	1453–1458
Hummel JM et al.	2005	A multicriteria decision analysis of augmentative treatment of upper limbs in persons with tetraplegia	Journal of Rehabilitation Research & Development	42	5	635–544
Hummel JM et al.	2000	Medical technology assessment: the use of the analytic hierarchy process as a tool for multidisciplinary evaluation of medical devices	International Journal of Artificial Organs	23	11	782–787
Hsu JC et al.	2015	Net clinical benefit of oral anticoagulants: a multiple criteria decision analysis	PLoS One	10	4	e0124806
Hsu JC, Tang DH, Lu CY	2015	Risk-benefit assessment of oral phosphodiesterase type 5 inhibitors for treatment of erectile dysfunction: a multiple criteria decision analysis	International journal of clinical practice	69	4	436–443
Ijzerman MJ, van Til JA, Bridges JFP	2012	A Comparison of Analytic Hierarchy Process and Conjoint Analysis Methods in assessing treatment alternatives for stroke Rehabilitation	The Patient	5	1	45–56
Ijzerman MJ, van Til JA, Snoek GJ	2008	Comparison of two multi-criteria decision techniques for eliciting treatment preferences in people with neurological disorders	The Patient	1	4	265–272
Jaberidoost M et al.	2015	Pharmaceutical supply chain risk assessment in Iran using analytic hierarchy process (AHP) and simple additive weighting (SAW) methods	Journal of pharmaceutical policy and practice	8	1	9

Table 3 List of all included studies *(Continued)*

Javalgi RG, Rao SR, Thomas EG	1991	Choosing a hospital: analysis of consumer tradeoffs	Journal of Health Care Marketing	11	1	12–22
Joshi V et al.	2011	Empirical investigation of radiologists' priorities for PACS selection: an analytical hierarchy process approach	Journal of Digital Imaging	24	4	700–708
Joshi V et al.	2014	PACS Administrators' and Radiologists' Perspective on the Importance of Features for PACS Selection	Journal of digital imaging	27	4	486–495
Kadohira M et al.	2015	Stakeholder prioritization of zoonoses in Japan with analytic hierarchy process method	Epidemiology and infection	143	7	1477–1485
Karagiannidis A et al.	2010	A multi-criteria assessment of scenarios on thermal processing of infectious hospital wastes: a case study for Central Macedonia	Waste Management	30	2	251–262
Karamouz M et al.	2007	Developing a master plan for hospital solid waste management: a case study	Waste Management	27	5	626–638
Katsumura Y et al.	2008	Relationship between risk information on total colonoscopy and patient preferences for colorectal cancer screening options: Analysis using the Analytic Hierarchy Process	BMC Health Services Research	8	1	
Kitamura Y	2010	Decision-making process of patients with gynecological cancer regarding their cancer treatment choices using the analytic hierarchy process	Japan Journal of Nursing Science	7	2	148–157
Koch T, Ridgley M	1998	Distanced Perspectives: AIDS, Anencephaly, and AHP	Theoretical Medicine and Bioethics	19	1	47–58
Koch T, Rowell M	1999	The dream of consensus: finding common ground in a bioethical context	Theoretical Medicine and Bioethics	20	3	261–273
Krishnamoorthy K, Mahalingam M	2015	Selection of a suitable method for the preparation of polymeric nanoparticles: multi-criteria decision making approach	Advanced pharmaceutical bulletin	5	1	57–67
Kunasekaran V, Krishnamoorthy K	2014	Multi criteria decision making to select the best method for the preparation of solid lipid nanoparticles of rasagiline mesylate using analytic hierarchy process	Journal of advanced pharmaceutical technology & research	5	3	115–121
Kuruoglu E, Guldal D, Mevsim V, Gunvar T	2015	Kuruoglu, Emel; Guldal, Dilek; Mevsim, Vildan; Gunvar, Tolga	BMC medical informatics and decision making	15		63
Kwak NK, McCarthy K, Parker GE	1997	A human resource planning model for hospital/medical technologists: an analytic hierarchy process approach	Journal of Medical Systems	21	3	173–187
Lambooij MS, Hummel MJ	2013	Differentiating innovation priorities among stakeholder in hospital care	BMC Medical Informatics and Decision Making	URL: http://www.biomedcentral.com/1472-6947/13/91 Accessed 31 Dec 2013.		

Table 3 List of all included studies *(Continued)*

Lee CW, Kwak NK	2011	Strategic Enterprise Resource Planning in a Health-Care System Using a Multicriteria Decision-Making Model	Journal of Medical Systems	35	2	265–275
Lee WC et al.	2015	A speedy cardiovascular diseases classifier using multiple criteria decision analysis	Sensors	15	1	1312–1320
Li A-T, Lin J-W	2014	Constructing core competency indicators for clinical teachers in Taiwan: a qualitative analysis and an analytic hierarchy process	BMC medical education	14		75
Li C, Yu C	2013	Performance Evaluation of Public Non-Profit Hospitals Using a BP Artificial Neural Network-The Case of Hubei Province in China	International Journal of Environmental Research and Public Health	10	8	3619–3633
Lin RH, Chuang CL	2010	A hybrid diagnosis model for determining the types of the liver disease	Computers in Biology and Medicine	40	7	665–670
Lu L et al.	2015	Assessment of regional human health risks from lead contamination in Yunnan province, southwestern China	PLoS One	10	3	e0119562
Maruthur NM et al.	2015	Use of the analytic hierarchy process for medication decision-making in type 2 diabetes	PLoS One	10	5	e0126625
Matsuda S	1996	An analysis of the Vietnamese system of occupational safety and health and setting priorities with the analytical hierarchy process	Occupational and Environmental Medicine	53	4	281–286
Matsuda S, Washino K	1998	How do the Japanese medical students evaluate the effectiveness of anti-smoking strategies- an application of the Analytic Hierarchy Process	Environmental Health and Preventive Medicine	3	2	73–77
Mok H-P et al.	2014	Development and validation of a convenient formula evaluating the value and applicability of medical literature in clinical practice	Pakistan journal of medical sciences	30	6	1377–1382
Moslehi S, Atefi Manesh P, Sarabi Asiabar A	2015	Quality measurement indicators for Iranian Health Centers	Medical journal of the Islamic Republic of Iran	29		177
Mühlbacher AC et al.	2015	Objective Criteria in the Medicinal Therapy for Type II Diabetes: An Analysis of the Patients' Perspective with Analytic Hierarchy Process and Best-Worst Scaling	Gesundheitswesen			online
Mühlbacher AC, Juhnke C, Kaczynski A	2015	Patients' Priorities in the Treatment of Neuroendocrine Tumours: An Analytical Hierarchy Process	Gesundheitswesen			online
Munoz DA, Nembhard HB, Kraschnewski JL	2014	Quantifying complexity in translational research: an integrated approach	International journal of health care quality assurance	27	8	760–776
Nuijten MJC, Kosa J	2004	Pricing of pharmaceuticals: Assessing the pricing potential by a pricing matrix model	The European Journal of Health Economics	5	2	110–115

Table 3 List of all included studies *(Continued)*

Oddershede Herrera A, Carrasco González R, Barham Abu-Muhor E	2008	Multi-criteria Decision Model for Assessing Health Service Information Technology Network Support Using the Analytic Hierarchy Process	Computatión y sistemas	12	2	173–182
Olivieri A et al.	2012	Proposed definition of 'poor mobilizer' in lymphoma and multiple myeloma: an analytic hierarchy process by ad hoc working group Gruppo ItalianoTrapianto di Midollo	Bone Marrow Transplantation	47	3	342–351
Page K	2012	The four principles: can they be measured and do they predict ethical decision making	BMC Medical Ethics	13	10	online
Papadopoulos A et al.	2015	TDS exposure project: application of the analytic hierarchy process for the prioritization of substances to be analyzed in a total diet study	Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association	76		46–53
Pecchia L et al.	2011	Analytic Hierarchy Process (AHP) for examining healthcare professionals' assessments of risk factors	Methods of Information in Medicine	50	5	435–444
Pecchia L et al.	2013	User needs elicitation via analytic hierarchy process (AHP). A case study on a Computed Tomography (CT) scanner	BMC medical informatics and decision making [electronic resource]	13	2	
Perseghin P et al.	2014	A policy for the disposal of autologous hematopoietic progenitor cells: report from an Italian consensus panel	Transfusion	54	9	2353–2360
Petit J et al.	2012	Softening the Rule of Five- where to draw the line?	Bioorganic & Medicinal Chemistry	20	18	5343–5351
Ramezanpour B et al.	2015	Market implementation of the MVA platform for pre-pandemic and pandemic influenza vaccines: A quantitative key opinion leader analysis	Vaccine	33	35	4349–4358
Reddy BP et al.	2014	Prioritising public health guidance topics in the National Institute for Health and Care Excellence using the Analytic Hierarchy Process	Public Health	128	10	896–903
Richman MB et a.l	2005	A novel computer based expert decision making model for prostate cancer disease management	The Journal of Urology	174	6	2310–2318
Riepe MW	2015	Clinical preference for factors in treatment of geriatric depression	Neuropsychiatric disease and treatment	11		25–31
Ross ME, Nydick RL	1992	Selection of licensing candidates in the pharmaceutical industry: an application of the analytic hierarchy process	Journal of Health Care Marketing	12	2	60–65
Sharma PS et al.	2011	Subjective risk vs. objective risk can lead to different post-cesarean birth decisions based on multiattribute modeling	Journal of Clinical Epidemiology	64	1	67–78
Shin T et al.	2009	The comparative evaluation of expanded national immunization policies in Korea using an analytic hierarchy process	Vaccine	27	5	792–802

Table 3 List of all included studies *(Continued)*

Shojaei P et al.	2014	Ranking the effects of urban development projects on social determinants of health: health impact assessment	Global journal of health science	6	5	183–195
Singh S, Dolan JG, Centor RM	2006	Optimal management of adults with pharyngitis-a multi-criteria decision analysis	BMC Medical Informatics and Decision Making	6	14	online
Smith J, Cook A, Packer C	2010	Evaluation criteria to assess the value of identification sources for horizon scanning	International Journal of Technology Assessment in Health Care	26	3	348–353
Šoltés V, Gavurová B	2014	The functionality comparison of the health care systems by the analytical hierarchy process method	E + M Ekonomie a Management	17	3	100–117
Suner A et al.	2012	Sequential decision tree using the analytic hierarchy process for decision support in rectal cancer	Artificial Intelligence in Medicine	56	1	59–68
Taghipour H et al.	2014	On-site or off-site treatment of medical waste: a challenge	Journal of environmental health science & engineering	12	1	68
Tan X et al.	2007	Evaluation of the Effect of a Health Education Campaign of HIV by Using an Analytical Hierarchy Process Method	International journal of Environmental Research and Public Health	4	3	254–259
Tarimcilar MM, Khaksari SZ	1991	Capital budgeting in hospital management using the analytic hierarchy process	Socio-economic Planning Sciences	25	1	27–34
Tu C et al.	2014	Application of the analytic hierarchy process to a risk assessment of emerging infectious diseases in Shaoxing city in southern China	Japanese journal of infectious diseases	67	6	417–422
Tzung TY et al.	2007	Decision factors and the recognition of medical specialty in patients receiving cosmetic laser and intense pulsed light treatment	Dermatologic Surgery	33	12	1488–1493
Uzoka FM et al.	2011	An experimental comparison of fuzzy logic and analytic hierarchy process for medical decision support systems	Computer Methods and Programs in Biomedicine	103	1	10–27
van Til JA et al.	2008	The use of the analytic hierarchy process to aid decision making in acquired equinovarus deformity	Archives of Physical Medicine and Rehabilitation	89	3	457–462
Velmurugan R, Selvamuthukumar S, Manavalan R	2011	Multi criteria decision making to select the suitable method for the preparation of nanoparticles using an analytical hierarchy process	Die Pharmazie	66	11	836–842
Wang KI et al.	2007	Analysis of senior medical students' preferences in specialty choice a survey in a medical school in northern Taiwan	Chang Gung Medical Journal	30	4	339–353
Weingarten MS et al.	1997	A pilot study of the use of analytic hierarchy process for the Selection of Surgery Residents	Academic medicine: journal of the Association of American Medical Colleges	72	5	400–402

Table 3 List of all included studies *(Continued)*

Wollmann D et al.	2012	Evaluation of health service providers by consumers through the Analytic Hierarchy Process Method	Revista de Saúde Pública	46	5	777–783
Wu C, Lin C, Chen H.	2007	Optimal selection of location for Taiwanese hospitals to ensure a competitive advantage by using the analytic hierarchy process and sensitivity analysis	Building and Environment	42		1431–1444
Xu X, Cao Y, Luan X	2014	Application of 4G wireless network-based system for remote diagnosis and nursing of stomal complications	International journal of clinical and experimental medicine	7	11	4554–4561
Xu Y et al.	2015	Comparison of patient preferences for fecal immunochemical test or colonoscopy using the analytic hierarchy process	BMC health services research	15		175
Zhang S et al.	2015	Indicators for Environment Health Risk Assessment in the Jiangsu Province of China	International journal of environmental research and public health	12	9	11012–11024
Zhu Q et al.	2014	The spatial distribution of health vulnerability to heat waves in Guangdong Province, China	Global Health Action	7		25051

Abbreviations
AHP: Analytic Hierarchy Process; CHERH: Center for Health Economics Research Hannover; CONSORT: Consolidated Standards of Reporting Trials; CR: Consistency Ratio; EV: Eigenvector method; IQWiG: Institute for Quality and Efficiency in Health Care; ISPOR: International Society for Pharmacoeconomics and Outcomes Research; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses.

Competing interests
The author(s) declare that they have no competing interests.

Authors' contributions
KS carried out the analyses and drafted the manuscript. IA and IH participated in the review process and decision making process for identifying relevant articles. IA made substantial contributions to conception of the article. IH collected and prepared the data adequately for the manuscript. KD participated in selection process of papers and she revised the manuscript. JMS revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements
The Center for Health Economics Research Hannover (CHERH) is founded by the Federal Ministry of Education and Research.

Author details
[1]Center for Health Economics Research Hannover (CHERH), Leibniz University of Hanover, Otto-Brenner-Str. 1, 30159 Hannover, Germany. [2]Institute for Risk and Insurance, Leibniz University of Hanover, Otto-Brenner-Str. 1, 30159 Hannover, Germany.

Received: 14 August 2015 Accepted: 15 December 2015
Published online: 24 December 2015

References

1. Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelneuordnungsgesetz –AMNOG).
2. Hailey D, Nordwall M. Survey on the involvement of consumers in health technology assessment programs. Int J Technol Assess Health Care. 2006; 22(4):497–9.
3. Buttorff C. What should be the role of patient preferences in making health care resource allocation decisions?; Available from: http://www.ispor.org/News/articles/August10/What-Should-Be-the-Role-of-Patient-Preferences.asp.
4. Bruera E, Sweeney C, Calder K, Palmer L, Benisch-Tolley S. Patient preferences versus physician perceptions of treatment decisions in cancer care. J Clin Oncol. 2001;19:2883–5.
5. Mühlbacher AC, Juhnke C. Patient preferences versus physicians' judgement: does it make a difference in healthcare decision making? Appl Health Econ Health Policy. 2013;11:163–80.
6. Gaston CM, Mitchell G. Information giving and decision-making in patients with advanced cancer: a systematic review. Soc Sci Med. 2005;61:2252–64.
7. Dolan JG. Multi-criteria clinical decision support: A primer on the use of multiple criteria decision making methods to promote evidence-based, patient-centered healthcare. Patient. 2010;3:229–48.
8. Institute for Quality and Efficiency in Health Care. Allgemeine Methoden: Entwurf für Version 4.2 vom 18.06.2014. [November 27, 2014]; Available from: https://www.iqwig.de/download/IQWiG_Methoden_Entwurf-fuer-Version-4-2.pdf.
9. Bridges JFP, Hauber AB, Marshall D, Lloyd A, Prosser LA, Regier DA, et al. Conjoint analysis applications in health–a checklist: a report of the ISPOR Good Research Practices for Conjoint Analysis Task Force. Value Health. 2011;14:403–13.
10. Marshall D, Bridges JFP, Hauber B, Cameron R, Donnalley L, Fyie K, et al. Conjoint Analysis Applications in Health - How are Studies being Designed and Reported?: An Update on Current Practice in the Published Literature between 2005 and 2008. The patient. 2010;3:249–56.
11. Ryan M. Using conjoint analysis to elicit preferences for health care. BMJ. 2000;320:1530–3.
12. Saaty TL. A scaling method for priorities in hierarchical structures. J Math Psychol. 1977;15:234–81.
13. Saaty TL. The analytic hierarchy process: planning, priority setting, resource allocation 1980.
14. Dolan JG. Medical decision making using the analytic hierarchy process: choice of initial antimicrobial therapy for acute pyelonephritis. Med Decis Making. 1989;9:51–6.
15. Dolan JG, Isselhardt Jr BJ, Cappuccio JD. The analytic hierarchy process in medical decision making: a tutorial. Med Decis Making. 1989;9:40–50.
16. Liberatore MJ, Nydick RL. The analytic hierarchy process in medical and health care decision making: A literature review. Eur J Oper Res. 2008;189: 194–207.
17. Cook DR, Staschak S, Green WT. Equitable allocation of livers for orthotopic transplantation: an application of the Analytic Hierarchy Process. Eur J Oper Res. 1990;48:49–56.
18. Dolan JG, Bordley DR, Miller H. Diagnostic strategies in the management of acute upper gastrointestinal bleeding: patient and physician preferences. J Gen Intern Med. 1993;8:525–9.
19. Cheever MA, Allison JP, Ferris AS, Finn OJ, Hastings BM, Hecht TT, et al. The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research. Clin Cancer Res. 2009;15:5323–37.
20. Joshi V, Narra VR, Joshi K, Lee K, Melson D. PACS Administrators' and Radiologists' Perspective on the Importance of Features for PACS Selection. J Digit Imaging. 2014;27:486–95.
21. de Bekker-Grob EW, Ryan M, Gerard K. Discrete choice experiments in health economics: a review of the literature. Health Econ. 2012;21:145–72.
22. Hummel M, IJzerman M (eds.). The past and future of the AHP in health care decision making; 2011.
23. Mühlbacher A, Kaczynski A. Der Analytic Hierarchy Process (AHP): Eine Methode zur Entscheidungsunterstützung im Gesundheitswesen. PharmacoEcon Ger Res Artic. 2013;11:119–32.
24. Saaty RW. The analytic hierarchy process—what it is and how it is used. Mathematical Modelling. 1987;9:161–76.
25. Dolan JG, Boohaker E, Allison J, Imperiale TF. Patients' preferences and priorities regarding colorectal cancer screening. Med Decis Making. 2013;33: 59–70.
26. Pecchia L, Martin JL, Ragozzino A, Vanzanella C, Scognamiglio A, Mirarchi L, et al. User needs elicitation via analytic hierarchy process (AHP). A case study on a Computed Tomography (CT) scanner. BMC Med Inform Decis Mak. 2013;13:2.
27. Srdjevic B. Combining different prioritization methods in the analytic hierarchy process synthesis. Comput Oper Res. 2005;32:1897–919.
28. Saaty TL. Decision making with the analytic hierarchy process. International journal of services sciences. 2008;1:83–98.
29. Meixner O, Haas R. Wissensmanagement und Entscheidungstheorie: Mit 35 Tabellen. Wien: Facultas.wuv; 2010.
30. Forman E, Peniwati K. Aggregating individual judgments and priorities with the analytic hierarchy process. Eur J Oper Res. 1998;108:165–9.
31. Chen H, Kocaoglu DF. A sensitivity analysis algorithm for hierarchical decision models. Eur J Oper Res. 2008;185:266–88.
32. Saaty TL, Vargas LG. Sensitivity analysis in the analytic hierarchy process. In: Saaty TL, Vargas LG, editors. Decision making with the analytic network process. Boston: Springer US; 2013. p. 345–60.
33. Arbel A. Approximate articulation of preference and priority derivation. Eur J Oper Res. 1989;43:317–26.
34. Moreno-Jimenez JM, Vargas LG. A probabilistic study of preference structures in the analytic hierarchy process with interval judgments. Math Comput Model. 1993;17:73–81.
35. Sugihara K, Tanaka H. Interval evaluations in the analytic hierarchy process by possibility analysis. Computational Intell. 2001;17:567–79.
36. Triantaphyllou E, Sánchez A. A sensitivity analysis approach for some deterministic multi-criteria decision-making methods. Decis Sci. 1997;28: 151–94.
37. Sowlati T, Assadi P, Paradi JC. Developing a mathematical programming model for sensitivity analysis in analytic hierarchy process. IJMOR. 2010;2:290.
38. Masuda T. Hierarchical sensitivity analysis of priority used in analytic hierarchy process. Int J of Systems Sc. 1990;21:415–27.
39. Huang Y. Enhancement on sensitivity analysis of priority in analytic hierarchy process. Int J Gen Syst. 2010;31:531–42.
40. Erkut E, Tarimcilar M. On sensitivity analysis in the analytic hierarchy process. IMA J Management Math. 1991;3:61–83.
41. Altuzarra A, Moreno-Jiménez JM, Salvador M. Consensus building in AHP-group decision making: A Bayesian approach. Oper Res. 2010;58:1755–73.
42. Wang Y, Luo Y. On rank reversal in decision analysis. Math Comput Model. 2009;49:1221–9.

43. Moher D, Schulz KF, Altman DG. The CONSORT statement. Revised recommendations for improving the quality of reports of parallel group randomized trails. BMC Med Res Methodol. 2001;1:2.
44. Moher D. Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. Ann Intern Med. 2009;151:264.
45. Bi Y, Lai D, Yan H. Synthetic evaluation of the effect of health promotion: impact of a UNICEF project in 40 poor western counties of China. Public Health. 2010;124:376–91.
46. Karagiannidis A, Papageorgiou A, Perkoulidis G, Sanida G, Samaras P. A multi-criteria assessment of scenarios on thermal processing of infectious hospital wastes: a case study for Central Macedonia. Waste Manag. 2010;30:251–62.
47. Kitamura Y. Decision-making process of patients with gynecological cancer regarding their cancer treatment choices using the analytic hierarchy process. Jpn J Nurs Sci. 2010;7:148–57.
48. Smith J, Cook A, Packer C. Evaluation criteria to assess the value of identification sources for horizon scanning. Int J Technol Assess Health Care. 2010;26:348–53.
49. Taghipour H, Mohammadyarei T, Asghari Jafarabadi M, Asl HA. On-site or off-site treatment of medical waste: a challenge. J Environ Health Sci Eng. 2014;12:68.
50. Cabrera-Barona P, Murphy T, Kienberger S, Blaschke T. A multi-criteria spatial deprivation index to support health inequality analyses. Int J Health Geogr. 2015;14:11.
51. Cancela J, Fico G, Arredondo Waldmeyer MT. Using the Analytic Hierarchy Process (AHP) to understand the most important factors to design and evaluate a telehealth system for Parkinson's disease. BMC Med Inform Decis Mak. 2015;15 Suppl 3:S7.
52. Lee WC, Hung FH, Tsang KF, Tung HC, Lau WH, Rakocevic V, et al. A speedy cardiovascular diseases classifier using multiple criteria decision analysis. Sensors (Basel). 2015;15:1312–20.
53. Lu L, Cheng H, Liu X, Xie J, Li Q, Zhou T. Assessment of regional human health risks from lead contamination in Yunnan province, southwestern China. PLoS One. 2015;10:e0119562.
54. Moslehi S, Atefi Manesh P, Sarabi AA. Quality measurement indicators for Iranian Health Centers. Med J Islam Repub Iran. 2015;29:177.
55. Mühlbacher AC, Bethge S, Kaczynski A, Juhnke C. Objective Criteria in the Medicinal Therapy for Type II Diabetes: An Analysis of the Patients' Perspective with Analytic Hierarchy Process and Best-Worst Scaling. Gesundheitswesen. 2015. https://www.thieme-connect.com/DOI/DOI?10.1055/s-0034-1390474.
56. Papadopoulos A, Sioen I, Cubadda F, Ozer H, Basegmez HIO, Turrini A, et al. TDS exposure project: application of the analytic hierarchy process for the prioritization of substances to be analyzed in a total diet study. Food Chem Toxicol. 2015;76:46–53.
57. Ramezanpour B, Pronker ES, Kreijtz JHCM, Osterhaus ADME, Claassen E. Market implementation of the MVA platform for pre-pandemic and pandemic influenza vaccines: A quantitative key opinion leader analysis. Vaccine. 2015;33:4349–58.
58. Xu X, Cao Y, Luan X. Application of 4G wireless network-based system for remote diagnosis and nursing of stomal complications. Int J Clin Exp Med. 2014;7:4554–61.
59. Xu Y, Levy BT, Daly JM, Bergus GR, Dunkelberg JC. Comparison of patient preferences for fecal immunochemical test or colonoscopy using the analytic hierarchy process. BMC Health Serv Res. 2015;15:175.
60. Mühlbacher AC, Juhnke C, Kaczynski A. Patients' Priorities in the Treatment of Neuroendocrine Tumours: An Analytical Hierarchy Process. Gesundheitswesen. 2015. https://www.thieme-connect.com/DOI/DOI?10.1055/s-0035-1548932.
61. Dou L, Yin A, Hao M, Lu J. An evaluation system for financial compensation in traditional Chinese medicine services. Complement Ther Med. 2015;23:637–43.
62. Zhu Q, Liu T, Lin H, Xiao J, Luo Y, Zeng W, et al. The spatial distribution of health vulnerability to heat waves in Guangdong Province. China Glob Health Action. 2014;7:25051.
63. Hsu JC, Tang DH, Lu CY. Risk-benefit assessment of oral phosphodiesterase type 5 inhibitors for treatment of erectile dysfunction: a multiple criteria decision analysis. Int J Clin Pract. 2015;69:436–43.
64. Kadohira M, Hill G, Yoshizaki R, Ota S, Yoshikawa Y. Stakeholder prioritization of zoonoses in Japan with analytic hierarchy process method. Epidemiol Infect. 2015;143:1477–85.
65. Hsu JC, Hsieh C, Yang YK, Lu CY. Net clinical benefit of oral anticoagulants: a multiple criteria decision analysis. PLoS One. 2015;10:e0124806.
66. Jaberidoost M, Olfat L, Hosseini A, Kebriaeezadeh A, Abdollahi M, Alaeddini M, et al. Pharmaceutical supply chain risk assessment in Iran using analytic hierarchy process (AHP) and simple additive weighting (SAW) methods. J Pharm Policy Pract. 2015;8:9.
67. Hou D, Ge X, Huang P, Zhang G, Loaiciga H. A real-time, dynamic early-warning model based on uncertainty analysis and risk assessment for sudden water pollution accidents. Environ Sci Pollut Res Int. 2014;21:8878–92.
68. Hu H, Liang W, Liu M, Li L, Li Z, Li T, et al. Establishment and evaluation of a model of a community health service in an underdeveloped area of China. Public Health. 2010;124:206–17.
69. Basoglu N, Daim TU, Topacan U. Determining patient preferences for remote monitoring. J Med Syst. 2012;36:1389–401.
70. Chen L, Chan C, Lee H, Chung Y, Lai F. Development of a decision support engine to assist patients with hospital selection. J Med Syst. 2014;38:59.
71. Chung K, Chen L, Chang Y, Chang Y, Lai M. Application of the analytic hierarchy process in the performance measurement of colorectal cancer care for the design of a pay-for-performance program in Taiwan. Int J Qual Health Care. 2013;25:81–91.
72. Danner M, Hummel JM, Volz F, van Manen JG, Wiegard B, Dintsios C, et al. 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;27:369–75.
73. Diaz-Ledezma C, Parvizi J. Surgical approaches for cam femoroacetabular impingement: the use of multicriteria decision analysis. Clin Orthop Relat Res. 2013;471:2509–16.
74. Joshi V, Lee K, Melson D, Narra VR. Empirical investigation of radiologists' priorities for PACS selection: an analytical hierarchy process approach. J Digit Imaging. 2011;24:700–8.
75. Lambooij MS, Hummel MJ. Differentiating innovation priorities among stakeholder in hospital care. BMC Med Inform Decis Mak. 2013;13:91.
76. Lee CW, Kwak NK. Strategic enterprise resource planning in a health-care system using a multicriteria decision-making model. J Med Syst. 2011;35:265–75.
77. Li A, Lin J. Constructing core competency indicators for clinical teachers in Taiwan: a qualitative analysis and an analytic hierarchy process. BMC Med Educ. 2014;14:75.
78. Li C, Yu C. Performance evaluation of public non-profit hospitals using a BP artificial neural network: the case of Hubei Province in China. Int J Environ Res Public Health. 2013;10:3619–33.
79. Olivieri A, Marchetti M, Lemoli R, Tarella C, Iacone A, Lanza F, et al. Proposed definition of 'poor mobilizer' in lymphoma and multiple myeloma: an analytic hierarchy process by ad hoc working group Gruppo ItalianoTrapianto di Midollo Osseo. Bone Marrow Transplant. 2012;47:342–51.
80. Page K. The four principles: can they be measured and do they predict ethical decision making? BMC Med Ethics. 2012;13:10.
81. Pecchia L, Bath PA, Pendleton N, Bracale M. Analytic Hierarchy Process (AHP) for examining healthcare professionals' assessments of risk factors. The relative importance of risk factors for falls in community-dwelling older people. Methods Inf Med. 2011;50:435–44.
82. Sharma PS, Eden KB, Guise J, Jimison HB, Dolan JG. Subjective risk vs. objective risk can lead to different post-cesarean birth decisions based on multiattribute modeling. J Clin Epidemiol. 2011;64:67–78.
83. Suner A, Celikoglu CC, Dicle O, Sokmen S. Sequential decision tree using the analytic hierarchy process for decision support in rectal cancer. Artif Intell Med. 2012;56:59–68.
84. Bahadori M, Ravangard R, Yaghoubi M, Alimohammadzadeh K. Assessing the service quality of Iran military hospitals: Joint Commission International standards and Analytic Hierarchy Process (AHP) technique. J Educ Health Promot. 2014;3:98.
85. Mok H, Zhou Y, Chen J, Gao Q. Development and validation of a convenient formula evaluating the value and applicability of medical literature in clinical practice. Pak J Med Sci. 2014;30:1377–82.
86. Reddy BP, Kelly MP, Thokala P, Walters SJ, Duenas A. Prioritising public health guidance topics in the National Institute for Health and Care Excellence using the Analytic Hierarchy Process. Public Health. 2014;128:896–903.
87. Shojaei P, Karimlou M, Nouri J, Mohammadi F, Malek Afzali H, Forouzan AS. Ranking the effects of urban development projects on social determinants of health: health impact assessment. Glob J Health Sci. 2014;6:183–95.
88. Šoltés V, Gavurová B. The functionality comparison of the health care systems by the analytical hierarchy process method. E + M 2014;17:100–17.

89. Tu C, Fang Y, Huang Z, Tan R. Application of the analytic hierarchy process to a risk assessment of emerging infectious diseases in Shaoxing city in southern China. Jpn J Infect Dis. 2014;67:417–22.
90. Hsu H, Tsai C, Chang M, Luh D. Constructing area-level indicators of successful ageing in Taiwan. Health Soc Care Community. 2010;18:70–81.
91. Lin R, Chuang C. A hybrid diagnosis model for determining the types of the liver disease. Comput Biol Med. 2010;40:665–70.
92. Ajami S, Ketabi S. Performance evaluation of medical records departments by analytical hierarchy process (AHP) approach in the selected hospitals in Isfahan: Medical records dep. & AHP. J Med Syst. 2012;36:1165–71.
93. Hilgerink MP, Hummel MJ, Manohar S, Vaartjes SR, Ijzerman MJ. Assessment of the added value of the Twente Photoacoustic Mammoscope in breast cancer diagnosis. Med Devices (Auckl). 2011;4:107–15.
94. Hummel JM, Boomkamp ISM, Steuten LMG, Verkerke BGJ, Ijzerman MJ. Predicting the health economic performance of new non-fusion surgery in adolescent idiopathic scoliosis. J Orthop Res. 2012;30:1453–8.
95. Ijzerman MJ, van Til JA, Bridges JFP. A comparison of analytic hierarchy process and conjoint analysis methods in assessing treatment alternatives for stroke rehabilitation. Patient. 2012;5:45–56.
96. Perseghin P, Marchetti M, Pierelli L, Olivieri A, Introna M, Lombardini L, et al. A policy for the disposal of autologous hematopoietic progenitor cells: report from an Italian consensus panel. Transfusion. 2014;54:2353–60.
97. Uzoka FE, Obot O, Barker K, Osuji J. An experimental comparison of fuzzy logic and analytic hierarchy process for medical decision support systems. Comput Methods Programs Biomed. 2011;103:10–27.
98. Kuruoglu E, Guldal D, Mevsim V, Gunvar T. Which family physician should I choose? The analytic hierarchy process approach for ranking of criteria in the selection of a family physician. BMC Med Inform Decis Mak. 2015;15:63.
99. Riepe MW. Clinical preference for factors in treatment of geriatric depression. Neuropsychiatr Dis Treat. 2015;11:25–31.
100. Krishnamoorthy K, Mahalingam M. Selection of a suitable method for the preparation of polymeric nanoparticles: multi-criteria decision making approach. Adv Pharm Bull. 2015;5:57–67.
101. Kunasekaran V, Krishnamoorthy K. Multi criteria decision making to select the best method for the preparation of solid lipid nanoparticles of rasagiline mesylate using analytic hierarchy process. J Adv Pharm Technol Res. 2014;5:115–21.
102. Velmurugan R, Selvamuthukumar S, Manavalan R. Multi criteria decision making to select the suitable method for the preparation of nanoparticles using an analytical hierarchy process. Pharmazie. 2011;66:836–42.
103. Wollmann D, Steiner MT, Vieira GE, Steiner PA. Evaluation of health service providers by consumers through the Analytic Hierarchy Process Method. Rev Saude Publica. 2012;46:777–83.
104. Fang L, Tung H. Comparison of nurse practitioner job core competency expectations of nurse managers, nurse practitioners, and physicians in Taiwan. J Am Acad Nurse Pract. 2010;22:409–16.
105. Maruthur NM, Joy S, Dolan J, Segal JB, Shihab HM, Singh S. Systematic assessment of benefits and risks: study protocol for a multi-criteria decision analysis using the Analytic Hierarchy Process for comparative effectiveness research. F1000Res. 2013;2:160.
106. Zhang S, Wei Z, Liu W, Yao L, Suo W, Xing J, et al. Indicators for Environment Health Risk Assessment in the Jiangsu Province of China. Int J Environ Res Public Health. 2015;12:11012–24.
107. Diaz-Ledezma C, Lichstein PM, Dolan JG, Parvizi J. Diagnosis of periprosthetic joint infection in medicare patients: Multicriteria decision analysis. Clin Orthop Relat Res. 2014;472(11):3275–84.
108. Petit J, Meurice N, Kaiser C, Maggiora G. Softening the rule of five. Where to draw the line? Bioorg Med Chem. 2012;20:5343–51.
109. Munoz DA, Nembhard HB, Kraschnewski JL. Quantifying complexity in translational research: an integrated approach. Int J Health Care Qual Assur. 2014;27:760–76.
110. Guariguata L, Whiting D, Weil C, Unwin N. The International Diabetes Federation diabetes atlas methodology for estimating global and national prevalence of diabetes in adults. Diabetes Res Clin Pract. 2011;94:322–32.
111. Maruthur NM, Joy SM, Dolan JG, Shihab HM, Singh S. Use of the analytic hierarchy process for medication decision-making in type 2 diabetes. PLoS One. 2015;10:e0126625.
112. Dyer RF, Forman EH. Group decision support with the Analytic Hierarchy Process. Decis Support Syst. 1992;8:99–124.
113. Saaty TL, Shang JS. Group decision-making: Head-count versus intensity of preference. Socio Econ Plan Sci. 2007;41:22–37.
114. Broekhuizen H, Groothuis-Oudshoorn CGM, van Til JA, Hummel JM, Ijzerman MJ. A review and classification of approaches for dealing with uncertainty in multi-criteria decision analysis for healthcare decisions. Pharmacoeconomics. 2015;33:445–55.
115. Saaty TL. How to make a decision: The analytic hierarchy process. Eur J Oper Res. 1990;48:9–26.
116. Maleki H, Zahir S. A comprehensive literature review of the rank reversal phenomenon in the analytic hierarchy process. J Multi-Crit Decis Anal. 2013;20:141–55.
117. Curran SS, Tkach VV, Overstreet RM. A new species of Homalometron (Digenea: Apocreadiidae) from fishes in the northern Gulf of Mexico. J Parasitol. 2013;99:93–101.

Modul 7

Measuring patients' priorities using the Analytic Hierarchy Process in comparison with Best-Worst-Scaling and rating cards: methodological aspects and ranking tasks

Schmidt, Katharina

Babac, Ana

Pauer, Frédéric

Damm, Kathrin

von der Schulenburg, J.-Matthias

Health Economics Review 2016;

6 (1): 50

Schmidt et al. Health Economics Review (2016) 6:50
DOI 10.1186/s13561-016-0130-6

Health Economics Review

RESEARCH Open Access

Measuring patients' priorities using the Analytic Hierarchy Process in comparison with Best-Worst-Scaling and rating cards: methodological aspects and ranking tasks

Katharina Schmidt[1*], Ana Babac[1], Frédéric Pauer[1], Kathrin Damm[1] and J-Matthias von der Schulenburg[1,2]

Abstract

Background: Identifying patient priorities and preference measurements have gained importance as patients claim a more active role in health care decision making. Due to the variety of existing methods, it is challenging to define an appropriate method for each decision problem. This study demonstrates the impact of the non-standardized Analytic Hierarchy Process (AHP) method on priorities, and compares it with Best-Worst-Scaling (BWS) and ranking card methods.

Methods: We investigated AHP results for different Consistency Ratio (CR) thresholds, aggregation methods, and sensitivity analyses. We also compared criteria rankings of AHP with BWS and ranking cards results by Kendall's tau b.

Results: The sample for our decision analysis consisted of 39 patients with rare diseases and mean age of 53. 82 years. The mean weights of the two groups of CR ≤ 0.1 and CR ≤ 0.2 did not differ significantly. For the aggregation by individual priority (AIP) method, the CR was higher than for aggregation by individual judgment (AIJ). In contrast, the weights of AIJ were similar compared to AIP, but some criteria's rankings differed. Weights aggregated by geometric mean, median, and mean showed deviating results and rank reversals. Sensitivity analyses showed instable rankings. Moderate to high correlations between the rankings resulting from AHP and BWS.

Limitations: Limitations were the small sample size and the heterogeneity of the patients with different rare diseases.

Conclusion: In the AHP method, the number of included patients is associated with the threshold of the CR and choice of the aggregation method, whereas both directions of influence could be demonstrated. Therefore, it is important to implement standards for the AHP method. The choice of method should depend on the trade-off between the burden for participants and possibilities for analyses.

Keywords: Decision making, Analytic Hierarchy Process, Best-worst-scaling, Method comparison, Patient preferences

* Correspondence: ks@cherh.de
[1]Center for Health Economics Research Hannover (CHERH), Leibniz University of Hannover, Otto-Brenner-Str. 1, D-30159 Hanover, Germany
Full list of author information is available at the end of the article

Background

Measurement of patient preferences and priorities has gained more relevance in health care. One reason is the increasing importance of patient participation in health care. In Germany, the Robert Koch-Institute used to call the patients "costumers" and "evaluators" in their Information System of the Federal Health Monitoring [1]. Patients also want to decide scope of service of statutory health insurances' and which services are covered. Several studies found differences between patients' and physicians' perceptions of preferences (e.g., [2–5]). It is relevant to assess the preferences of the (potential) patients instead of proxy reports. Another reason for the increasing importance is the integration of preferences as utility in health economics evaluations and reimbursement decisions for pharmaceuticals. Knowledge of patients' preferences or priorities could be a chance for optimizing the health care system according to patients' requirements.

Decisions regarding treatment preferences must consider a variety of characteristics, so called multi-criteria decision problems. Possible options for solving decision problems are value-based methods, strategy based methods, and Conjoint Analyses (CA). The German Institute for Quality and Efficiency in Health Care (IQWiG) tested and confirmed the Analytic Hierarchy Process (AHP) method as decision making tool in health technology assessments [6]. Application of AHP for the measurement of preferences has increased during the last five years, but is still a less researched approach in health care decision making [7]. It remains unclear whether the AHP method and established decision making methods yield comparable results. Recent studies already examined the direct comparisons of AHP and CA, as seen in [8–11]. Other studies conducted comparisons between CA and Best-Worst Scaling (BWS) [12–16]. Mühlbacher and Kaczynski (2016) demonstrated the similarity of BWS results and ratings, but did not compare directly the results from AHP with BWS [17]. Although another study published by Mühlbacher et al. showed similar results for BWS and AHP methods, some of the subgroups differed in their rankings obtained by BWS and AHP method [18]. However, we found no further evidence about the similarity or differences in priorities raised by AHP, BWS, or ranking cards.

This study accompanied a research project designed to gather patient needs concerning the establishment of a central information portal about rare diseases (Zentrales Informationsportal über seltene Erkrankungen, ZIPSE). Since the available space on the website was limited, the most important information categories for patients occupy the most space followed by the less important information categories. Various information requirements on diagnosis, therapy, self-help, research, and specialized care facilities for people living with rare diseases, their relatives, and health care professionals were identified in qualitative interviews (see [19]). However, the ranking of the information criteria remained unclear. AHP was a suitable method for prioritizing these information categories in the next step (see [20]). Since AHP is a relatively new approach in health care and it is rarely been used in health care research compared to BWS and DCE, several methodological aspects remain unstandardized. Forman et al. (1998) described different aggregation methods for group decisions with the AHP method: aggregating individual judgments (AIJ) and aggregating individual priorities (AIP) by arithmetic mean or geometric mean [21]. The choice of aggregation method depends on the circumstances and the aim of the study. We wanted to examine and compare the resulting differences in decisions of the aggregation methods in our study. This paper shows outcomes for the different Consistency Ratio (CR) thresholds, aggregations methods, and sensitivity analyses. Furthermore, the study tries to identify how to validate the AHP outcomes. Outcomes were compared with the results of questionnaires using the following well established methods: BWS Case 1, and ranking cards. The first aim of this study was to demonstrate the impact of the non-standardized AHP method on priorities. Does the aggregation method influence the resulting group priority rankings? The second aim was to compare the AHP outcomes with the outcomes achieved by BWS and ranking methods to validate the resulting priorities from patient perspective (convergence validity).

Methods

AHP method and application

The AHP method originates from the marketing sector, invented by Thomas Saaty in the late 1970s. Dolan et al. applied the method of AHP the first time in the health care sector several years later in 1989 [22, 23]. Nevertheless, the AHP remains a rarely used decision making method in health care research compared to BWS, ranking cards, and DCE. The following methodological explanations are in accordance with Saaty [24]. The AHP decomposes the decision problem at different levels of hierarchy. The first level describes the aim of the decision making. This is then explained in further detail at a lower level using sub-criteria. The last level contains possible alternatives with their characteristics. In the interview, the participant compares all criteria pairwise at each level (15 comparisons in total) using a scale ranging from 9 to 1 to 9. Thereafter, the judgments of the pairwise comparisons set up a matrix. This method presumes that the reciprocal request results in reciprocal weights of judgments; therefore, only the upper half of the matrix has to be queried. The matrices are used to calculate weights by the

Eigenvector Method. Additionally, the Consistency Ratio (CR) can be computed from the matrices to examine whether the participants' answers are random. Following Saaty, the CR has to be ≤ 0.1. Other authors suggested a CR ≤ 0.2, but the threshold value is not defined consistently [8, 25]. Higher CR values indicate exclusion of answers and questionnaires due to inconsistency.

First, we briefly report the results of information requirements of patients with rare diseases. Second, we compare the results of CR ≤ 0.1 and CR ≤ 0.2 for median, quartiles, and extreme values (as box-plots). Third, different aggregation methods (geometric mean, arithmetic mean, and median) are used and the differences in results noted. Saaty suggested to calculate group priorities by aggregating judgments or final outcomes by geometric mean to satisfy the reciprocal property of the AHP [26]. Reciprocal properties present the first axiom for the AHP, meaning that the strength of one criterion's dominance over a second criterion is inversely proportional to the second criterion's dominance over the first. This implies that if criterion A is five times more important than criterion B, criterion B is one-fifth the importance of criterion A (for all axioms see [27]). This relationship must be preserved after aggregation and can be achieved by the geometric mean method. The geometric mean is always smaller than the arithmetic mean, except for one observation is zero [28]. In this sub-section, we also examine differences in the results for aggregating individual judgments (AIJ) in contrast to aggregating individual priorities (AIP). Additionally, a sensitivity analysis estimates the stability of weights. As most AHPs combine specific criteria combinations into overall alternatives (e.g., criteria combinations to describe three different cars), the sensitivity analyses focus on the stability of these alternatives. Because no standard method for the AHP without combining the attributes to alternatives was implemented, we looked at the confidence intervals (CIs) for each global weight of the criteria, and identified the stability of the ranking positions for each criterion. Therefore, we determined the BC_a bootstrap 95%-CI because our sample was small and in this case bootstrap CI were more accurate and correct than the standard CI [29]. All our analyses were conducted with the R statistic software program and the package "pmr" [30].

Methodological background of the BWS and ranking cards

As a second method in this paper, we applied BWS Case 1 in the same study population population [31]. Here, different combinations of the criteria built up the sets. The interviewee selected the best and the worst criteria in each set, resulting in two decisions per set. Each person answered seven sets. The BWS method is based on random utility theory, and uses the choice models or the count analysis. Methods used in choice approaches are multinomial logit model, conditional logit model, maximum-likelihood, or weighted least square method population [31]. Since we were not interested in predictors for the decision, but rather in rankings, we emphasized the count analysis method and rankings.

Using ranking cards resulted in an ordinal ranking of criteria, implying that distances between criteria could not be measured. Besides, it was a well-established warm-up task [32], and could support the interviewee to remain consistent with their prior ranking throughout all tasks. This survey included the ranking cards method before the AHP tasks.

Comparison of results from AHP, BWS, and ranking cards

Furthermore, the results from AHP, BWS, and ranking cards were compared. We placed the results in a table and examined differences in the rank. The AHP's weights could not be compared with the weights from the BWS, because they are based on deviating mathematical calculation methods and scales. In addition, we conducted tests for correlation between the ranks with the help of Kendall's tau b coefficient. This coefficient was used for rank ordered data, and identifies concordant and discordant rankings between two or more variables [33]. The Kendall's tau b makes adjustments for ties in the data, in contrast to Kendall's tau a.

Survey design

The study sample consisted of randomly selected participants from the qualitative main study of the ZIPSE project [19]. A positive vote was obtained from the ethics committee of Albert-Ludwigs-University Freiburg (number 53/14). As it was an accompanying research project, inclusion and exclusion criteria for participants were equal to those of the main study sample. Therefore, participants were at least 18 years old and were either suffering from a rare disease, or were the near relative of a sick individual. In this study participants were interviewed either face-to-face, or via phone with a paper-pencil questionnaire that contained AHP, BWS, and ranking tasks. Criteria development is described in detail by Babac et al. [20]. Additionally, socio-demographic and disease specific data were collected. A ranking task of cards with the criteria's descriptions should support consistent answering. Therefore, participants arranged the cards according to their preferred order, and left them next to the questionnaire during the rest of the interview. The interviewer indicated inconsistencies between ranking cards. Hence, participants could adjust either the order of the cards, or the judgment in the questionnaire.

Financial support for this study was provided in part by a grant from the Federal Ministry of Health. The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

Results

Initially, we report the AHP results including the criteria description and their hierarchical arrangement. Then, we show the information criteria priorities evaluated by patients with rare diseases or their relatives. The following subsections investigate the outcomes of different methodological approaches in the AHP method. Finally, we report the comparison of AHP results with BWS and ranking tasks.

Figure 1 shows the final hierarchy for the AHP. It consists of four levels with the aim of study on the first level. The aim decomposes into information about *medical issues*, *research*, *current events,* and *social advisory and support services*. The topic of *medical issues* was again subdivided into *diagnosis*, *treatment,* and *disease patterns*. The first two were split into *provider* and *methods* at the fourth level. *Disease patterns* contained *aetiology*, *frequency*, *typical symptoms,* and *progression* at the lowest level. At the third level *research* implied *current studies*, *study results,* and *registries. Current events* at level two contained no further subcategories. The last category at level two was divided into *social law counseling*, *psychosocial counseling,* and *self-help* at level three. *Self-help* further held the subcategories of *personal contacts* and *online contacts* (fourth level). Further details and descriptions can be found in Additional file 1.

The sample for our decision analysis consisted of 31 women and 8 men with mean age of 53.82 years. The inequitable distribution of gender was due to the fact of unequal proportions in the qualitative main study.

In the first scenario, all participants who reached a CR at second level exceeding 0.1 were excluded from the analyses. Then 22 included participants (19 women, 3 men; mean age: 52.50 years) remained for further analytical steps. In this scenario, we calculated weights for each included participant and then aggregated the weights (AIP method). The first approach was aggregating the weights by median. In Fig. 2, the results are shown as boxplots including the quartiles and distribution of weights for each criterion at second level.

The boxplots show that *medical issues* were the most important criteria for the participants with a median weight of 0.4548 (SD = 0.1728), followed by *social support* (weight (w) = 0.1575, SD = 0.1777), and *research* (w = 0.1314, SD = 0.1462). The least criterion was information about *current events* with a median weight of 0.0913 (SD = 0.1550). The SDs of *social support, research,* and *current events* indicated high variations of the priorities in the sample.

Figure 3 shows the local weights of sub-criteria at the lower third level. The gray boxplots indicated the subcriteria of *medical issues* with the highest weight for *diagnosis* (median weight (mw) = 0.4517, SD = 0.2240), followed by *treatment* (mw = 0.3512, SD = 0.2223), and

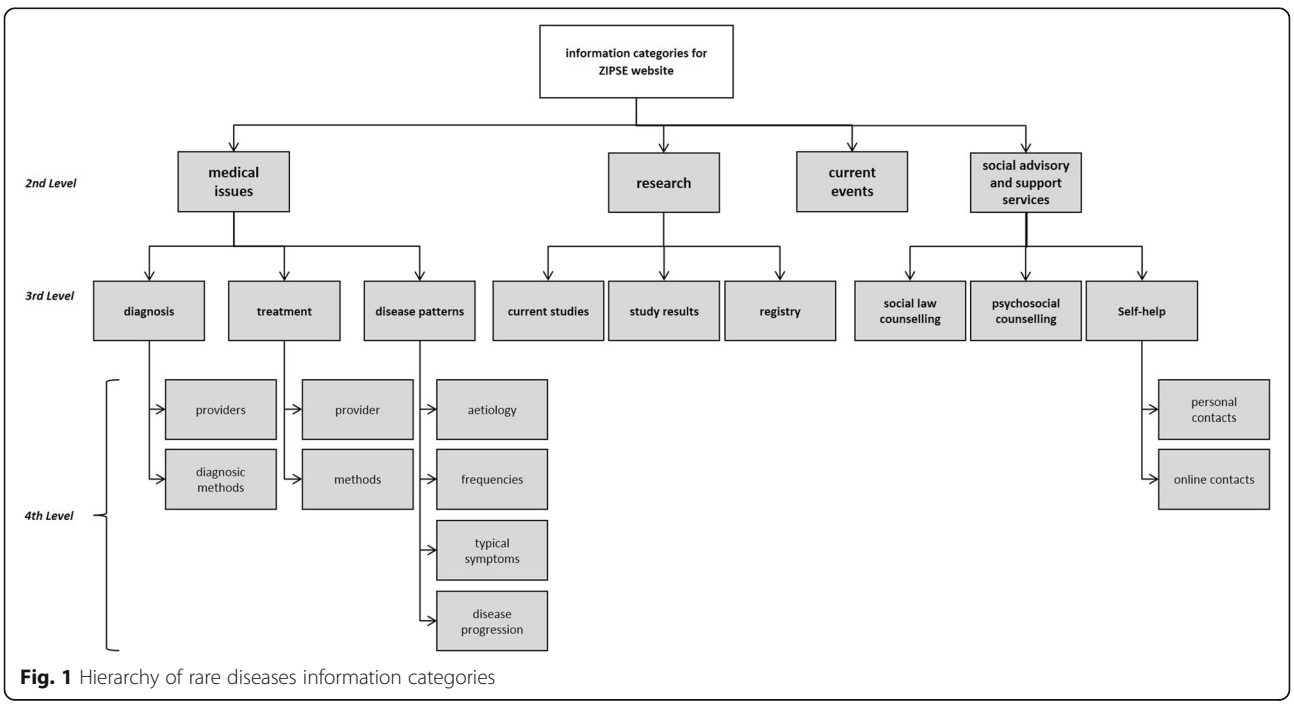


Fig. 1 Hierarchy of rare diseases information categories

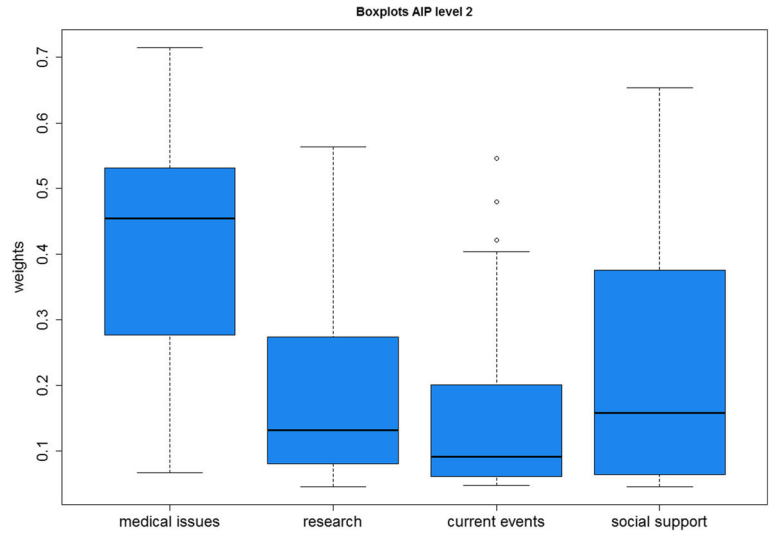


Fig. 2 Boxplots of global weights from criteria at second level

disease patterns (mw = 0.1492, SD = 0.0763). The second information criterion of *research* (blue boxplots) included *current studies*, *study results,* and *registry*. The most important sub-criterion was *study results* with a local weight of 0.4416 (SD = 0.2015), the second *current studies* (w = 0.3184, SD = 0.1955), and the third was the information about *registries* (w = 0.1429, SD = 0.2142). The green boxplots displayed the local weights for the category of *social support. Self-help* (w = 0.4663, SD = 0.2307) reached the highest weight followed by *psychosocial counseling* (w = 0.2845, SD = 0.1801), and *law counseling* with the lowest weight of 0.2167 (SD = 0.1768). We did not compare the global weights of subcriteria against each other because high weights at the second level (e.g., for *medical issues*) would highly influence the weights at the third level. Therefore, we used the sub-criteria's local weights for comparisons within each criterion because the global weights were not important for our methodological considerations.

Comparison of consistency thresholds

Figure 4 shows the boxplots for all global weights separated by level. Additionally, it compares the boxplots for a threshold of included participants with high consistency (CR ≤ 0.1) and a threshold of lesser consistency (CR ≤ 0.2). All graphs show an almost equal median for the two groups of CR and a *t*-test indicate no significant differences of median for each criterion (not shown here). However, a difference in the ranking by median occurs at level three: *law counseling* gained a higher weight for an extended threshold and received rank 9 (w = 0.0310) instead of the 13th and last rank (w = 0.0452). At the same

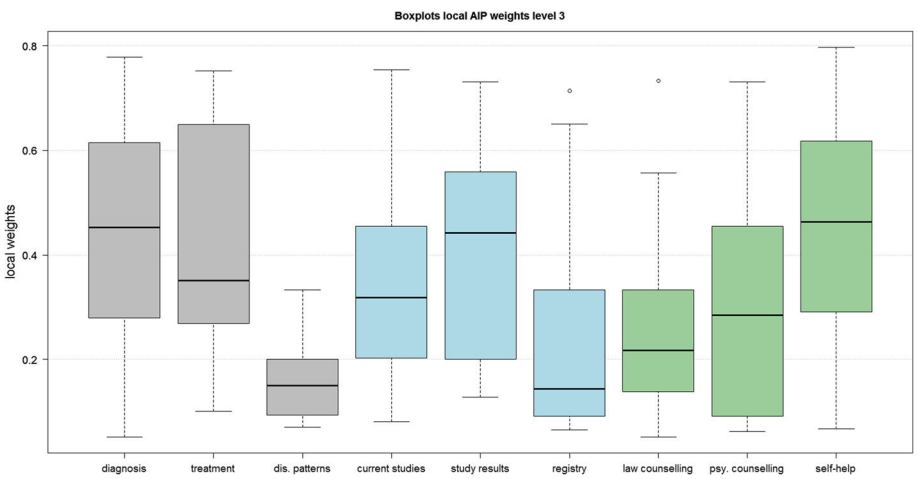


Fig. 3 Boxplots local AIP weights at third level

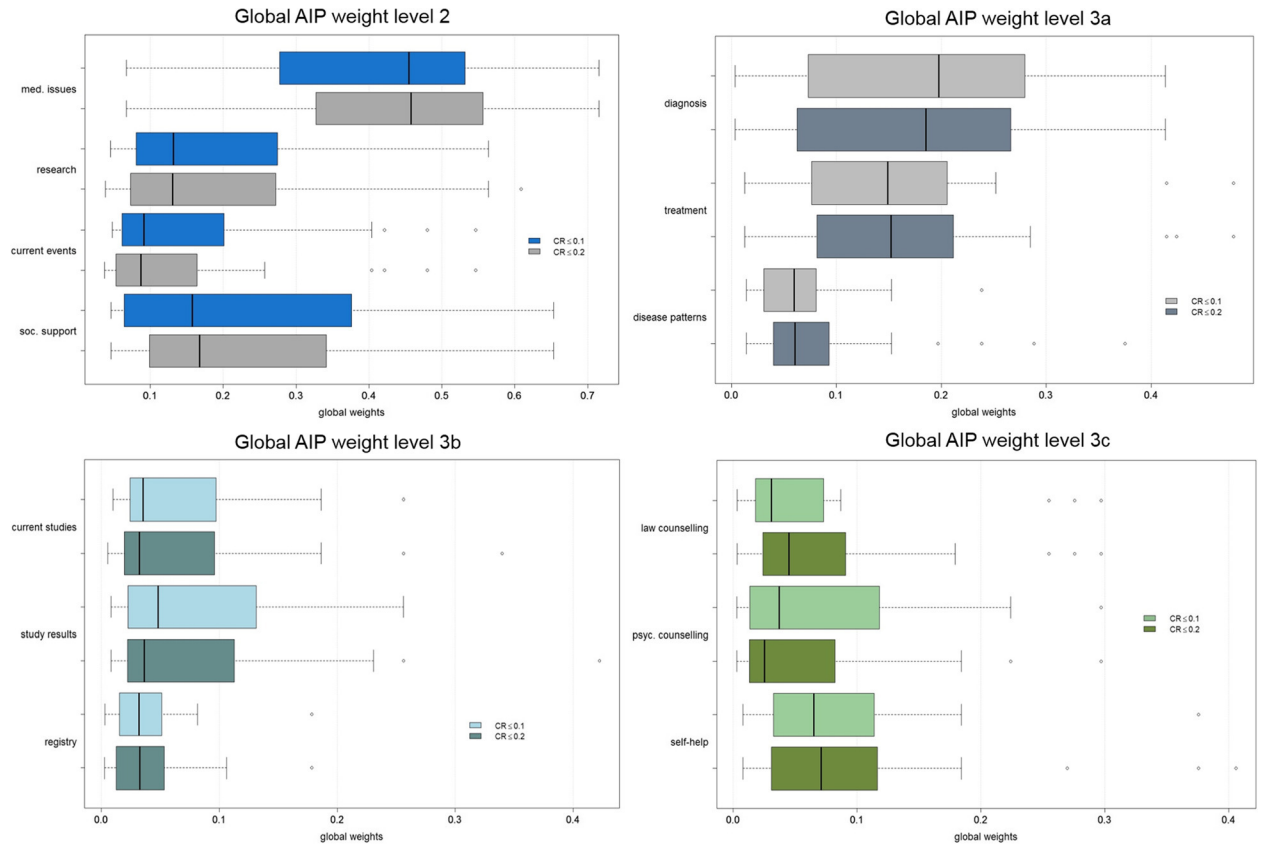


Fig. 4 Boxplots global AIP weights separated by CR

time, *psychosocial counseling* fell from rank 10 to 13 (weight 0.0372 onto 0.0254). A rank reversal occurs for *current studies* (weight 0.0353 onto 0.0324) and *registries* (weight 0.0319 onto 0.0325). In summary, the medians between a lower and a higher CR threshold did not differ significantly. Nevertheless, when small differences in weights occurred, rank reversals could be observed. In this study, rank reversals occurred only for the last four rankings.

Comparison of aggregation methods

In the next step, we analyzed differences in global weights by different aggregation methods. All mean calculations were based on geometric mean calculation as it serves the Pareto Principle and therefore seems to be the correct approach in theory [10, 34]. In the first scenario, the AIJ was applied. This method aggregated the comparison matrices first. In a second step, priority weights were calculated for each criterion. An overall CR was calculated for level two after the aggregation of all individual opinions. In the second scenario the AIP method was applied. This methodology calculated eigenvectors and priorities for each participant first. Only participants with a CR smaller than or equal to 0.1 were included in the aggregation. Afterwards, resulting priority weights were aggregated through geometric mean calculation.

Figure 5 displays the results of the two scenarios that comprised all 31 participants for scenario 1 and 22 for scenario 2. The aggregated judgments (scenario 1) show similar global weights for most of the criteria compared to the aggregated weights (scenario 2). Rank reversal occurrs between *diagnosis, treatment,* and *research,* because for scenario 1, *research* (w1 = 0.2038) and *treatment* (w1 = 0.1862) were more important than *diagnosis* (w1 = 0.1691), whereas in scenario 2, *research* (w2 = 0.1916) and *treatment* (w2 = 0.1892) were less important than *diagnosis* (w2 = 0.1955). Likewise, the ranking differs for *self-help, study results,* and *disease patterns*: in scenario 1, *disease patterns* (w1 = 0.0940) were more important than *self-help* (w1 = 0.0871) and *study results* (w1 = 0.0860), and in scenario 2, it was the other way round (*self-help* w2 = 0.0906, *study results* w2 = 0.0786, *disease patterns* w2 = 0.0785). A third rank reversal can be seen for the two scenarios between *current studies* (w1 = 0.0721, w2 = 0.0704, rank 11 vs. 10), *psychosocial counseling* (w1 = 0.0568, w2 = 0.0547, rank 12 vs. 11), and *law counseling* (w1 = 0.0729, w2 = 0.0531, rank 10

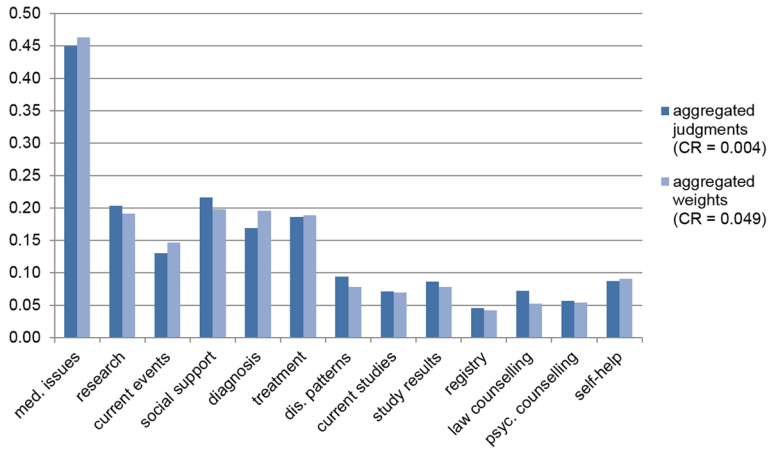


Fig. 5 Comparison of global weights for different aggregation levels

vs. 12). The CR for the second level was 0.004 in the first scenario, whereas the CR was 0.05 in the second scenario.

In the next step, the AIJ and AIP were compared by median. The table for these comparisons can be found in Additional file 2. The results are nearly identical to Fig. 5. The differences are small deviations in the weights and a few higher weights for the AIP than the AIJ (*current events*, *registries,* and *self-help*). The last comparison of AIP and AIJ was conducted by their means. Here, the AIP were markedly higher than most of the AIJ, also in comparison with the AIPs of the previously mentioned aggregation methods. Additionally, the weights summed up to 1 at first level, and they yielded the appropriate weights at lower levels. However, the most important question in this context was whether the ranking position changed through the different aggregation methods. Table 1 answers this question.

The noticeable difference occurs for the criterion *self-help*, which took the ranking positions from 7 to 13 over the different methods. Another striking criterion is *current studies*, which obtains ranking positions between 5 and 11. Two less intensive varying criteria were *social support* and *disease pattern* that differed between 5 positions. The further 9 criteria varied between 3 ranking positions, so a relatively stable valuation could be assumed.

Finally, the influence of aggregation method on CR had to be examined. The CR in the scenario of aggregation by geometric mean was markedly lower for AIJ than for AIP (CR AIJ: 0.0045; CR AIP: 0.0490), although only participants with a CR ≤ 0.1 were included for the AIP. By using the median (CR AIJ: 0.0683; CR AIP: 0.0674)

Table 1 Comparison of aggregation methods and weights

	Geometric mean ranking		Median ranking		Mean ranking	
	AIJ	AIP	AIJ	AIP	AIJ	AIP
Med. issues	1	1	1	1	1	1
Research	3	3	5	5	3	3
Current events	6	6	9	6	6	5
Social support	**2**	**2**	4	3	**7**	**2**
Diagnosis	5	4	2	2	2	4
Treatment	4	5	3	4	4	6
Disease patterns	7	8	**6**	8	9	**11**
Current studies	**11**	10	7	**11**	**5**	10
Study results	9	9	8	9	8	8
Registry	13	13	13	12	11	13
Law counseling	10	12	10	13	10	12
Psychosocial counseling	12	11	11	10	12	9
Self-help	8	**7**	12	**7**	**13**	7

The bold data highlights the results in the following text passage

or mean scenario (CR AIJ: 0.0745; CR AIP: 0.0587), the CRs were similar, but still much higher than the CR from AIJ by geometric mean, as expected.

Sensitivity analysis of AHP results

Usually AHP examine a combination of (sub-)criteria weights resulting in decision alternatives. Thereafter, the sensitivity of alternatives can be analyzed. However, the underlying study does not integrate a hierarchy level with decision outcomes, but only criteria and sub-criteria. Therefore, we looked at the stability of the criteria's ranking positions. Consequently, we calculated the CIs for each global weight (see Fig. 6). In addition, we show the mean weight of the underlying sample. The CIs distributed over three ranges for global weights. The seven lowest criteria in the figure from *self-help* to *results* showed CIs from approximately 0.03 to 0.14, and the CIs were rather small, particularly *social support*. Then, the criteria of *current studies*, *research*, *disease patterns*, *therapy,* and *diagnosis* covered a CI from approximately 0.11 to 0.30. A markedly higher CI arose for *medical issues* (CI: 0.34–0.49). It could be concluded that within the first two groups, the criteria were likely vulnerable to rank reversal. In contrast, the first rank for *medical issues* was assumed to be robust.

Comparison of methods

In the next section, we wanted to contrast the results of the AHP and the BWS. Table 2 compares the results of the methods. The most important criterion at level two was information about *medical issues* in all three methods, followed by *social support* and *research*. The least important criterion, *current events*, was also equal for AHP and BWS, but for the ranking cards it was also ranked position 3. At level three for *medical issues*, the most important criterion was *treatment* in the BWS, and diagnosis in the AHP. *Disease patterns* took the third position in both cases. The sub-criteria for *research* were ranked as followed for BWS and also AHP: 1) *study results*, 2) *current studies*, 3) *registry*. In the category of *social support,* the most important sub-criterion was *self-help*. The positions 2 and 3 differed between BWS and AHP. In the BWS, the second important sub-criterion was *law counseling,* whereas it was *psychosocial counseling* in the AHP. The ranking cards results showed doubled ranking positions at all levels, particularly when BWS and AHP were indifferent.

Because the ranking cards gave orientation for the AHP in the interviews, we assumed that there was a correlation between their results. Therefore, we did not evaluate the correlations for AHP and ranking. We examined the correlation between AHP and BWS rankings by Kendall's tau coefficient, for each hierarchical level. We found significant moderate to strong correlation between the two methods in the rankings (see Table 3).

Discussion

In this paper, we focused on methodological aspects of AHP and comparison of methods. The first step was to compare the results for different CR thresholds. Thereby, we considered the weights for including all interviewees with CR ≤ 0.1 or CR ≤ 0.2. We found that the mean weights between these two groups did not differ significantly. However, rank reversal could occur if the criteria's weights are close. For clarification, another phenomenon in AHP is also called "rank reversal": it occurs when adding or deleting an alternative leads to a shift in the previous alternatives' ranking order [35, 36]. The latter phenomenon was not investigated in our study.

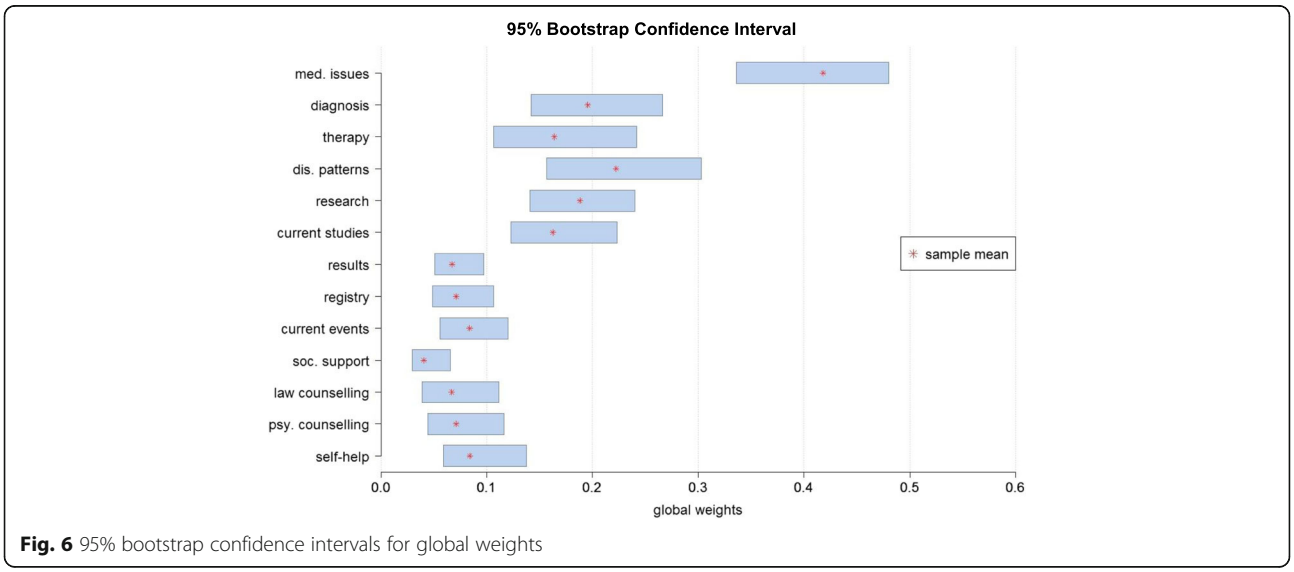


Fig. 6 95% bootstrap confidence intervals for global weights

Table 2 Comparison of BWS, AHP, and ranking cards

Criteria	BWS values	AHP local weights	BWS ranking	AHP ranking	Ranking cards[a]
Med. issues	1.000	0.368	1	1	1
Research	0.322	0.152	3	3	3
Current events	0.000	0.117	4	4	3
Social support	0.372	0.158	2	2	2
Diagnosis	0.855	0.354	2	1	1
Treatment	1.000	0.342	1	2	1
Dis. patterns	0.000	0.142	3	3	2
Current studies	0.279	0.304	2	2	2
Study results	1.000	0.339	1	1	1
Registry	0.000	0.184	3	3	2
Law counseling	0.421	0.213	2	3	2
Psyc. counseling	0.000	0.220	3	2	2
Self-help	1.000	0.363	1	1	1

[a]Equal ranking for multiple criteria permitted

The second step was to compare different aggregation methods. Therefore, we calculated the geometric means of the AIJ method (scenario 1) as well as the AIP method (scenario 2). The first difference was the number of participants that were included with a CR ≤ 0.1. In the first scenario, we included 31 participants, and in the second scenario, we had to exclude 9 participants because they showed CRs > 0.1. In the first scenario, we had a CR of 0.004 for the second level calculated after aggregating the judgments. In the second scenario, the CR at the second level was 0.05, and thus higher than in scenario 1, although the participants with CRs > 0.1 were excluded from the final CR calculation. The results received from scenario 1 showed almost the same weights compared to the results from scenario 2. Besides, the criteria's rankings differed between the scenarios, due to short distances between the weights. The AIJ method implies that the group decides as a new individual whereas the AIP method is based on the assumption that each individual decides on her or his own and the resulting decisions are aggregated [21]. Therefore, the aggregating method should depend on whether the sample is seen as one unit or a group of individuals. Forman et al. (1998) argued that for AIJ the geometric mean must be used because otherwise two social choice theory axioms (Pareto optimality and homogeneity) are not satisfied [21, 37]. The Pareto optimality axiom describes that the most frequently preferred alternative in the individual decisions must be the preferred one in the group decision. The homogeneity axiom states that the ratio between the criteria weights is the same for individual and aggregated group judgments. Our study supported Forman's demand as we saw violations of the Pareto axiom in Table 1, but not for the most preferred criterion. The homogeneity axiom was not investigated in our study. In future AHP studies, following Forman et al. (1998) and Saaty (2008) the geometric mean should be used in AIJ method.

Table 3 Correlation between AHP ranking and BWS ranking for each level

	Kendalls tau	*p*-value
Level two	0.585	<0.001
Level three a	0.543	<0.001
Level three b	0.613	<0.001
Level three c	0.668	<0.001

In the third step, we opposed the criteria's rankings received from aggregated weights and judgments by geometric mean, median, and mean. Here, the ranking positions showed deviating results and rank reversals. These aspects should be considered when results derived by different aggregation methods in studies are compared.

As no sensitivity analysis is suggested for AHPs that do not include alternatives, we tried to find an appropriate one. The aim of sensitivity analysis in AHP is to find instable criteria that could cause rank reversal. Therefore, we illustrated the 95%-CIs for all criteria. Where CIs overlap because of similar weights, the risk for rank reversal increased.

Finally, we evaluated the criteria's rankings for the different methods (AHP, BWS, ranking cards). However, we could not compare the weights from AHP with the weights from the BWS, because they use different scales. Therefore, only the rankings could be compared between the methods. Here, we found moderate to strong correlations between the AHP and BWS.

Correlated results between the methods were similarly reported by prior studies. Pignone et al. (2012) investigated differences in value elicitations with CA, rating,

and ranking tasks [38]. They concluded that the CA produced different values compared with ranking and rating, but the latter two led to similar results. Van Til et al. analyzed the differences between pairwise comparisons, BWS, five point rating scales, point allocation and ranking [39]. There were no differences between the methods at group level; however, differences occurred at the individual level and the largest differences were between pairwise comparisons and the five point rating scale. The correlation between the methods for individual weights was moderate. Furthermore, the order of the methods shown in the questionnaire influenced the weights. We did not examine this aspect in our study, because we had a small sample, and could not expect significant results regarding this question. Therefore, the order of tasks could also influence the results.

A major problem was the inconsistent response behavior of the participants in the AHP. Our sample consisted of patients with different rare diseases. The diverse clinical pictures and disease stages could have led to different priorities in the evaluation of the information criteria. Although in our study the participants used ranking cards for assistance during the AHP, the CRs were not all below the defined threshold. This phenomenon raised the question, whether the AHP method was not applicable in certain participant groups or in a heterogeneous sample. Therefore, future research projects should investigate the requirements for their participants, because this could bias the results. Further studies should also examine whether the aggregation of judgments always leads to higher values than the aggregation of weights, as detected in our study.

Another aspect was the small number of participants. Although we neglected this aspect in our study, the number of participants could also be an influencing factor of the results. Recent literature suggests that AHP is particularly useful for small groups, because priorities can be calculated for each participant [40]. As we used the sample from the main study, a larger proportion of women was included. Nevertheless, by aggregating the individual judgments or weights the researcher gave a statement for a (heterogeneous) group. Thus, we should present the results from the AHP under the restriction of their study population. The results were representative for this study population only.

Conclusion

In the AHP method, the number of patients is influenced by the CR aggregation method and the threshold of the CR, which could bias the results. Therefore, it is important to establish guidelines and investigate the differences for each study as also mentioned by Schmidt (2015) [7]. The comparison between the different methods (AHP, BWS, ranking tasks) resulted in similar outcomes.

The AHP seemed to be a challenge for some participants. Reasons could be the unusual scale and the need for consistency over several questions. However, we could not identify special groups because our sample was too small and homogenous. The BWS also forced the participants to make decisions. However, here only the best and worst decision had to be made. Therefore, the cognitive burden is reduced compared to other methods, for example, the DCE [41]. The researcher should consider the trade-off between methods that are easy to understand, and the method's gain of information as well as the method's theoretical basis. In addition, the sensitivity of each method should be calculated for each research question. In sum, the choice of method depends on the trade-off between the burden for participants and possibilities for analyses. Consequently, the method should be chosen according to the characteristics of the study sample and the aim of the study.

Additional files

Additional file 1: Description of the AHP criteria. (DOCX 15 kb)

Additional file 2: Aggregation level and different means. (DOCX 13 kb)

Abbreviations

AHP: Analytic Hierarchy Process; AIJ: Aggregation by individual judgment; AIP: Aggregation by individual priority; BWS: Best-worst-scaling; CA: Conjoint analyses; CI: Confidence interval; CR: Consistency ratio; DCE: Discrete choice experiment; IQWiG: Institute for Quality and Efficiency in Health Care; MW: Median weight; SD: Standard deviation; W: Weight; ZIPSE: Zentrales Informationsportal über seltene Erkrankungen (English: central information portal about rare diseases)

Acknowledgements

The publication of this article was funded by the Open Access fund of Leibniz Universität Hannover.

Authors' contributions

KS carried out the analyses and drafted the manuscript. AB recruited and interviewed the participants and participated in finalizing the manuscript. FP, KD, and JMS supported the study conduct and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]Center for Health Economics Research Hannover (CHERH), Leibniz University of Hannover, Otto-Brenner-Str. 1, D-30159 Hanover, Germany. [2]Biomedical Research in Endstage and Obstructive Lung Disease Hannover (BREATH), Member of the German Center for Lung Research (DZL), Hanover, Germany.

Received: 13 June 2016 Accepted: 25 October 2016
Published online: 14 November 2016

References

1. Robert Koch-Institut (ed.). Bürger- und Patientenorientierung im Gesundheitswesen [Internet]. Berlin. 2006. Available from: https://www.rki.de/DE/Content/Gesundheitsmonitoring/Gesundheitsberichterstattung/GBEDownloadsT/buergerorientierung.pdf?__blob=publicationFile. Accessed 25 Apr 2016.

2. Caocci G, Voso MT, Angelucci E, et al. Accuracy of physician assessment of treatment preferences and health status in elderly patients with higher-risk myelodysplastic syndromes. Leuk Res. 2015;39(8):859–65.
3. Hageman MG, Kinaci A, Ju K, et al. Carpal Tunnel Syndrome: Assessment of Surgeon and Patient Preferences and Priorities for Decision-Making. J Hand Surg. 2014;39(9):1799–1804.e1.
4. Bruera E, Sweeney C, Calder K, et al. Patient preferences versus physician perceptions of treatment decisions in cancer care. J Clin Oncol Off J Am Soc Clin Oncol. 2001;19(11):2883–5.
5. Hummel MJM, Volz F, van Manen JG, et al. Using the analytic hierarchy process to elicit patient preferences: prioritizing multiple outcome measures of antidepressant drug treatment. Patient. 2012;5(4):225–37.
6. Danner M, Hummel JM, Volz F, et al. 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.
7. Schmidt K, Aumann I, Hollander I, et al. Applying the Analytic Hierarchy Process in healthcare research: A systematic literature review and evaluation of reporting. BMC Med Inform Decis Mak. 2015;15(1):497.
8. Ijzerman MJ, van Til JA, Snoek GJ. Comparison of two multi-criteria decision techniques for eliciting treatment preferences in people with neurological disorders. Patient. 2008;1(4):265–72.
9. Benaïm C, Perennou D-A, Pelissier J-Y, Daures J-P. Using an analytical hierarchy process (AHP) for weighting items of a measurement scale: a pilot study. Rev Epidemiol Sante Publique. 2010;58(1):59–63.
10. Neidhardt K, Wasmuth T, Schmid A. Die Gewichtung multipler patientenrelevanter Endpunkte, Ein methodischer Vergleich von Conjoint Analyse und Analytic Hierarchy Process unter Berücksichtigung des Effizienzgrenzenkonzepts des IQWiG [Internet]. 2012. Available from: http://www.fiwi.uni-bayreuth.de/de/download/WP_02-12.pdf. Accessed Aug 4 2015.
11. Ijzerman MJ, van Til JA, Bridges JFP. A comparison of analytic hierarchy process and conjoint analysis methods in assessing treatment alternatives for stroke rehabilitation. Patient. 2012;5(1):45–56.
12. Whitty JA, Walker R, Golenko X, et al. A Think Aloud Study Comparing the Validity and Acceptability of Discrete Choice and Best Worst Scaling Methods. PLoS ONE. 2014;9(4):e90635.
13. Whitty JA, Ratcliffe J, Chen G, Scuffham PA. Australian Public Preferences for the Funding of New Health Technologies: A Comparison of Discrete Choice and Profile Case Best-Worst Scaling Methods. Med Decis Mak. 2014;34(5):638–54.
14. Xie F, Pullenayegum E, Gaebel K, et al. Eliciting preferences to the EQ-5D-5 L health states: discrete choice experiment or multiprofile case of best–worst scaling? Eur J Health Econ. 2014;15(3):281–8.
15. Potoglou D, Burge P, Flynn T, et al. Best–worst scaling vs. discrete choice experiments: An empirical comparison using social care data. Soc Sci Med. 2011;72(10):1717–27.
16. Severin F, Schmidtke J, Mühlbacher A, Rogowski WH. Eliciting preferences for priority setting in genetic testing: a pilot study comparing best-worst scaling and discrete-choice experiments. Eur J Hum Genet. 2013;21(11):1202–8.
17. Mühlbacher AC, Kaczynski A. The Expert Perspective in Treatment of Functional Gastrointestinal Conditions: A Multi-Criteria Decision Analysis Using AHP and BWS. J Multi-Criteria Decis Anal. 2016;23(3-4):112-125.
18. Mühlbacher AC, Bethge S, Kaczynski A, Juhnke C. Objective Criteria in the Medicinal Therapy for Type II Diabetes: An Analysis of the Patients' Perspective with Analytic Hierarchy Process and Best-Worst Scaling [Zielkriterien der medikamentosen Therapie des Diabetes Typ II: Eine Analyse der Patientenperspektive mit Analytic Hierarchy Process und Best-Worst Scaling]. Gesundheitswesen. 2016;78(5):326–36.
19. Litzkendorf S, Babac A, Rosenfeldt D, Schauer F, Hartz T, et al. Information Needs of People with Rare Diseases - What Information Do Patients and their Relatives Require? J Rare Dis Diagn Ther. 2016;2(2):40.
20. Dolan JG. Multi-criteria clinical decision support: A primer on the use of multiple criteria decision making methods to promote evidence-based, patient-centered healthcare. The Patient. 2010;3(4):229–48.
21. Forman E, Peniwati K. Aggregating individual judgments and priorities with the analytic hierarchy process. Eur J Oper Res. 1998;108(1):165–9.
22. Dolan JG, Isselhardt BJ, Cappuccio JD. The analytic hierarchy process in medical decision making: a tutorial. Med Decis Making. 1989;9(1):40–50.
23. Dolan JG. Medical decision making using the analytic hierarchy process: choice of initial antimicrobial therapy for acute pyelonephritis. Med Decis Making. 1989;9(1):51–6.
24. Saaty TL. Modeling unstructured decision problems. The theory of analytical hierarchies. Math Comput Simul. 1978;20(3):147–58.
25. Dolan JG. Shared decision-making – transferring research into practice: The Analytic Hierarchy Process (AHP). Patient Educ Couns. 2008;73(3):418–25.
26. Saaty TL. Decision making with the analytic hierarchy process. Int J Services Sciences. 2008;1(1):83–98.
27. Saaty RW. The analytic hierarchy process—what it is and how it is used. Math Model. 1987;9(3–5):161–76.
28. Colquhoun D. Lectures on biostatistics, An introduction to statistics with applications in biology and medicine. Oxford: Clarendon Press; op. 1971. XVIII, 425 str. ISBN: 978–0198541196.
29. DiCiccio TJ, Efron B. Bootstrap Confidence Intervals. Stat Sci [Internet]. 1996;11(3):189–212. Available from: http://www.jstor.org/stable/2246110.
30. Lee PH, Yu, Philip L. H. Probability Models for Ranking Data [Internet]. 2014. Available from: http://cran.r-project.org/web/packages/pmr/index.html. Accessed 4 Feb 2015.
31. Flynn TN, Louviere JJ, Peters TJ, Coast J. Estimating preferences for a dermatology consultation using Best-Worst Scaling: Comparison of various methods of analysis. BMC Med Res Methodol. 2008;8(1):76.
32. Ali S, Ronaldson S. Ordinal preference elicitation methods in health economics and health services research: using discrete choice experiments and ranking methods. Br Med Bull. 2012;103(1):21–44.
33. Benninghaus H. Deskriptive Statistik, Eine Einführung für Sozialwissenschaftler. 11th ed. Wiesbaden: VS, Verl. für Sozialwiss; 2007. p. 285S. Studienskripten zur Soziologie. ISBN 9783531146072.
34. Ishizaka A, Labib A. Review of the main developments in the analytic hierarchy process. Expert Systems with Applications. 2011;38(11):14336-45.
35. Wang Y-M, Luo Y. On rank reversal in decision analysis. Math Comput Model. 2009;49(5–6):1221–9.
36. Belton V, Gear T. On a short-coming of Saaty's method of analytic hierarchies. Omega. 1983;11(3):228–30.
37. Ramanathan R, Ganesh LS. Group preference aggregation methods employed in AHP, An evaluation and an intrinsic process for deriving members' weightages. Eur J Oper Res. 1994;79(2):249–65.
38. Pignone MP, Brenner AT, Hawley S, et al. Conjoint Analysis Versus Rating and Ranking for Values Elicitation and Clarification in Colorectal Cancer Screening. J Gen Intern Med. 2012;27(1):45–50.
39. van Til J, Groothuis-Oudshoorn C, Lieferink M, et al. Does technique matter; a pilot study exploring weighting techniques for a multi-criteria decision support framework. Cost Eff Resour Alloc. 2014;12(1):22.
40. Mühlbacher AC, Kaczynski A. Der Analytic Hierarchy Process (AHP): Eine Methode zur Entscheidungsunterstützung im Gesundheitswesen. PharmacoEcon Ger Res Artic. 2013;11(2):119–32.
41. Flynn TN, Louviere JJ, Peters TJ, Coast J. Best–worst scaling: What it can do for health care research and how to do it. J Health Econ. 2007;26(1):171–89.

Modul 8

Comparison of different approaches applied in Analytic Hierarchy Process - an example of information needs of patients with rare diseases

Pauer, Frédéric

Schmidt, Katharina

Babac, Ana

Damm, Kathrin

Frank, Martin

von der Schulenburg, J.-Matthias

BMC medical informatics and decision making 2016;

16 (1): 117

Pauer et al. BMC Medical Informatics and Decision Making (2016) 16:117
DOI 10.1186/s12911-016-0346-8

BMC Medical Informatics and Decision Making

RESEARCH ARTICLE Open Access

Comparison of different approaches applied in Analytic Hierarchy Process – an example of information needs of patients with rare diseases

Frédéric Pauer*, Katharina Schmidt, Ana Babac, Kathrin Damm, Martin Frank and J.-Matthias Graf von der Schulenburg

Abstract

Background: The Analytic Hierarchy Process (AHP) is increasingly used to measure patient priorities. Studies have shown that there are several different approaches to data acquisition and data aggregation. The aim of this study was to measure the information needs of patients having a rare disease and to analyze the effects of these different AHP approaches. The ranking of information needs is then used to display information categories on a web-based information portal about rare diseases according to the patient's priorities.

Methods: The information needs of patients suffering from rare diseases were identified by an Internet research study and a preliminary qualitative study. Hence, we designed a three-level hierarchy containing 13 criteria. For data acquisition, the differences in outcomes were investigated using individual versus group judgements separately. Furthermore, we analyzed the different effects when using the median and arithmetic and geometric means for data aggregation. A consistency ratio ≤0.2 was determined to represent an acceptable consistency level.

Results: Forty individual and three group judgements were collected from patients suffering from a rare disease and their close relatives. The consistency ratio of 31 individual and three group judgements was acceptable and thus these judgements were included in the study. To a large extent, the local ranks for individual and group judgements were similar. Interestingly, group judgements were in a significantly smaller range than individual judgements. According to our data, the ranks of the criteria differed slightly according to the data aggregation method used.

Conclusions: It is important to explain and justify the choice of an appropriate method for data acquisition because response behaviors differ according to the method. We conclude that researchers should select a suitable method based on the thematic perspective or investigated topics in the study. Because the arithmetic mean is very vulnerable to outliers, the geometric mean and the median seem to be acceptable alternatives for data aggregation. Overall, using the AHP to identify patient priorities and enhance the user-friendliness of information websites offers an important contribution to medical informatics.

Keywords: Decision-making, Analytic Hierarchy Process, Rare disease, Patient priorities, Internet homepage

* Correspondence: frp@cherh.de
Center for Health Economics Research Hannover (CHERH), Leibniz University of Hannover, Otto-Brenner-Straße 1, Hannover 30159, Germany

Background

The number of studies measuring patient priorities by using the Analytic Hierarchy Process (AHP) has increased significantly in the last few years [1]. The AHP was developed by Thomas L. Saaty in the 1970s to solve complex problems of multiple criteria decision-making [2], based on the idea that it is more reliable to judge the relative importance of several criteria with the help of respective pairwise comparison in a hierarchical structure than to judge their absolute importance [3]. The method was originally applied in the marketing sector and later in healthcare research. In addition, the AHP can be used to relate subjective criteria, which can be both quantitative and qualitative. As implied, it has been demonstrated that the AHP is a useful method for healthcare delivery as well as medical informatics decision-making [1, 4–7]. In this study, we ranked the information needs of people having a rare disease and their relatives using different AHP methods. This ranking of information needs is then transferred accordingly to display information categories on a web-based information portal about rare diseases in Germany. Because the available space on a user-friendly website homepage is restricted, the most important categories should be more accessible than less important categories. To present information categories on this website according to the user's priorities, this paper consulted both experts in medical informatics and patient-reported outcomes.

Today, approximately 4 million people in Germany suffer from rare diseases. The level in the United States is similar to that in Europe, with approximately 30 million people living with rare diseases. It is estimated that 400 million people worldwide suffer from a rare disease. Currently, international definitions of rare diseases vary greatly. For example in the EU, a disease is considered rare if it affects fewer than one in 2000 citizens, whereas in the United States a disease is considered rare if it affects fewer than 200,000 people, or about one in 1500 people [8, 9]. To improve patients' well-being, a national action plan for people with rare diseases was adopted by the Federal Government in Germany in 2013 that is supposed to coordinate national efforts invested in rare diseases. The establishment of a rare diseases information portal is one component of a broader set of planned measures, which includes 52 policy proposals [10]. Although conditions may differ significantly, patients having rare diseases and their relatives frequently face similar challenges [10, 11], which include protracted diagnosis processes as well as a deficient information base. To address these deficiencies, both medical experts and experts on medical informatics consider it relevant to assess the priorities of the (potential) patients and relatives.

As part of the development of an information portal for rare diseases, we used the AHP to identify the importance of several information types, e.g., information about therapy and social-legal advice. However, there are no best practices or a common gold standard available for applying the methods [1]. More precisely, it is noticeable that there are several methodological differences in the published studies concerning data acquisition and aggregation [1]. In some studies, single participants were interviewed (e.g. [12–14]), whereas in others, group discussions were used to analyze the priorities (e.g. [15, 16]). It therefore remains unknown which data acquisition method is more suitable for the AHP. To determine whether two methods (individual and group decisions) yield the same outcomes, we implemented them separately. The goals of this study were on the one hand to analyze the different influences of individual and group judgements on data acquisition, and on the other hand, to examine the different effects on the AHP results of using the arithmetic and geometric mean as well as the median for the data aggregation. We also discuss the degree to which the results of this study can be transferred to other disciplines. Finally, we fulfill our objective of providing a recommendation on choosing appropriate methods for further studies using the AHP.

Methods

Participants

Patients suffering from a rare disease were eligible to participate in the study. In addition, the relatives of these patients, for example, the parents of a child suffering from such a disease, were eligible to participate. The inclusion of both patient and relatives is necessary because many patients suffering from a rare disease are diagnosed as children, and the information priorities of the parents appear as a proxy for the children's priorities. Moreover, both patients and relatives will use the information portal. Patients were excluded if they were unable to concentrate continuously on the questionnaire or did not adequately understand the German language. Participants were recruited by the Freiburg Centre for Rare Diseases (Medical Center of the University Freiburg, Germany) and through rare disease self-help groups.

Analytic Hierarchy

The AHP is a stepwise problem-solving procedure. First, the decision-makers have to construct a hierarchical structure of the criteria. To achieve this, the multiple criteria decision problem must be broken down into its component parts [17]. The information needs of people suffering from a rare disease were identified by an Internet research study, including a review of already existing websites providing information on rare diseases. Furthermore, a preliminary qualitative study, the subjects of which were patients suffering from a rare disease, yielded important findings about the wording of the identified items that were regarded as

the defined targets. We designed a three-level hierarchy by grouping these items into information fields and information types.

The next step was to analyze the priorities. Patients and relatives were asked to compare every two information fields in the second level at each time with respect to the target. The information types in the third level were also compared pairwise with respect to the corresponding information field. Participants were asked to judge the importance of one endpoint as compared with another on a 9-point scale [18]. The participants also received printed ranking cards with the information fields and information types, which helped them provide consistent answers to the pairwise comparison questions. One example of a pairwise comparison is displayed in Fig. 1. It can be seen that "1" indicates that the two endpoints are of equal importance and "9" that the importance of one endpoint is extremely different from that of the other. Based on matrices of the pairwise comparisons, the standard AHP eigenvector method was used to calculate the patient's priorities using Microsoft Windows Excel [18]. The questionnaire used in the studies is avaliable as Additional file 1.

The final operation was consistency verification, which is listed as one of the key benefits of the AHP [19]. Saaty demonstrated that the consistency ratio (CR) can be calculated using the consistency index and the random index [18]. The CR value of a perfectly cardinal consistency matrix is 0. The CR value reflects the internal consistency of an observed set of judgements, and CR ≤ 0.2 has been determined to be an acceptable level of consistency [20, 21]. The results of participants who answered consistently were included in the analyses. Finally, the priorities of individual participants were aggregated to analyze the priorities of all the participants. The different data acquisition and aggregation methods are described in the following section.

Data acquisition

For data acquisition on individual decision-making, patients and relatives were interviewed. The interviews were conducted by telephone or in a face-to-face situation in a place familiar to the participant. In the case of telephone interviews, the AHP questionnaire was mailed to the participants a few days before the appointment. At the beginning of the interview, the structure of the AHP and the broad outline of the method, as well as all the quality criteria, were explained. Thereafter, the participants completed a guided AHP. Finally, the calculated individual weights (priorities of each criterion) were aggregated (Fig. 2) when the answers were consistent, as described above.

The same AHP questionnaire was used for the face-to-face group discussions. The group meetings were held at the Universities of Hannover, Frankfurt am Main, and Freiburg im Breisgau. After the interviewer presented a description of the structure and method of the AHP, each group member judged the relative priorities of each comparison. Then, the individual judgements (on a 9-point scale) were gathered and displayed anonymously on a screen. The group members discussed each pairwise comparison, as well as the rationales behind the individual judgements. Finally, for each pairwise comparison, a common group decision (consensus) was reached. The calculated group priorities were aggregated with all the other group priorities (Fig. 2) when the answers were consistent, as described above. The distribution of the priorities of individual and group weights was analyzed in separate box plots for each category using the statistics software R.

Data aggregation

Priorities can be aggregated using the arithmetic mean. According to a frequently used method for aggregating the priorities of individuals into a consensus rating, we also used the geometric mean [21–23]. In addition, we used the median to calculate the mean value of the priorities. The median divides the data set into two equal parts and indicates the mean value. The individual priorities were aggregated using each of these methods

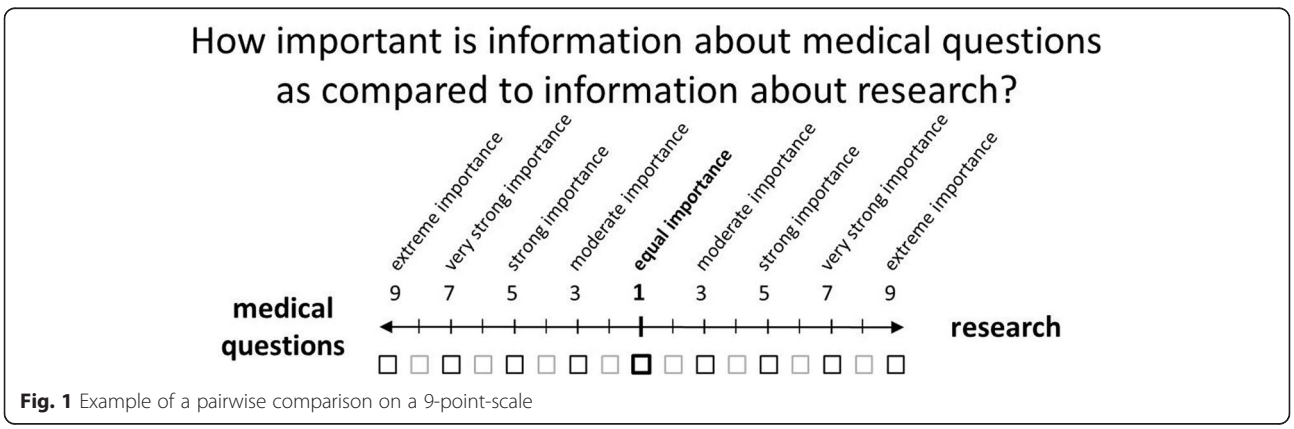


Fig. 1 Example of a pairwise comparison on a 9-point-scale

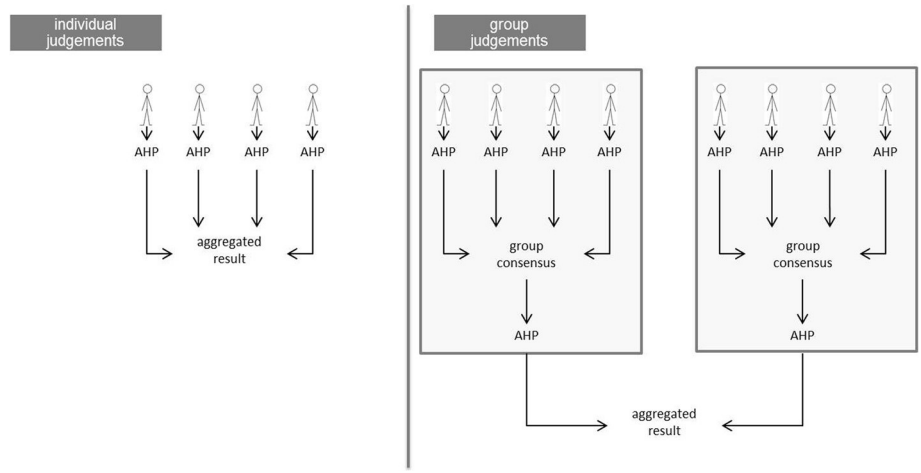


Fig. 2 Individual and group Analytic Hierarchy Process

independently to consider the different distributions resulting from the different methods. These results are presented in the "Data aggregation" subsection of the Results section.

Results

Participants

Thirty-six patients suffering a rare disease and four relatives ($n = 40$) having an average age of 50.7 years (ages ranged from 18 to 74 years) participated in the AHP in which the individual method was applied. In addition, for the group method, eight patients and three relatives were divided into three groups having a size of three or four participants. The average age of the group members was 52.2 years (ages ranged from 40 to 85 years). There were more female than male members in both populations. The average ages are relative high for both samples because adult relatives acted as a proxy for their children. Related to the issue, these relatives would search for information about rare diseases in the information portal. The following numbers of patients were suffering from the following rare diseases (note: the assignment to the orpha.net classification of rare diseases is not clearly regulated): rare skin diseases (five patients/two relatives), rare tumors (six patients), rare metabolic diseases (four patients), rare immunodeficiencies (seven patients), rare eye diseases (one patient), rare lung diseases (two patients/one relative), rare muscular diseases (two patients), rare blood count disorders (seven patients), rare genetic diseases (four patients/one relative), rare kidney diseases (two patients), rare skeletal dysplasia (one relative) and rare neurological diseases (four patients/two relatives). The demographic statistics of all the participants are displayed in Table 1. In addition to the information in the table, the average age at the time of diagnosis was 33.8 years for the individual AHP and 34.3 years for the group AHP; some patients were diagnosed at birth. The patients in the individual AHP had lived an average of 16.9 years since the diagnosis of a rare disease, and the group members had lived an average of 19 years since diagnosis. The marital status of the study population of the individual AHP was as follows: 27 of the 40 participants declared that they were married, six were divorced, and seven were living without a partner. Five of the group members were living with a partner, two were widowed, and four had no partner.

Analytic Hierarchy

The informational content of 300 websites maintained by providers of information about rare diseases was analyzed to identify the important items. These items were structured into a three-level hierarchy by grouping them into information fields and information types. We included four information fields: *medical questions, research*, *current events*, and *social counselling and assistance services*. Subsequently, we included nine information types: *diagnostics*, *therapy*, *disease pattern*, *new studies, study results*, *registers*, *social-legal advice*, *psychosocial counselling*, and *self-help*. The hierarchical structure (Fig. 3) contains the target on the first level, the information fields on the second level, and the information types on the third level. Consequently, for analyzing the priorities, 15 pairwise comparisons in each questionnaire were conducted: six comparisons of the four information fields on the second level and three times three comparisons of information types on the third level. An explanation of each information criterion was given to all participants, as shown in the Appendix.

Consistency ratio

The study sample showed a wide range of CRs. When the acceptable CR was set at a lower level, fewer participants could be included in the analyses. Moreover, the number

Table 1 Demographic statistics of the study population

Variable	Characteristics	Individual		Group	
		Frequency	Rate	Frequency	Rate
Sex	male	11	27.5 %	4	36.4 %
	female	29	72.5 %	7	63.6 %
Age	$x < 30$	2	5.0 %	0	0.0 %
	$30 \leq x < 50$	18	45.0 %	6	54.6 %
	$50 \leq x < 70$	16	40.0 %	4	36.4 %
	$x > 70$	3	7.5 %	1	9.1 %
Labor status	employed	17	42.5 %	6	54.6 %
	retired	11	27.5 %	2	18.2 %
	disabled	10	25.0 %	2	18.2 %
	student	1	2.5 %	0	0.0 %
	n/a	1	2.5 %	0	0.0 %
Estimated severity of the disorder	low	6	15.0 %	2	18.2 %
	medium	19	47.5 %	4	36.4 %
	high	15	37.5 %	5	45.5 %
Status	patient	36	90.0 %	8	72.7 %
	relative	4	10.0 %	3	27.3 %

of included participants decreased if consistency was required at all the investigated levels. Figure 4 shows an overview of the sample sizes according to the different levels of consistency. We determined an acceptable level of consistency to be a CR of 0.2 on the second level of the hierarchy. These parameters led to 31 individual judgements and all three group judgements being included in the analysis. However, the following results differed only slightly by determining a CR of 0.1.

Data acquisition

Further analyses were conducted by comparing individual and group priorities on the same level of consistency. The comparisons were conducted between individual and group priorities that were included in the CR = 0.2 category on the second level of the hierarchy. Figure 5 presents the corresponding local ranks of the information types (second level) and information fields (third level). To a large extent, the local ranks for individual and group judgements were similar. In both, *Information about medical questions* was the most relevant information type. In addition, the order of information fields (*diagnostics*, *therapy,* and *disease pattern*) in this information type was the same. Furthermore, in the second rank, information about *social counselling and assistance services* can be evaluated for individual and group priorities. Moreover, we found differences between individual and group judgements: *information about current events* was ranked higher by the group participants, and the order of the information fields *registers*, *new studies*, and *study results* differed.

In addition to the comparison above, we analyzed the weights of each category for the individual and group

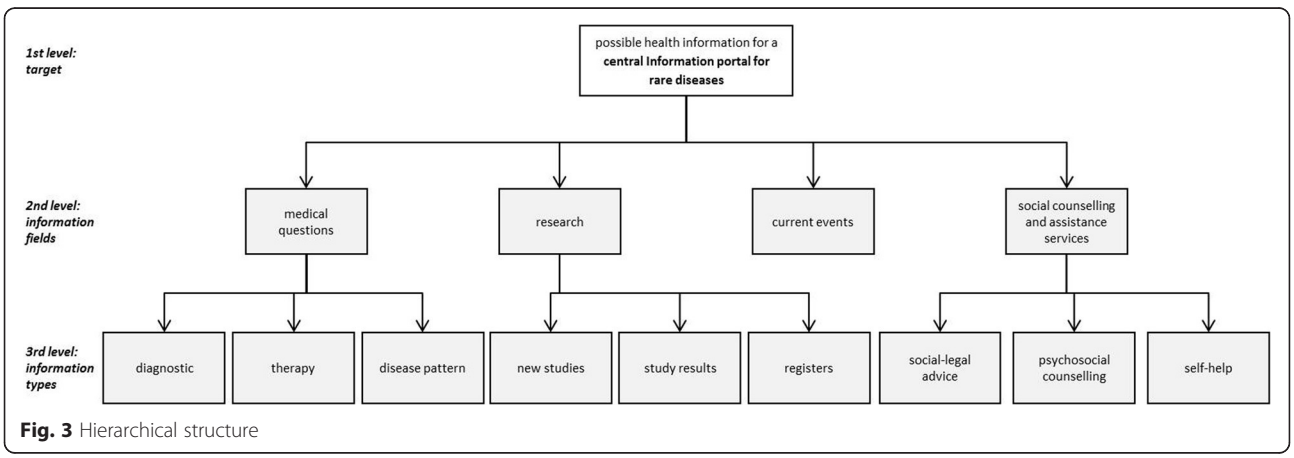


Fig. 3 Hierarchical structure

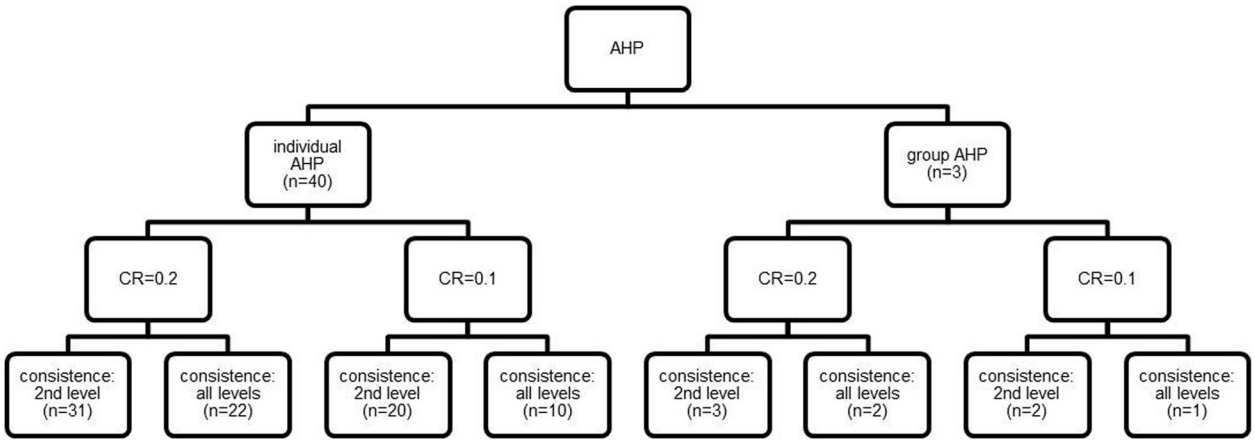


Fig. 4 Sample sizes by different levels of consistency ratio

priorities separately. The (global) weights quantify the priorities and allow all the information categories to be compared. The distribution of priorities for each category is displayed in Fig. 6. For each category, the distribution of group priorities (*group*) and individual priorities (*ind*) is shown. Based on the median, the differences between the individual and group priorities were small. For example, the weight of the category *information about medical questions* was noticeably higher for individual priorities. For the category *information about registers*, the weight was higher for group priorities. Moreover, we determined that the data span from minimum to maximum was most frequently greater for the individual priorities than for the group priorities.

Furthermore, we analyzed the answers given as individual judgements compared to those given as group judgements. The cumulative relative value distribution indicates the response behavior of individuals and groups. Figure 7 shows that group judgements frequently were in a narrower range than individual judgements; in particular, most of the judgements were located between 1 = equally important and 5 = very important. Stronger priorities (7 = very strongly important to 9 = extremely important) were not used in group judgements. The 45°-line symbolizes an equal distribution of the judgements between 1 = equally important and 9 = extremely important. Statistically significant differences between individual and group judgements ($p = 0.0027$) were found using a t-test analysis.

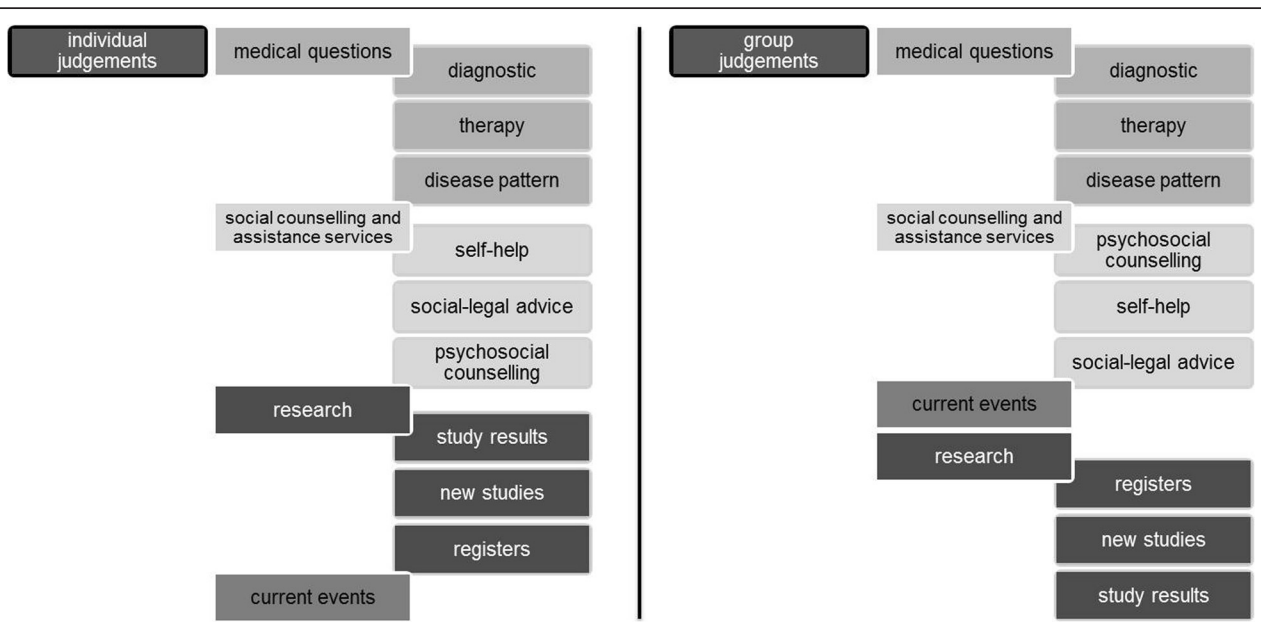


Fig. 5 Local ranks of individual and group judgements

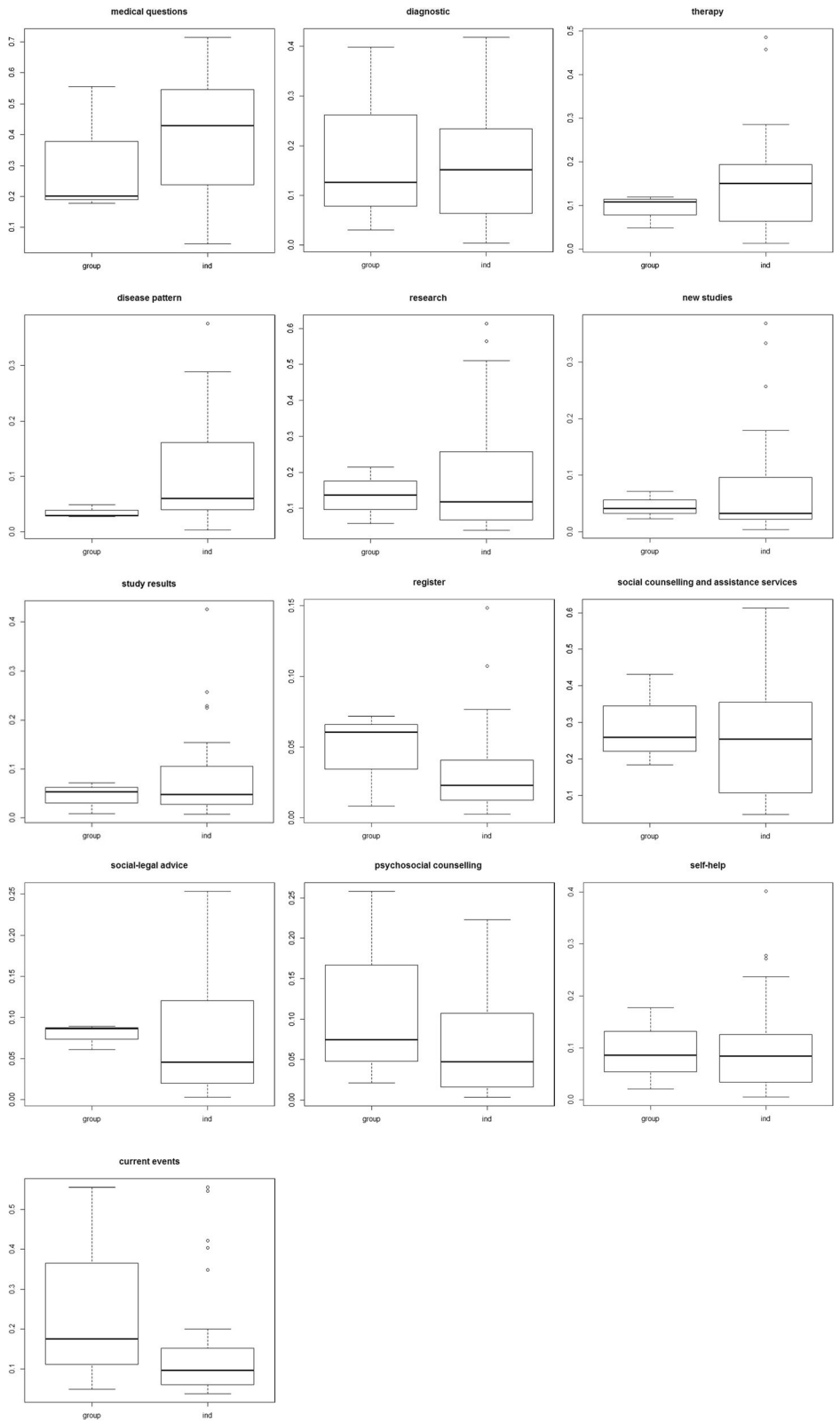


Fig. 6 Distribution of priorities of individual and group judgements

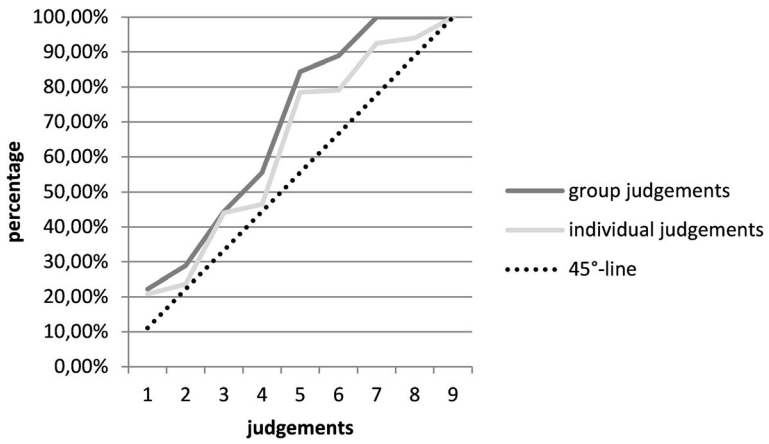


Fig. 7 Distribution of the given answers

Data aggregation

Aggregating single priorities is required to generate a summary of the study results. Depending on the data aggregation method, the ranks of the information criteria and the corresponding weights differ slightly. An advantage of using different methods separately is that the different distributions of the data sets can be considered and results can be compared between the methods.

Figure 8 shows the global ranks of the items grouped by the methods used for data aggregation (arithmetic and geometric mean, as well as the median). A comparison of the global ranks of the aggregation by the arithmetic mean with the aggregation by the geometric mean reveals that the criterion *information about diagnostics* had a lower priority if the data were aggregated by geometric mean. The same result was obtained for *information about new studies*. Other information criteria showed the same global ranking for both aggregation methods. A comparison of the global ranks of the aggregation by median with the aggregation by arithmetic mean showed that the criteria *information about self-help* and *information about disease patterns* changed ranks, as did the criteria *information about psychosocial counselling* and *information about new studies*. In summary, according to our data, there is no strong difference between the ranking of information criteria when the data are aggregated by the median or by the arithmetic or geometric mean.

Discussion

We have demonstrated that the AHP can be used to identify patient priorities with regard to the information needs of people having rare diseases. For this purpose,

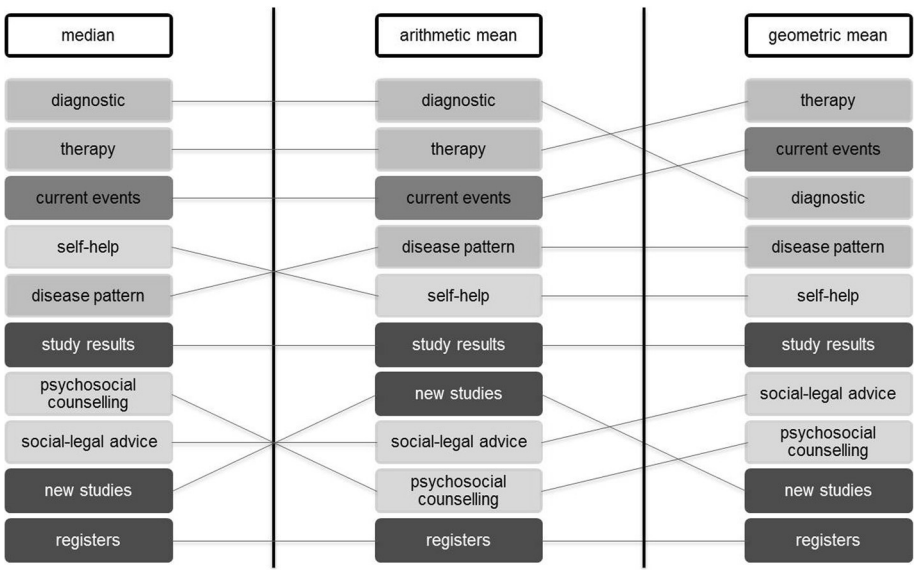


Fig. 8 Comparison of data aggregation by median and arithmetic and geometric mean

group decisions were as suitable as individual decisions. Although the local rank of the information types resulted in a similar order of individual and group decisions, their global weights varied slightly. Interestingly, we found another important aspect: group judgements were in a significantly smaller range than individual judgements. This result may be correlated with the fact that group judgements are more frequently consistent. Hence, it could conceivably be hypothesized that using smaller ranges, e.g., a 7- or 5-point scale, would lead to more consistent answers. Unfortunately, we cannot compare the response behavior with that reported in other published studies, because such an analysis was not conducted in these studies [1]. Furthermore, it can be argued that group decisions frequently represent the compromise solution of the group participants, and therefore, the group judgements are a mean of the individual judgements and consequently the group's priorities have a more limited range. We attempted to avoid a situation in which the group participants gave only the mean of their individual judgements as their answer. Frequently, the group participants discussed the rationales behind the individual judgements and decided on a common group priority that was not the mean of the individual judgements. Sometimes, the group judgement was even outside the range of the individual minimum and maximum judgements. There are, however, other possible explanations that should be investigated in further studies.

The findings of this study suggest that there is no "gold standard" method for data acquisition. According to our data, both the individual and group methods lead to very similar results. Moreover, there is no right or wrong ranking of the priorities of information needs. Researchers should select the most suitable method using other criteria, such as the thematic perspective of the study or the properties of the goods or topics that are addressed. It can be argued that, on the one hand, for free or non-rival goods, methods that involve individual decision-making are more suitable, because there is no need for the participants to be prepared to compromise; other people will not face disadvantages or advantages because of one individual's decision. On the other hand, group decisions are suitable for scarce or rival goods. Another aspect that should be considered is the peer pressure exerted in group discussions. The group situation can lead to particular disadvantages when intimate insights should be given in the interview, in which case, individual participants do not dare to answer truthfully or do not state their personal opinions. With regard to the implementation of the rare disease information portal or other websites, the order of information categories should not be influenced by other users. Therefore, an individual user's priorities shall be used to identify which information categories are more important and should be more accessible on the website than less important categories. In summary, the use of patient priorities to expand the user-friendliness of information websites using the AHP offers an important contribution for medical informatics.

According to our data, aggregations by median, arithmetic mean, and geometric mean lead to very similar rankings of information criteria. Because the arithmetic mean is very vulnerable to outliers, the median and the geometric mean appear to be acceptable alternatives for data aggregation, although the differences between the two methods depend on additional factors, such as the number of criteria in the hierarchy and the number of participants. Nevertheless, comparing the analyses using different methods offers the advantage of enabling consideration of the different distributions of the data sets.

The AHP method can lead to judgements that do not meet the defined CR requirement. We determined that the use of ranking cards prior to pairwise comparison of each category may help participants answer more consistently. Furthermore, we noticed that a comparison of four aspects of a category (such as the comparison of four information fields) is more challenging for participants than a comparison of three aspects of a category (such as the comparison of three information types) in terms of cardinal consistency. This fact was used to confirm the conditions for participation in this study: patients who were unable to concentrate on the questionnaire continuously were excluded, as well as children. This participation bias may lead to a non-representative ranking of the information needs of people suffering from a rare disease. Further applications of the AHP should consider restricting the number of pairwise comparisons in each category. Moreover, by setting a CR at ≤ 0.2, we could include a sufficient number of judgements in our analysis. If we had set a lower CR value, the number of included judgements would have been lower, and consequently, the informative value of this study would have been more limited.

Assumptions and limitations

The number of patients living with any one rare disease is limited. For this reason, we pooled patients with heterogeneous rare diseases, who frequently face similar challenges and have similar information needs. However, because of the relatively low number of participants interviewed in this study, the results may not be representative. Furthermore, a bias exists regarding the information criteria *current events*, because no information types were grouped in this information field. In addition, we attempted to minimize the interviewer bias, as well as the bias between telephone and face-to-face interviews.

Conclusions

To the best of our knowledge, this is the first study to investigate the differences in individual and group

judgements when conducting an AHP. Our study demonstrated the need for better strategies for choosing an appropriate method. Both methods led to similar outcomes; however, the response behavior differed. In brief, we demonstrated that the AHP can be used to identify the importance of several information types to people having a rare disease, and to order these information types on a website that presents information on rare diseases. Using the results of the AHP, we could rank the information needs of people suffering from a rare disease and their relatives according to their priorities. These priorities can be used to constitute information categories that are more important and should be more accessible on the website than less important categories. Overall, the use of an AHP to identify patient priorities and expand the user-friendliness of information websites offers an important contribution to medical informatics. According to our data, the use of different methods for data aggregation had no distinct influence on the ranking of the information criteria.

The strength of our study is in the transparent comparison of the different approaches applied in the AHP. The study indicates appropriate methods for conducting an AHP in other healthcare settings and in the field of medical informatics. Even if the results of the data acquisition methods do not differ, as was shown in our data, it is important that the researcher explain and justify the choice of method. We suggest that researchers select a suitable method based on the thematic perspective of the study or the properties of the goods or topics they are addressing. For example, it can be argued that group judgements should be used for studies addressing goods with limited availability. This investigation yielded important findings for subsequent studies that use the AHP method as a tool for medical decision-making and identifying patients' priorities.

Appendix

Definitions of the information criteria

Medical questions: Information that contains medical background information about rare diseases, e.g., information about diagnostics, therapy, or disease pattern.

Diagnostics: Information about diagnostic procedures using which a healthcare professional can identify rare diseases and make a diagnosis. In addition, contact information about specialized healthcare professionals or centers for rare diseases.

Therapy: Information about treatment procedures. In addition, contact information about healthcare professionals who can treat people suffering from a rare disease.

Disease pattern: Information about reasons for, symptoms, and progression of rare diseases.

Research: Information and results of scientists or pharmaceutical companies about new findings related to rare diseases.

New studies: Investigations of medical treatments of rare diseases that are scheduled or starting immediately for which participants are still being sought.

Study results: Results of current medical research.

Registers: Collections of disease data in the long term to improve the treatment opportunities and to monitor the distribution of the diseases.

Current events: Information and important appointments for public meetings where patients and affected persons can talk to healthcare staff.

Social counselling and assistance services: Contact data for and information about counselling centers that can help people suffering from a rare disease.

Social-legal advice: Here, answers can be found to questions concerned with the services of statutory health insurance, labor laws, or statutory pension funds.

Psychosocial counselling: Information and contact data that can provide psychosocial counselling in the case of illness-related problems of family, friends, or coworkers.

Self-help: Contact information about support groups of patients and close relatives.

Additional file

Additional file 1: Questionnaire. (PDF 556 kb)

Abbreviations

AHP, analytic hierarchy process; CHERH, center for health economics research hannover; CR, consistency ratio; Ind, individual.

Acknowledgements

The authors wish to thank the patients and their families for participating in this study.

Funding

The Center for Health Economics Research Hannover (CHERH) is funded by the Federal Ministry of Education and Research. This study was funded in part by the Federal Ministry of Health Germany. The study sample consisted of randomly selected participants from the qualitative main study of the ZIPSE project.

Availability of data and materials

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Authors' contributions

FP carried out the analyses and drafted the manuscript. FP and KS prepared the data adequately. FP and AB collected the data. KS revised the manuscript. KD and MF made substantial contributions to the conception of the article. JMS revised the manuscript for important intellectual content. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate
Ethical approval was issued by the Ethics Committee of the Albert-Ludwigs-Universität Freiburg (53/14). Informed consent was obtained from all participants prior to the survey and interviews.

Received: 28 January 2016 Accepted: 3 August 2016
Published online: 09 September 2016

References

1. Schmidt K, Aumann I, Hollander I, Damm K, Schulenburg JM. Applying the Analytic Hierarchy Process in healthcare research: A systematic review and evaluation of reporting. BMC Med Inform Decis Mak. 2015;15:115.
2. Saaty TL. A Scaling Method for Priorities in Hierarchical Structures. J Math Psychol. 1977;15:234–81.
3. Pecchia L, Melillo P. Analytic Hierarchy Process for Health Technology Assessment: A Case Study for Selecting a Maintenance Service Contract. In: Ventre AGS, Maturo A, Hošková-Mayerová S, Kacprzyk J, editors. Multicriteria and Multiagent Decision Making with Applications to Economics and Social Sciences Studies in Fuzziness and Soft Computing. Heidelberg: Springer; 2013. p. 275–88.
4. Dolan JG. Are Patients Capable of using the Analytic Hierarchy Process and Willing to Use It to Help Make Clinical Decisions? Med Decis Mak. 1995;15:76–80.
5. Pecchia L, Bath PA, Pendleton N, Bracale M. Analytic Hierarchy Process (AHP) for Examining Healthcare Professionals' Assessments of Risk Factors. Methods Inf Med. 2011;50:435–44.
6. Liberatore MJ, Nydick RL. The Analytic Hierarchy Process in Medical and Health Care Decision Making: A Literature Review. Eur J Oper Res. 2008;189(1):194–207. doi:10.1016/j.ejor.2007.05.001.
7. Hajrahimi N, Dehaghani SM, Hajrahimi N, Sarmadi S. Quality assessment of Isfahan Medical Faculty web site electronic services and prioritizing solutions using analytic hierarchy process approach. J Educ Health Promot. 2014;3:117. doi:10.4103/2277-9531.145920.
8. Kaplan W, Wirtz VJ, Mantel-Teeusisse A, Stolk P, Duthey B, Laing R. Priority Medicines for Europe and the World 2013 Update, World Health Organization. 2013. http://www.who.int/medicines/areas/priority_medicines/MasterDocJune28_FINAL_Web.pdf. Accessed 21 Dec 2015.
9. Regulation (EC) No 141/2000. Regulation (EC) No 141/2000 of the European Parliament and of the Council of 16 Dec 1999 on Orphan Medicinal Products. Off J Europ Com. 1999. http://ec.europa.eu/health/files/eudralex/vol-1/reg_2000_141/reg_2000_141_en.pdf. Accessed 21 Dec 2015.
10. National Action Alliance for People with Rare Diseases. National Action Plan for People with Rare Diseases. 2013. http://www.namse.de/images/stories/Dokumente/Aktionsplan/national%20plan%20of%20action.pdf. Accessed 21 Dec 2015.
11. Eidt D, Frank M, Reimann A, Wagner TOF, Mittendorf T, Schulenburg JM. Maßnahmen zur Verbesserung der gesundheitlichen Situation von Menschen mit seltenen Erkrankungen in Deutschland. 2009. http://www.bmg.bund.de/uploads/publications/BMG-G-09050-Bericht-Massnahmen-seltene-Krankheiten_200908.pdf. Accessed 21 Dec 2015.
12. Chung KP, Chen LJ, Chang YJ, Chang YJ, Lai MS. Application of the Analytic Hierarchy Process in the Performance Measurement of Colorectal Cancer Care for the Design of a Pay-for-performance Program in Taiwan. Int J Qual Health Care. 2013;25:81–91.
13. Dolan JG, Boohaker MD, Allison J, Imperiale TF. Patients Preferences and Priorities regarding Colorectal Cancer Screening. Med Decis Mak. 2013;33:59–70.
14. Suner A, Celikoglu CC, Dicle O, Sökmen S. Sequential Decision Tree using the Analytic Hierarchy Process for Decision Support in Rectal Cancer. Artif Intell Med. 2012;56:59–68.
15. Danner M, Hummel JM, Volz F, van Manen JG, Wiegard B, Dintsios CH, Bastian H, Gerber A, Ijzerman AJ. 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;27:369–75.
16. Taghipour H, Mohammadyarei T, Jafarabadi MA, Hashemi AA. On-Site or Off-side Treatment of Medical Waste: A Challenge. J Environ Health Sci Eng. 2014;12:86.
17. Dolan JG, Isselhardt BJ, Cappuccio JD. The Analytic Hierarchy Process in Medical Decision Making: A Tutorial. Med Decis Mak. 1989;9:40–50.
18. Saaty TL. The Analytic Hierarchy Process. New York: McGraw Hill; 1980.
19. Ho W. Integrated Analytic Hierarchy Process and its Applications – A Literature Review. Eur J Oper Res. 2008;186:211–28.
20. Dolan JG. Shared Decision-making - Transferring Research into Practice: The Analytic Hierarchy Process (AHP). Patient Educ Couns. 2008;73:418–25.
21. Brinkmeyer D, Müller R. Entscheidungsunterstützung mit dem AHP. Z Agrarinformatik. 1994;2:82–92.
22. Saaty TL, Shang JS. Group Decision-making: Head-count versus Intensity of Preference. Socioecon Plann Sci. 2007;41:22–37.
23. Aczel TL, Saaty T. Procedures for Synthesizing Ratio Judgement. J Math Psychol. 1983;27:93–102.