



Nutzenbewertung mittels MCDA

Ein umfassendes Bewertungstool für neue Krebstherapien
(Aktualisierung Bevölkerungsbefragung)

Studie im Auftrag von Bristol-Myers Squibb SA



Nutzenbewertung mittels MCDA

Studie im Auftrag von Bristol-Myers Squibb SA

Marc Bill

Dr. Barbara Fischer

Dr. Florian Kuhlmei

Dr. Harry Telser

23. Mai 2023

Inhaltsverzeichnis

Das Wichtigste in Kürze	4
1 Ausgangslage und Projektziel	7
2 Literatur und methodische Grundlagen	9
2.1 MCDA und HTA	9
2.2 ISPOR Good-Practice-Richtlinien	10
3 MCDA für nichtkurative Krebstherapien	14
3.1 Problemdefinition	14
3.2 Kriterienwahl: Bestimmung des Wertebaums für nichtkurative Therapien	15
3.2.1 Vorgehen und Wertebaum	15
3.2.2 Ausgestaltung Attribute	16
3.3 Leistung messen	20
3.4 Bewertung der Attribute	21
3.4.1 Bestimmung der Minima und Maxima	22
3.4.2 Lineare und nichtlineare Wertefunktionen	24
3.4.3 Massgebende Präferenzen für die Wertefunktionen	26
3.4.4 Methodenwahl	26
3.5 Gewichtung der Attribute	27
3.6 Zusammenfassung der Einzelnutzen und Kategorisierung	28
3.7 Umgang mit Unsicherheit: Sensitivitätsanalyse	29
4 Anwendungsbeispiel	31
4.1 Stichprobenauswahl und Befragungsdesign	31
4.2 Statistische Auswertung und Resultate DCE	33
4.3 Präferenzheterogenität	36
4.4 Nutzenbewertung: Fortgeschrittenes Nierenzellkarzinom	37
5 Fazit	44
6 Bibliografie	45
Anhang A: Beteiligte Akteure	50
Anhang B: Methoden zur Präferenzmessung	51
Anhang C: Discrete-Choice-Experimente	53
Anhang D: Resultat der Expertenbefragung	72

Das Wichtigste in Kürze

Fortschritte in der Behandlung von Krebs haben in den letzten Jahren unter anderem dazu geführt, dass Krebspatienten heute eine deutlich bessere Prognose haben. Die zunehmenden Innovationen bei Onkologika stellen das Zulassungsverfahren zur Vergütung durch die obligatorische Krankenpflegeversicherung (OKP) jedoch vor Herausforderungen. Dem Bundesamt für Gesundheit (BAG) wurde vorgeworfen, dass bei Aufnahmegesuchen auf die Spezialitätenliste insbesondere die Kriterien bei der Prüfung von Wirksamkeit, Zweckmässigkeit und Wirtschaftlichkeit (WZW) nur unzureichend präzisiert seien und namentlich der Nutzen der Medikamente ungenügend und nicht nach einheitlichen Kriterien bewertet werde.

Ziel des vorliegenden Projekts ist es, ein Instrument zu entwickeln, welches die vielfältigen Nutzenkomponenten von innovativen Medikamenten besser reflektiert und eine einheitliche Bewertung neuer Behandlungsmethoden allgemein und damit ausserhalb der Einzelfallbeurteilung ermöglicht. Dabei soll die Bewertung schneller und einfacher umgesetzt werden können als das derzeitige Bewertungsverfahren zur Aufnahme in die Spezialitätenliste. Es soll zudem die Kritik am heutigen Verfahren adressieren und durch Standardisierungen zu einer Konkretisierung der WZW-Kriterien beitragen. Durch den Einbezug weiterer, heute nicht berücksichtigter Nutzenkomponenten sowie auch des potenziellen Schadens einer Therapie, gilt dies vor allem für die beiden Kriterien Wirksamkeit und Zweckmässigkeit. Das Kriterium der Wirtschaftlichkeit ist nicht Ziel des vorliegenden Projekts. Primärer Anwendungsbereich des Bewertungsinstruments ist die Onkologie.

Der vorliegende Bericht beschreibt die Entwicklung und Modellierung eines einheitlichen Nutzenbewertungsmodells für die Schweiz. Dazu wird auf die Multi-Criteria Decision Analysis (MCDA) als Methode zurückgegriffen, welche es erlaubt, die Werte mehrerer unterschiedlicher Nutzenkomponenten auf einer wissenschaftlichen Basis zu einem Gesamtnutzenwert zu aggregieren. Das Potenzial von MCDA für solche Bewertungen ist seit längerem bekannt und zu ihrer Umsetzung existieren internationale Richtlinien der «International Society for Pharmacoeconomics and Outcomes Research» (ISPOR). Eine Übersicht über die Themen dieser Richtlinien und der Behandlung im vorliegenden Bericht bietet Tabelle 1.

Tabelle 1 Vorgehen zur Durchführung einer MCDA gemäss ISPOR-Checkliste

Schritt gemäss ISPOR-Checkliste		Behandlung in Kapitel
1	Definition des Entscheidungsproblems	3.1
2	Auswahl und Strukturierung der relevanten Attribute für die MCDA	3.2
3	Leistung messen	3.3
4	Bewertung der Attribute	3.4
5	Gewichtung der Attribute	3.5
6	Zusammenfassung der Einzelnutzen	3.6
7	Umgang mit Unsicherheit	3.7
8	Berichterstattung	4 (Anwendungsbeispiel)

Quelle: Marsh et al. (2016), eigene Darstellung Polynomics.

Das für die Schweiz entwickelte Modell leiten wir zunächst möglichst allgemein her. Den Fokus legen wir dabei auf nichtkurative Therapien für Krebserkrankungen. Für andere Therapieformen (adjuvante und potenziell kurative Therapien sowie Therapien für seltene Erkrankungen) müssten

angepasste Modelle entwickelt werden. Grundsätzlich vergleichen wir im Modell jeweils die Standardbehandlung mit einer neuen Behandlungsmethode. Beide werden getrennt voneinander bewertet; der Nutzengewinn der neuen Therapie wird dann durch den Vergleich der beiden Ergebnisse ermittelt. Die resultierende Nutzendifferenz wird anschliessend in grobe Kategorien eingeteilt, um nicht eine Genauigkeit vorzutäuschen, die mit dem Bewertungstool nicht erreicht werden kann.

In Absprache mit Experten aus der Onkologie, der Gesundheitsökonomie und der Pharmaindustrie entstand der in Tabelle 2 dargestellte Wertebaum für nichtkurative Therapien (Schritt 2). Die jeweiligen Messgrössen (Attribute) für die Nutzenkriterien wurden mit Hilfe einer Fokusgruppe weiter konkretisiert. Um den Einsatzbereich des Modells möglichst breit zu halten, sind teilweise mehrere Varianten zugelassen, um ein Attribut zu messen. So kann das Attribut Überleben entweder als Medianüberlebenszeit oder als progressionsfreies Überleben gemessen werden.

Tabelle 2 Wertebaum für nichtkurative Therapien

Dimension	Kriterium	Attribut
Patientennutzen	Behandlungserfolg	▪ Überleben
		▪ Gesundheitsbezogene Lebensqualität
	Sicherheitsprofil	▪ Unerwünschte Ereignisse
		▪ Patientenkomfort
Gesellschaftlicher Nutzen	Krankheitslast	▪ Disability-adjusted life years (DALY)
	Kosten	▪ Indirekte Kosten ▪ Weitere Behandlungskosten

Quelle: Eigene Darstellung, Polynomics.

Die Resonanz zum hier vorgestellten Nutzenbewertungssystem war bei den konsultierten Experten grossmehrheitlich positiv. Den Nutzenbegriff breiter zu fassen und insbesondere auch den sozialen Nutzen in die Bewertung einfließen zu lassen, wurde als äusserst relevant erachtet.

Das vorgeschlagene Modell ist als Hilfsmittel zur Bestimmung der Kostenübernahme durch die OKP gedacht, ist jedoch nicht zur spezifischen Preissetzung konzipiert. Grundsätzlich lässt sich das Modell aber überall verwenden, wo Nutzenbewertungen von Arzneimitteln im regulatorischen Kontext gebraucht werden, zum Beispiel beim therapeutischen Quervergleich (TQV), bei Priorisierungen, bei wertebasierten Vergütungsmodellen wie Value Based Pricing. Das Konzept ist gut anpassbar und erlaubt für solche Anwendungsformen den einfachen Einbezug weiterer Nutzendimensionen.

Um zu zeigen, dass ein Instrument zur standardisierten Nutzenbewertung von Onkologika mittels einer MCDA für die Schweiz praxistauglich umsetzbar ist, haben wir ein konkretes Anwendungsbeispiel erstellt. Für die Gewichtung und Bewertung der Attribute haben wir die Präferenzen der Schweizer Bevölkerung mittels eines Discrete-Choice-Experimentes (DCE) ermittelt. Wir messen also nicht den Nutzen betroffener Patientinnen und Patienten, sondern eines Querschnitts über alle in der Schweiz lebenden Personen. Diese Perspektive ist geeignet, um den Wert von Behandlungen aus Sicht der OKP zu bestimmen.

Basierend auf der repräsentativen Bevölkerungsbefragung ergeben sich die in Tabelle 3 dargestellten Gewichte. Das wichtigste Attribut ist die Überlebenszeit nach Therapiebeginn mit einem Gewicht von 44 Prozent. Die langfristige gesundheitsbezogene Lebensqualität erhält ein Gewicht von 26 Prozent und die während der Behandlung auftretenden schweren Nebenwirkungen ein

Gewicht von 14 Prozent. Der Patientenkomfort bestehend aus Behandlungshäufigkeit und Behandlungsort kommen gemeinsam auf ein Gewicht von rund 9 Prozent. Damit resultiert für den Patientennutzen insgesamt ein Gewicht von 93 Prozent, die Attribute, die den gesellschaftlichen Nutzen messen, erhalten insgesamt ein Gewicht von 7 Prozent. Es ist festzuhalten, dass die resultierende Gewichtung für das Setting nichtkurativer Krebstherapien ermittelt wurde und nicht einfach in einen anderen Kontext übertragen werden darf. Im vorliegenden Kontext ist die Gewichtung jedoch als robust anzusehen. Es zeigen sich keine grossen Unterschiede zwischen verschiedenen Bevölkerungsgruppen und eine frühere Expertenbefragung kam ebenfalls zu einem vergleichbaren Resultat.

Tabelle 3 Gewichtung der Attribute im Setting nichtkurativer Krebstherapien

Dimension	Attribut	Gewicht
Patientennutzen	Überlebenszeit nach Therapiebeginn (Median)	44.4%
	Langfristige Lebensqualität	26.0%
	Schwere unerwünschte Ereignisse	13.8%
	Behandlungshäufigkeit	3.4%
	Behandlungsort	5.3%
Gesellschaftlicher Nutzen	Krankheitslast in der Gesellschaft	2.4%
	Indirekte Kosten für die Gesellschaft	0.5%
	Weitere Behandlungskosten	4.2%

Quelle: Eigene Berechnungen, Polynomics.

Das so kalibrierte Modell wurde anschliessend für die nichtkurative Therapie des fortgeschrittenen Nierenzellkarzinoms angewendet. Als Arzneimittelinnovation diente uns die Kombinations-therapie aus den Wirkstoffen Nivolumab plus Ipilimumab, als Komparator die Therapie mit Sunitinib. Gemäss dem Modell resultiert aufgrund der Innovation eine «grosse Nutzenverbesserung». Das Beispiel zeigt auf, dass MCDA eine geeignete Methode ist, um zuverlässig und einfach eine standardisierte Nutzenwertbestimmung verschiedener Therapien durchzuführen.

1 Ausgangslage und Projektziel

Aus Patientensicht haben sich die Aussichten nach einer Krebsdiagnose in den letzten Jahrzehnten deutlich verbessert (Bill et al., 2020). Dieser Erfolg ist das Produkt von vielen Faktoren. So wird Krebs heute häufig früher erkannt und ist dann besser behandelbar, Risikofaktoren für besonders gefährliche Krebsarten wurden identifiziert, Lebensgewohnheiten (wie Rauchen und Alkoholkonsum) haben sich geändert und Arbeitsbedingungen wurden verbessert. Und natürlich wurde auch die Behandlung von Krebs deutlich erfolgreicher – sowohl durch die Verbesserung und Kombination bestehender Behandlungsschemata als auch durch die Etablierung neuer Therapieansätze. Zu den wichtigsten Fortschritten in der Onkologie zählen die Immun- und zielgerichtete Therapien sowie zuletzt auch die biomarkergetriebenen Kombinationstherapien. Unter den vielversprechendsten neuen Therapieansätzen, die vereinzelt bereits im Einsatz sind und in naher Zukunft zur Behandlung mehrerer Krebsarten greifbar erscheinen, sind zelluläre Therapien und Gentherapien.

Diese innovativen Onkologika stellen hohe Anforderungen an das Gesundheitssystem und insbesondere an die Verfahren zur Marktzulassung und zur Kostenübernahme durch die OKP. Die Regeln dieser Verfahren sind teilweise noch vergangenheitsorientiert und wurden von den medizinischen Fortschritten überholt. Dadurch verzögert sich der Zugang zu Innovationen für den Patienten.

Die bestehenden Herausforderungen haben sich nicht zuletzt darin geäussert, dass gegenwärtig Hersteller, Leistungserbringer und Krankenkassen in der Schweiz vermehrt in die Einzelfallvergütung (nach Art. 71a-d KVV) «flüchten», wenn die Behandlung eines Patienten mit einer neuen Therapie zwar möglich ist, aber der langwierige Vergütungsprozess nach der Zulassung noch nicht abgeschlossen ist. Auch wenn Art. 71 a-d KVV ursprünglich eine Ausnahmeregelung darstellen sollte, ist die Zahl der Kostengutsprachegehalte in den letzten Jahren für von Swissmedic zugelassene und nicht zugelassene Therapien stark angestiegen (Kägi et al., 2020). Dieser sogenannte off-label (Art.71a KVV) und off-list/limitatio (Art.71b KVV) use (OLU) von Medikamenten ermöglicht den gewünschten raschen Zugang zu neuen und innovativen Behandlungen. Die Einzelfallvergütung ist aber für Versicherer, Hersteller und die Leistungserbringer in der Abwicklung aufwändig und die Beurteilung ist uneinheitlich (Fries et al., 2018). So ist nicht sichergestellt, dass jeder Patient den gleichen Zugang zu Behandlungen hat. Um diesem Defizit entgegenzuwirken, hat die Schweizerische Gesellschaft der Vertrauens- und Versicherungsärzte (SGV) 2018 eine einheitliche Bewertung der Einzelfälle vorgeschlagen und dieses Modell speziell für Onkologika verfeinert.¹ Es ist jedoch nur im Kontext einer Einzelfallbeurteilung und für die Anwendung durch Ärzte ausgelegt und damit kein allgemeines Bewertungstool.

Auch die Einhaltung der WZW-Kriterien, wonach die von der obligatorischen Krankenversicherung übernommenen Leistungen wirksam, zweckmässig und wirtschaftlich (WZW) sein müssen (Art. 32 KVG), ist bei einer Einzelfallentscheidung schwieriger umzusetzen. In der Vergangenheit wurden dem BAG aber gerade hier Vorwürfe gemacht; nämlich, dass die WZW-Kriterien bei der Prüfung der Aufnahmegesuche auf die Spezialitätenliste nur unzureichend präzisiert seien und namentlich der Nutzen der Medikamente ungenügend und nicht nach einheitlichen Kriterien

¹ Siehe «Nutzenbewertung zur Vergütung von Arzneimitteln im Einzelfall (Art. 71a – 71d KVV)», verfügbar unter <http://www.vertrauensaezte.ch/expertcom/71kvv/updmay18/>, zuletzt aufgerufen am 23.08.2021. Das «OLUtool Onko» genannte Modell ist verfügbar unter http://www.vertrauensaezte.ch/expertcom/71kvv/updmay18/olutool-onko_3_2.pdf, zuletzt abgerufen am 23.08.2021.

bewertet werde (Parlamentarische Verwaltungskontrolle, 2013). Bemängelt wurde auch der fehlende relative Vergleich der Wirksamkeit zu anderen Medikamenten sowie der fehlende bzw. nur situative Einbezug von sozialen Präferenzen. Damit gemeint ist die Erfassung und Berücksichtigung des «Willens» der Bevölkerung.

Ziel des vorliegenden Projekts ist es, ein Instrument zu entwickeln, welches den Nutzen von innovativen Medikamenten besser reflektiert und eine einheitliche Bewertung neuer Behandlungsmethoden allgemein und damit ausserhalb der Einzelfallbeurteilung ermöglicht. Dabei soll die Bewertung schneller und einfacher umgesetzt werden können als das derzeitige Bewertungsverfahren zur Aufnahme in die Spezialitätenliste. Es soll zudem die Kritik am heutigen Verfahren adressieren und durch Standardisierungen zu einer Konkretisierung der WZW-Kriterien beitragen. Durch den Einbezug weiterer Nutzenkomponenten sowie auch des potenziellen Schadens einer Therapie, gilt dies vor allem für die beiden Kriterien Wirksamkeit und Zweckmässigkeit. Das Kriterium der Wirtschaftlichkeit ist nicht Ziel des vorliegenden Projekts. Das Bewertungsinstrument soll nicht dazu dienen, einen wirtschaftlichen Preis für eine neue Therapie zu ermitteln, es ist lediglich als Hilfsmittel gedacht, um den Nutzen von neuen Therapien standardisiert, umfassend und einheitlich in Health Technology Assessments (HTA) zu bestimmen. Idealerweise lassen sich durch ein solches Instrument die oben genannten Probleme reduzieren. Primärer Anwendungsbereich des Bewertungsinstruments ist die Onkologie.

Der vorliegende Bericht beschreibt die Entwicklung und Modellierung eines einheitlichen Nutzenbewertungsmodells für die Schweiz. Kapitel 2 enthält die methodischen Grundlagen der verwendeten Methode, die Multi-Criteria Decision Analysis (MCDA), und eine Übersicht relevanter Anwendungen in HTA. Das für die Schweiz entwickelte Modell ist im Kapitel 3 dargestellt. Eine Anwendung dieses Modells mit einer konkreten Nutzenbewertung einer Arzneimittelinnovation findet sich in Kapitel 4. Der Bericht schliesst mit einem Fazit in Kapitel 5.

2 Literatur und methodische Grundlagen

2.1 MCDA und HTA

Bei Entscheidungen müssen oftmals viele Kriterien berücksichtigt und bewertet werden. Entscheidungsprobleme sind dabei unterschiedlich komplex. Bei einfachen individuellen Entscheidungen, wie zum Beispiel der Auswahl der Kleidungsstücke für den bevorstehenden Tag oder der Wahl des Abendessens im Restaurant, sind wir fähig, die Alternativen unter Berücksichtigung der relevanten Kriterien gegeneinander abzuwägen und eine optimale Entscheidung zu treffen. Bei komplexen kollektiven Entscheidungen jedoch, bei welchen oftmals verschiedene Interessensgruppen zu berücksichtigen sind und die weitreichende Auswirkungen auf die Gesellschaft haben, können Methoden zur Entscheidungsfindung dazu beitragen, dass die wichtigsten Kriterien korrekt und konsistent berücksichtigt werden. Die Multi-Criteria Decision Analysis (MCDA) ist solch ein formaler Ansatz, der Entscheidungsträgern helfen kann, Information zu organisieren und synthetisieren, um komplexe Entscheidungen zu erkunden und die optimale Entscheidung zu treffen (vgl. Belton and Stewart, 2002).

In vielen Gesundheitssystemen haben sich Health Technology Assessments (HTA) zur Nutzenbewertung von neuen Therapien und Arzneimitteln etabliert (vgl. Baltussen et al., 2019). Dabei haben sich Methoden durchgesetzt, die insbesondere die Lebensdauer der Patienten in Kombination mit der gesundheitsbezogenen Lebensqualität bei der Nutzenbewertung berücksichtigen, also den Patientennutzen bestimmen (vgl. bspw. Angelis and Kanavos, 2017). Der soziale Nutzen, der durch eine Behandlung entsteht, wird jedoch meist ausser Acht gelassen oder zumindest nicht systematisch und vollständig zu erfassen versucht. Dazu zählen alle Nutzenkomponenten, die einen Einfluss nicht nur auf den Patienten, sondern auch auf sein Umfeld haben – also etwa reduzierter Pflegebedarf durch Angehörige, eine schnellere Rückkehr zum Arbeitsplatz, eine geringere Belastung des Gesundheitssystems durch die Reduzierung von Begleiterkrankungen, aber auch gesundheitliche Ungleichheiten zu reduzieren oder die Menschen vor den Verarmungseffekten von Krankheit zu schützen. Die MCDA bietet hier Entscheidungsträgern einen strukturierten Weg, die verschiedenen Nutzenkomponenten der Gesellschaft in HTA einzuschliessen und konsistentere Entscheidungen zu treffen.

Das Potenzial von MCDA für HTA wurde auch von der «International Society for Pharmacoeconomics and Outcomes Research» (ISPOR) erkannt, welche die MCDA Emerging Good Practices Task Force schuf. Diese Arbeitsgruppe hat in zwei Papieren Richtlinien für die Nutzung von MCDA zur Entscheidungsfindung im Gesundheitswesen erarbeitet (vgl. Marsh et al., 2016; Thokala et al., 2016). Ihre Erkenntnisse, Einordnungen und Empfehlungen bilden die Grundlage für viele Anwendungen von MCDA bei HTA.

Diese Anwendungen hielten in den letzten Jahren in vielen Gesundheitssystemen und Fachgebieten Einzug. Ein Übersichtsartikel von Oliveira et al. (2019) fand, dass der Fokus der gefundenen Studien zwischen 1990 und 2017 auf Arzneimittel (42%), allgemeine Gesundheitstechnologien (28%), Gesundheitsinterventionen (24%), medizinische Geräte (15%) und Impfungen (3%) lag. Im Bereich der Arzneimittel fanden am meisten MCDA-Anwendungen zu seltenen Krankheiten (vgl. Baran-Kooiker et al., 2018; Schey et al., 2020) und Krebs (vgl. Angelis et al., 2020; Ezeife et al., 2020; Garau, 2018; Hsu et al., 2019; Kwon, 2017; Wagner, 2017) statt. In den letzten Jahren kam es auch vermehrt zu Bewertungen von Arzneimitteln im off-patent oder off-label use (vgl. Abdullah et al., 2019; Brixner et al., 2017; Inotai et al., 2018). Weiter befassten sich einige Autoren generell mit der Umsetzung von MCDA als HTA-Instrument in den Gesundheitssystemen von beispielsweise Frankreich (vgl. Ghabri et al., 2019) oder Kanada (vgl. Laba et al., 2020).

2.2 ISPOR Good-Practice-Richtlinien

Obwohl MCDA vermehrt als ein HTA-Instrument diskutiert wird, sind die tatsächlich umgesetzten Anwendungen nach wie vor eher rar. Aus methodologischer Sicht wurden jedoch wichtige Eckpfeiler für die Anwendung von MCDA als HTA in den letzten Jahren geschaffen. Die MCDA Emerging Good Practices Task Force hat zwei Papiere mit Richtlinien für die Nutzung von MCDA zur Entscheidungsfindung im Gesundheitswesen erarbeitet (vgl. Marsh et al., 2016; Thokala et al., 2016).

Auch die hier vorliegende Arbeit baut auf der darin erarbeiteten achtstufigen Checkliste für die Gestaltung, Durchführung und Bewertung von MCDA auf und hält sich an die empfohlenen Richtlinien. Die Task Force hat jedoch ganz bewusst auf weitere Konkretisierungen für die Anwendung von MCDA als HTA verzichtet, da das geeignete Vorgehen im Detail kontextabhängig ist und sehr unterschiedlich sein kann. Tabelle 4 fasst die acht Schritte einer MCDA zusammen. Im Folgenden gehen wir detailliert auf die einzelnen Schritte ein.

Tabelle 4 ISPOR-Checkliste zur Durchführung einer MCDA

Schritt	Empfehlung
1 Definition des Entscheidungsproblems	▪ Klare Beschreibung des Entscheidungsproblems entwickeln ▪ Validierung und Bericht des Entscheidungsproblems
2 Auswahl und Strukturierung der relevanten Attribute für die MCDA	▪ Bericht und Begründung der Methodenauswahl zur Identifizierung der Attribute ▪ Bericht und Begründung der Attributdefinitionen ▪ Validierung und Bericht der Attribute und des Wertebaums (Value Tree)
3 Leistung messen	▪ Bericht und Begründung der Quellen zur Messung der Leistung ▪ Validierung und Bericht der Leistungsmatrix
4 Bewertung der Attribute	▪ Bericht und Begründung der Methodenauswahl zur Bewertung ▪ Validierung und Bericht der Wertefunktionen (Value Functions)
5 Gewichtung der Attribute	▪ Bericht und Begründung der Methodenauswahl zur Gewichtung ▪ Validierung und Bericht der Gewichte
6 Zusammenfassung der Einzelnutzen	▪ Bericht und Begründung der Aggregierungsmethode ▪ Validierung und Bericht der Gewichte
7 Umgang mit Unsicherheit	▪ Bericht über die Quellen der Unsicherheit ▪ Bericht und Rechtfertigung der Sensitivitätsanalyse
8 Berichterstattung	▪ Bericht über die MCDA-Methode und die Resultate ▪ Überprüfung der Resultate

Quelle: Marsh et al. (2016), eigene Darstellung Polynomics.

Definition des Entscheidungsproblems

Der erste Schritt in einer MCDA sollte die klare Beschreibung eines Entscheidungsproblems sein. Das Entscheidungsproblem muss auf die Ziele der Entscheidungsträger abgestimmt sein. Soll eine Rangfolge oder eine Bewertung von Alternativen resultieren? Handelt es sich um einen einmaligen Entscheid oder muss das Modell wiederverwendbar sein? Soll das Modell für verschiedene Entscheidungen, Alternativen, Anspruchsgruppen und Beschränkungen (z. B. Budgets) verwendet werden? Zur Definition des Entscheidungsproblems sollten die Entscheidungsträger, Experten und Anspruchsgruppen eingebunden werden. Es kann aus früheren Entscheidungen gelernt werden oder auf Instrumente wie zum Beispiel die «Criteria, Alternative, Stakeholder,

Uncertainty, and Environment (CAUSE)»-Checkliste zurückgegriffen werden (Marsh et al., 2016).

Franco und Montibeller (2011) sehen zwei Schritte, die bei der Problemdefinition durchlaufen werden müssen. Erstens müssen die Anspruchsgruppen definiert werden und ebenso das Mass der Einbindung der verschiedenen Gruppen. Schlüsselakteure sind zwingend in die Problemdefinition einzubinden. Jedoch auch für die anderen Akteure muss ersichtlich sein, dass die MCDA fair und rational ausgestaltet ist. Zweitens müssen die unterschiedlichen Auffassungen des Problems der Akteure identifiziert und daraus eine Problemdefinition abgeleitet werden, die dem Verständnis der Schlüsselakteure entspricht.

Kriterienwahl

Im zweiten Schritt werden die für die Entscheidung relevanten Kriterien definiert. Diese können beispielsweise in einem mehrstufigen Wertebaum (Value Tree) dargestellt werden. Der Wertebaum unterteilt die relevanten Nutzenkomponenten in Subkriterien (Attribute). Das ausgewählte Set an Attributen sollte vier Eigenschaften erfüllen:

- *Vollständigkeit:*
Die Attribute müssen alle relevanten Faktoren der Entscheidung abdecken.
- *Nicht-Redundanz:*
Attribute sollten entfernt werden, wenn sie als nicht notwendig oder nicht wichtig beurteilt werden.
- *Nicht-Überlappung:*
Attribute, die das Gleiche messen, sollten entfernt werden.
- *Unabhängigkeit von Präferenzen:*
Die Bewertung eines Attributes sollte nicht von der Bewertung eines anderen Attributes abhängig sein.

Ist das Set an Attributen ausgewählt, müssen die einzelnen Attribute ausgestaltet werden. Die ISPOR nennt auch dazu einige Eigenschaften, welche die Attribute aufweisen sollten. Insbesondere sollten sie eindeutig, umfassend, direkt, operationalisierbar und verständlich sein.

Leistung messen

Sind die Attribute ausgewählt und ausgestaltet, müssen für die zu bewertenden Alternativen die Ausprägungen der Attribute ermittelt werden. Die Daten können beispielsweise erhoben oder aus bestehenden Studien extrahiert werden. Daten sollen falls möglich mit Methoden, die auf den Prinzipien der evidenzbasierten Medizin beruhen, gesammelt worden sein (z. B. klinische Phase-III-Studien). Datenlücken können mit Expertenmeinungen gefüllt werden.

Bewertung der Attribute

Bei der Bewertung der Attribute (Scoring) geht es darum, zwischen Attributen vergleichbare Wertefunktionen (Value Functions) zu erhalten. Dabei wird für jedes Attribut eine Wertefunktion geschätzt, die zwischen 0 für den schlechtesten Wert und 100 für den besten (erzielbaren) Wert normalisiert werden kann. Das Scoring definiert damit für alle möglichen Ausprägungen der Attribute den entsprechenden Wert auf einer normalisierten Werteskala.

Die verschiedenen Methoden zur Bestimmung der Wertefunktion lassen sich grundsätzlich in kompositionelle und dekompositionelle Methoden unterteilen. Im Gegensatz zu kompositionellen

Methoden, bei welchen die Beurteilungen einzelner Eigenschaften und Ausprägungen erfragt werden, die anschliessend zu einer Gesamtbewertung zusammengesetzt werden, erheben dekompositionelle Verfahren Gesamturteile von den Befragten. Diese werden dann mittels einer statistischen Analyse in die Teilpräferenzen für die zugrundeliegenden Eigenschaften zerlegt. In der wissenschaftlichen Literatur gilt das Discrete-Choice-Experiment als Goldstandard der dekompositionellen Methoden. Bei den kompositionellen Methoden ist die Auswahl an Methoden grösser. Beispielsweise kommen SWING-Weighting, MACBETH oder Direct Rating zum Einsatz, um nur einige zu nennen. Eine Übersicht der wichtigsten Methoden zur Bestimmung der Wertefunktion findet sich in Tabelle 19 im Anhang.

Bei der Methodenauswahl und dem Erhebungsdesign sollte darauf geachtet werden, dass den Befragten klar kommuniziert wird, welche Ausprägungen des Attributs mit den Enden der Werteskala korrespondieren. Diese minimalen und maximalen Referenzwerte beeinflussen die Interpretation der Wertefunktion und der Gewichte. Insbesondere wenn sich Laien (z. B. Patienten oder die Bevölkerung) unter den Befragten befinden, ist für eine konsistente Erhebung der Wertefunktionen darauf zu achten, dass keine kognitive Überforderung stattfindet.

Gewichtung der Attribute

Die Gewichtung der Attribute offenbart ihre relative Bedeutung. Es gilt festzulegen, zu welchem Teil die Nutzenkomponenten zum Gesamtnutzen beitragen. Bei der Auswahl der Erhebungsmethode sollte wiederum darauf geachtet werden, die Befragten kognitiv nicht zu überlasten. Zudem muss eine repräsentative Erhebung die Heterogenität zwischen aber auch innerhalb der Anspruchsgruppen berücksichtigen. Eine Übersicht der wichtigsten Methoden zur Bestimmung der Gewichte findet sich in Tabelle 19 im Anhang. Einige Methoden erlauben auch die gemeinsame Bestimmung von Bewertung und Gewichtung der Attribute. Dazu zählt zum Beispiel das Discrete-Choice-Experiment.

Die Gewichtung kann grundsätzlich mehrstufig für die einzelnen Ebenen des Wertebaums erfolgen oder einstufig (für alle Attribute auf der untersten Ebene gleichzeitig). Ein mehrstufiges Vorgehen bietet sich besonders dann an, wenn es viele Attribute auf der untersten Ebene gibt (Bana e Costa et al., 2012).

Zusammenfassung der Einzelnutzen

In diesem Schritt wird sichergestellt, dass die Wertefunktionen und Gewichte in Übereinstimmung mit den Präferenzen der Anspruchsgruppen zu einem Gesamtnutzen kombiniert werden. Grundsätzlich stehen verschiedene Aggregationsmethoden zur Verfügung, die schnell kompliziert werden (multiplikative Modelle oder multilineare Schätzmodelle). Die additive Nutzenfunktion ist das einfachste und am weitest verbreitete Konzept. Das Konzept ist jedoch wegen der starken zugrundeliegenden Annahmen und Limitationen umstritten (vgl. z. B. Marsh et al., 2018). Der Umgang mit den damit verbundenen Problemen ist eine der zwölf Herausforderungen, die Oliveira et al. (2019) bei der Verwendung von MCDA für HTA identifiziert haben. Das additive Nutzenkonzept trifft zwar starke Annahmen, reduziert jedoch die Komplexität der MCDA entscheidend.

Umgang mit Unsicherheit: Sensitivitätsanalyse

Für jeden der drei vorangegangenen Schritte (Bewertung, Gewichtung, Aggregation) müssen weitreichende Entscheidungen getroffen und die damit jeweils verbundenen Annahmen und Ein-

schränkungen berücksichtigt werden. Die Validierung einer gegebenen Bewertung und Gewichtung ist schwierig, da sie immer subjektive Werturteile enthalten. Ein robustes Ergebnis erhält man deshalb nur, wenn es umfangreichen Sensitivitätsanalysen standhält. So kann sichergestellt werden, dass ein Ergebnis nicht von spezifischen Werturteilen oder Unsicherheit getrieben ist. In der Sensitivitätsanalyse sollten folgende Quellen von Unsicherheit behandelt werden (vgl. Marsh et al., 2016; und Briggs et al., 2012):

- *Strukturelle Unsicherheit*: Design und Eignung des Modells für die Fragestellung
- *Unsicherheit über (klinische) Evidenz*: Qualität der «Fakten»
- *Heterogenität*: Unterschiede in den Modellinputs für Subgruppen von Patienten
- *Stochastische Unsicherheit*: Varianz der Modellinputs
- *Parameterunsicherheit*: Standardfehler der Modellergebnisse

Berichterstattung

Die MCDA ist dazu gedacht, Entscheidungsträger bei einer Entscheidungssituation zu unterstützen. Der Entscheid selbst ist nicht das Resultat der MCDA. Damit das Resultat die Entscheidungsträger effektiv unterstützt, ist es essenziell das MCDA-Modell, deren Annahmen und Sensitivität transparent und verständlich zu dokumentieren. Der Bericht soll die Punkte der ISPOR Checkliste enthalten.

3 MCDA für nichtkurative Krebstherapien

3.1 Problemdefinition

Der Projektauftrag besteht darin, ein Instrument zur standardisierten Nutzenbewertung als Hilfe im Überprüfungsprozess zur Vergütung durch die OKP zu entwerfen. Das Modell soll:

- eine gesamtheitliche Nutzenbetrachtung ermöglichen, d. h. sowohl der Patientennutzen als auch der gesellschaftliche Nutzen sollen im Modell abgebildet werden.
- den Vergleich einer Arzneimittelinnovation mit einem Komparator zulassen (relative Nutzenbewertung).
- für Laien verständlich und transparent sein.
- im Bereich der Onkologie breit einsetzbar und schneller als die bisherigen Instrumente des BAG sein.

Die Nutzenkategorien in Zahlungsbereitschaften oder Frankenbeträge zu übertragen, ist nicht Ziel des Modells. Der Zusatznutzen der Innovation gegenüber dem Komparator soll in einer Kategorisierung dargestellt werden.

Anwendungsgebiet und Modellvarianten

Das Anwendungsgebiet ist die Onkologie. Das MCDA-Modell soll möglichst generisch ausgestaltet sein, sodass es für viele Onkologika anwendbar ist. Es soll wissenschaftlichen Standards entsprechen und einfach verständlich sein. Anwender sollen sowohl die Verantwortlichen bei Pharmaunternehmen als auch beim BAG sein.

Ein Modell ist immer eine Abstraktion der Realität. Ein komplett generisches Modell für alle Onkologika würde der Diversität der bestehenden Behandlungen zu wenig Rechnung tragen. Dies geht aus den bisherigen internen Arbeiten des Auftraggebers sowie aus Expertengesprächen mit medizinischen Onkologen hervor. Bereits etablierte Bewertungsrahmen für Onkologika unterscheiden deshalb verschiedene Modellvarianten je nach Behandlungsziel. In Anlehnung an die «Magnitude of Clinical Benefit Scale» der European Society for Medical Oncology (ESMO) und aufgrund der Diskussionen in den Expertengesprächen schlagen wir drei Modellvarianten vor:

- adjuvante und potenziell kurative Therapien
- nichtkurative Therapien
- Therapien gegen seltene Krankheiten

Adjuvante und kurative Therapien haben ein anderes Ziel als nichtkurative Therapien. Kurative Therapien dürfen beispielsweise ein anderes Sicherheitsprofil aufweisen als nichtkurative Therapien, bei welchen die Verträglichkeit sehr wichtig ist. Bei Arzneimittelinnovationen gegen seltene Krankheiten ist typischerweise die Datenverfügbarkeit und deren Qualität wesentlich schlechter und häufig werden in den klinischen Studien andere Endpunkte als relevant betrachtet. Ob sich hier eine weitere Unterteilung je nach Therapieziel aufdrängt, kann noch nicht abgeschätzt werden.

Im vorliegenden Projekt kann nur für eine Modellvariante ein MCDA-Modell entwickelt werden. Das vorgeschlagene Modell wird an einem Anwendungsbeispiel exemplarisch getestet. Wir verwenden für das Anwendungsbeispiel eine nichtkurative Therapie und fokussieren deshalb im weiteren Projektverlauf auf diese Modellvariante.

3.2 Kriterienwahl: Bestimmung des Wertebaums für nichtkurative Therapien

3.2.1 Vorgehen und Wertebaum

Bei der Bestimmung des Wertebaums sind wir auf zwei Arten vorgegangen. Zum einen haben wir die verschiedenen Nutzenkomponenten definiert, aus denen der Baum bestehen soll (top down). Zum anderen haben wir in der relevanten Literatur nach möglichen Eigenschaften gesucht, mit welchen die Nutzenkomponenten abgebildet werden können (bottom up). Nachdem wir die Resultate dieser zwei Vorgehensweisen zusammengeführt hatten, besprachen wir unseren Vorschlag des Wertebaums mit zwei Experten der Onkologie, zwei Experten der Gesundheitsökonomie und einem Experten der Pharmaindustrie.² Aufgrund deren Rückmeldungen überarbeiteten wir den besprochenen Wertebaum zur finalen Version. In den Gesprächen und der Überarbeitung wurde stets auf die von der ISPOR vorgegebenen Eigenschaften für das finale Set von Attributen geachtet (vgl. Abschnitt 2.2).

Der finale Wertebaum für nichtkurative Therapien berücksichtigt den Patientennutzen und den gesellschaftlichen Nutzen. Diese zwei Nutzenkomponenten stellen die oberste Ebene unseres Wertebaums, die Dimensionen, dar (vgl. Tabelle 5). Die Dimensionen werden weiter in Kriterien unterteilt, die wiederum durch Attribute charakterisiert sind. Die Nutzenkomponenten werden anhand der Attribute auf der untersten Ebene gemessen.

Der Patientennutzen ist in die Kriterien Behandlungserfolg und Sicherheitsprofil gegliedert, wobei der Behandlungserfolg am durchschnittlichen Überleben und der gesundheitsbezogenen Lebensqualität gemessen wird. Während diese zwei Eigenschaften den Nutzen messen, soll das Sicherheitsprofil den Disnutzen der Therapie anhand des Auftretens von unerwünschten Ereignissen und des Patientenkomforts (Patient Convenience) messen. Der gesellschaftliche Nutzen teilt sich auf in Nutzen aufgrund der Krankheitslast sowie Kosten für die Gesellschaft. Der Nutzen (respektive Disnutzen) aus der Krankheitslast wird mit den Disability-Adjusted Life Years (DALY) gemessen. Bei den Kosten berücksichtigen wir neben den indirekten Kosten, d. h. dem Produktivitätsverlust für die Gesellschaft aufgrund der durch die Krankheit und der Behandlung verlorenen Zeit, auch die weiteren Behandlungskosten der Therapie. Diese umfassen beispielsweise die Kosten der Konditionierung des Patienten für die Therapie oder der ärztlichen Betreuung während und nach der Therapie. Sie enthalten aber nicht die Kosten des betrachteten Medikaments an und für sich, da diese für das neue Arzneimittel bei der Bewertung noch nicht bekannt ist. Die übrigen direkten Kosten können sich je nach Arzneimittel stark unterscheiden, da durch eine Änderung der Behandlung Substitutionseffekte bei den anderen Kostenblöcken auftreten können. Deshalb sollten neben den indirekten, gesellschaftlichen Kosten ebenfalls die übrigen direkten Kosten in die Nutzenbewertung für den Vergütungsentscheid einfließen.

² Diese meisten dieser Experten waren auch Teil der Fokusgruppe, die noch breiter aufgestellt war und auch Experten mit langjähriger Behördenerfahrung als Mitglieder hatte. Die Fokusgruppe begleitete den weiteren Verlauf des Projekts. Siehe auch Anhang A: Beteiligte Akteure.

Tabelle 5 Wertebaum für nichtkurative Therapien

Dimension	Kriterium	Attribut
Patientennutzen	Behandlungserfolg	▪ Überleben
		▪ Gesundheitsbezogene Lebensqualität
	Sicherheitsprofil	▪ Unerwünschte Ereignisse
		▪ Patientenkomfort
Gesellschaftlicher Nutzen	Krankheitslast	▪ Disability-adjusted life years (DALY)
	Kosten	▪ Indirekte Kosten ▪ Weitere Behandlungskosten

Quelle: Eigene Darstellung, Polynomics.

In einigen theoretischen Papieren zur Anwendung von MCDA für HTA wird der Innovationsgehalt als Nutzenkomponente einer Arzneimittelinnovation proklamiert (vgl. z. B. Angelis and Kanavos, 2017). Aus unserer Sicht ist der Innovationsgehalt keine Komponente, die in die Nutzenbewertung eines Arzneimittels zur Vergütung durch die OKP einfließen sollte. Stellt ein Arzneimittel tatsächlich eine bahnbrechende Innovation dar, wird sich dies im Patientennutzen und/oder im gesellschaftlichen Nutzen widerspiegeln. Wenn eine Innovation zum Beispiel im Wirkmechanismus keinen Effekt auf den Behandlungserfolg, das Sicherheitsprofil, die Produktivitätsverluste etc. hat, stellt sich die berechtigte Frage, wieso eine soziale Krankenversicherung dies mit einem höheren Preis finanzieren sollte. Auch das Argument, dass mit einer Innovation die Grundlage für weitere Innovationen und Neuentwicklungen geschaffen wird, verfängt nicht. Sind diese nachgelagerten Innovationen nutzenstiftend, sollten diese dann entsprechend vergütet werden. Dadurch bleiben die Anreize bestehen, an vielversprechenden Technologien weiter zu forschen. Das Innovationspotenzial bereits frühzeitig abzuschätzen, bleibt eine Herausforderung. Zudem ist der (potenzielle zukünftige) Effekt der Innovation zum Zeitpunkt des Vergütungsentscheids noch nicht abschätzbar und somit nicht adäquat einzupreisen. Innovation lässt sich an anderen Stellen im System effektiver fördern und würdigen wie zum Beispiel durch die (staatliche) Subvention von Grundlagenforschung oder die Ausschreibung von Forschungspreisen.

Der Wertebaum enthält somit aus unserer Sicht alle relevanten Nutzenkomponenten, um eine gesamtheitliche Nutzenbewertung von Arzneimittelinnovationen zu ermöglichen. Zur besseren Übersichtlichkeit und Komplexitätsreduktion beschränken wir uns auf die wichtigsten Attribute. Das hilft auch dabei, die Anforderungen der ISPOR-Richtlinien zu erfüllen. Eine Herausforderung bei der Ausgestaltung einer MCDA ist die Einhaltung der Anforderung von Nicht-Überlappung der Attribute, d. h. Attribute, die das Gleiche messen, sollen entfernt werden. Wir sehen betreffend Überlappung im vorgeschlagenen Wertebaum keine Probleme. Allerdings sind die Attribute nicht zwingend unabhängig voneinander. Beispielsweise ist es denkbar, dass die Bewertung des Attributes «Überleben» vom Niveau der «gesundheitsbezogenen Lebensqualität» abhängt. Längeres Überleben bei sehr guter Lebensqualität könnte anders bewertet werden als bei schlechter Lebensqualität. Solchen Präferenzabhängigkeiten versuchen wir bei der Ausgestaltung der Eigenschaften Rechnung zu tragen (vgl. Abschnitt 3.2.2).

3.2.2 Ausgestaltung Attribute

In diesem Kapitel beschreiben wir detailliert, wie wir die Attribute des Wertebaums messen wollen. Die Attribute bilden die Nutzenkomponenten ab, sind jedoch nur teilweise direkt messbar.

Demzufolge brauchen wir eine Operationalisierung, die festlegt, wie die Attribute gemessen werden sollen. Bevor wir die Messgrösse bestimmen, führen wir uns das Ziel nochmals vor Augen. Das MCDA-Modell hat zum Ziel, den Nutzen einer Arzneimittelinnovation im Vergleich zu einer bestehenden Therapie (Komparator) zu bewerten. Grundsätzlich gibt es zwei Möglichkeiten einen solchen Vergleich in einem MCDA-Modell umzusetzen:

1. Getrennte Bewertung der Innovation und des Komparators anhand des MCDA-Modells. Anschliessender Vergleich der beiden Gesamtnutzen.
2. Bewertung der Veränderung in den einzelnen Attributen durch die Innovation relativ zum Komparator. Aus dem MCDA-Modell resultiert direkt der Zusatznutzen der Innovation gegenüber dem Komparator.

Beide Möglichkeiten haben ihre Vor- und Nachteile. Bei Möglichkeit (1) vereinfacht sich insbesondere die Ermittlung der Wertefunktionen wesentlich, was schlussendlich auch die Komplexität des Modells für den Endnutzer reduziert. Bei Möglichkeit (2) hingegen könnte es für den Endnutzer einfacher sein, die Ausprägungen der Attribute für seinen Vergleich zu ermitteln (z. B. die Veränderung der indirekten Kosten). Für die weiteren Schritte zur Ausgestaltung des MCDA-Modells birgt Möglichkeit (1) aus unserer Sicht mehr Vorteile. Wir bewerten deshalb die Attribute aller Kriterien für die Innovation und den Komparator getrennt und addieren diese zu einem Gesamtnutzen auf. Aus dem Vergleich des Gesamtnutzens zwischen Innovation und Komparator resultiert der Zusatznutzen der Innovation.

Zur Operationalisierung der Attribute haben wir uns als Erstes auf die Literatur gestützt. Den daraus abgeleiteten Vorschlag besprachen wir mit den gleichen Experten in persönlichen Gesprächen und zusätzlich im Plenum einer Experten-Fokusgruppe, in der die wichtigsten Anspruchsgruppen vertreten waren.³ Im Folgenden zeigen wir die Resultate dieses Prozesses. Die Operationalisierung umfasst die Messgrösse, an welchen die Attribute gemessen werden.

Überleben

Das Attribut Überleben messen wir entweder am Overall Survival (OS) oder am Progression-Free Survival (PFS), je nachdem welches Mass der primäre Endpunkt der verwendeten Studie ist.⁴ Bleibt unklar, welches Mass der primäre Endpunkt ist oder sind mehrere primäre Endpunkte definiert, wird OS gegenüber PFS priorisiert. Gemessen wird zum Zeitpunkt des Medians, d. h. wie lange genau 50 Prozent der Patienten überleben.

Gesundheitsbezogene Lebensqualität

Für die Messung gesundheitsbezogener Lebensqualität verwenden wir Patient-Reported Outcome Measures (PROM). Es gibt eine Vielzahl von PROM, die unterschiedlich umfassend und verbreitet sind. Wir berücksichtigen folgende PROM und priorisieren diese nach der genannten Reihenfolge: SF-36, SF-12, EQ-5D, QLQ-C30 mit seinen krebsartenspezifischen Modulen (z. B. BR23). Sind in einem Anwendungsfall mehrere PROM verfügbar, wird gemäss obenstehender Liste priorisiert. Es können auch alternative Masse verwendet werden. Die Qualität und die Zweckmässigkeit müssen jedoch von einem Experten bestätigt sein. Wichtig ist, dass die Masse mit Werten zwischen 0 und 100 gemessen werden oder auf diesen Bereich übertragen werden können.

³ Die Zusammensetzung der Fokusgruppe ist in Anhang A dargelegt.

⁴ Wenn die verglichenen Behandlungen unterschiedliche primäre Endpunkte aufweisen, wird über die jeweils im Modell hinterlegte Bewertungsfunktion die Vergleichbarkeit hergestellt.

Das Attribut soll die langfristige Lebensqualität messen. Wir schlagen als Kennzahl deshalb die durchschnittliche Lebensqualität innerhalb des Zeitraums halbem Medianüberleben und Medianüberleben vor. Bei einem Medianüberleben von beispielsweise 22 Monaten wird die durchschnittliche Lebensqualität der Monate 11 bis 22 ermittelt. Die Messgrösse zur Bestimmung des Medianüberlebens ergibt sich aus dem Attribut «Überleben», d. h. entweder OS oder PFS.

Um die Resultate der verschiedenen PROM vergleichbar zu machen, werden die damit verbundenen Nutzenwerte, die sogenannten Utilities, verwendet. Dabei werden die Resultate auf eine normierte Kardinalskala mit der minimalen Ausprägung 0 (für Tod) und der maximalen Ausprägung 100 (für perfekte Gesundheit) übertragen. Trotz dieser Normierung können sich der Erwartungswert und die Verteilung der Werte zwischen den PROM nach wie vor unterscheiden. In der Literatur finden sich entsprechende Umrechnungsfunktionen, welche die Resultate verschiedener PROM vergleichbar machen (vgl. z. B. Richardson et al., 2015).

Unerwünschte Ereignisse

Der Nutzen durch unerwünschte Ereignisse wird anhand des Anteils der Patienten, bei welchen mindestens ein unerwünschtes Ereignis der Stufe 3, 4 oder 5 auftrat, gemessen. Unerwünschte Ereignisse der Stufe 1 und 2 werden nicht berücksichtigt, da sie kaum klinische Relevanz aufweisen respektive nur minimale, meist lokale Interventionen zur Folge haben. Der Patientennutzen sollte in der Regel durch solche Ereignisse nicht betroffen sein.

Patientenkomfort

Patientenkomfort wird durch die zwei Messgrössen Ort und Häufigkeit der Verabreichung gemessen. Der Ort der Verabreichung kann die Ausprägungen «zu Hause» oder «ambulant» annehmen. Die Häufigkeit der Verabreichung wird auf einer kontinuierlichen Zeitskala gemessen (z. B. täglich bis halbjährlich).

Patientenkomfort hat viele weitere Ausprägungen, wie zum Beispiel Verabreichungsart (oral, intravenös etc.), Dauer der Verabreichung und Behandlung etc. Wir sind der Meinung, dass sich eine relevante Veränderung in den übrigen Eigenschaften auch in den zwei gewählten Messgrössen zeigt und diese im Bereich der nichtkurativen Krebstherapie besondere Relevanz aufweisen.

Krankheitslast (DALY)

Die Krankheitslast misst die Belastung der Gesellschaft durch eine Krankheit. Sie wird in Anzahl verlorener Lebensjahre gemessen. Diese setzen sich zusammen aus den verlorenen Lebensjahren bei frühzeitigem Versterben und der mit der Krankheit, sprich mit reduzierter Lebensqualität, gelebten Lebensjahre. Die Disability-Adjusted Life Years (DALY) sind ein weit verbreitetes Mass, welches diese zwei Komponenten enthält. Eine hohe Anzahl DALY kann auf eine hohe Mortalität oder eine hohe Prävalenz (Häufigkeit) einer Krankheit zurückzuführen sein. Als Messgrösse verwenden wir den Anteil DALY einer Indikation relativ zur Gesamtanzahl krebsbedingter DALY und der totalen Anzahl DALY in der Schweiz. Die relative Messgrösse stellt die Relevanz der Indikation für die Gesellschaft klarer dar als die absolute Anzahl DALY.

Die DALY für die Schweiz sind auf der Webseite der Global Burden of Disease (GBD) verfügbar. Es fließen die Anzahl DALY der Indikation in das Modell ein. Sind für die Indikation keine DALY vorhanden, werden als zweite Priorität die «Disability Weights» von GBD verwendet.⁵

⁵ Diese sind verfügbar unter <https://ghdx.healthdata.org/record/ihme-data/gbd-2019-disability-weights>, zuletzt aufgerufen am 31.3.2023.

Diese sind für eine grössere Anzahl und präzisere Indikationen verfügbar als die DALY. Um die Anzahl DALY zu erhalten, werden die «Disability Weights» mit der Prävalenz der Indikation multipliziert sowie die Anzahl der Todesfälle, multipliziert mit der mittleren verbleibenden Lebenserwartung, addiert. Die Prävalenz kann im besten Fall datenbasiert ermittelt werden. Alternativ muss eine Abschätzung durch Experten vorgenommen werden. Die aktuelle Anzahl der DALY für die Schweiz und für Krebs sind ebenfalls der Webseite der GBD zu entnehmen.⁶ Anhand der ermittelten Daten ist der Anteil DALY der Indikation an den totalen DALY von Krebs und der Schweiz zu berechnen.

Die Ausprägung dieses Attributs verändert sich nicht zwischen Komparator und Innovation, solange beide Arzneimittel für die gleiche Indikation zugelassen sind. Das Attribut soll die allgemeine Krankheitslast für die Gesellschaft messen, d. h. wie viele Personen sind von der Krankheit betroffen oder wie hoch sind die Einschränkungen in der allgemeinen Lebensqualität. Auch wenn die Innovation die DALYs nicht verändert, ist die allgemeine Krankheitslast einer Indikation für die Nutzenbewertung einer Innovation äusserst relevant. Der Effekt betrifft nicht die relative Nutzenbewertung, die absolute Nutzenbewertung ist jedoch davon abhängig.

Indirekte Kosten

Die indirekten Kosten einer Krankheit und ihrer Behandlung setzen sich zusammen aus Produktivitätsverlusten durch Präsentismus und Absentismus:

- Präsentismus misst die Produktivitätsverluste aufgrund der krankheitsbedingt verringerten Leistungsfähigkeit eines Arbeitnehmers.
- Absentismus misst die Produktivitätsverluste aufgrund Abwesenheit vom Arbeitsplatz. Die Abwesenheit kann durch Krankheit (Morbidität), vorzeitigen Tod (Mortalität) oder der Pflege einer erkrankten Person (informelle Pflege) begründet sein.

Die indirekten Kosten werden nach dem Friktionskostenansatz bewertet. Dahinter steht die Annahme, dass am Arbeitsmarkt keine Vollbeschäftigung herrscht. Es wird angenommen, dass die erkrankte Person nach einer gewissen Zeit durch einen neuen (bisher arbeitslosen) Arbeitnehmer ersetzt werden kann. Da das Attribut für sich allein steht und nicht zum Beispiel gegen direkte Arzneimittelkosten verrechnet werden muss, verzichten wir auf eine Monetarisierung. Wir weisen die indirekten Kosten in Anzahl verlorene Arbeitsmonate pro Patienten über die gesamte Behandlungsdauer aus. In der Realität werden sehr selten Informationen zu indirekten Kosten verfügbar sein. Hier muss eine Einschätzung durch einen Experten (der Gesundheitsökonomie) erfolgen.

Falls in der Literatur die indirekten Kosten in Geldeinheiten ausgewiesen sind, muss zur Umrechnung in die Anzahl verlorener Arbeitsmonate der aktuelle Medianlohn der Lohnstrukturerhebung des Bundesamts für Statistik verwendet werden.⁷ Zur Umrechnung unbezahlter Arbeit, zum Beispiel informeller Pflege sind die vom Bundesamt für Statistik geschätzten Arbeitskosten der unbezahlten Arbeit zu verwenden.⁸ Falls die indirekten Kosten in einer anderen Zeiteinheit ausgewiesen sind, ist für die Umrechnung von 42 Arbeitsstunden pro Woche, verteilt über fünf Tage und vier Arbeitswochen pro Monat auszugehen.

⁶ Verfügbar unter <https://vizhub.healthdata.org/gbd-results/>, zuletzt aufgerufen am 31.03.2023. Suche eingrenzbar auf «Neoplasms» und «Switzerland».

⁷ Verfügbar unter <https://www.bfs.admin.ch/bfs/de/home/statistiken/arbeit-erwerb/loehne-erwerbseinkommen-arbeitskosten/lohniveau-schweiz.html>, zuletzt aufgerufen am 23.08.2021.

⁸ Verfügbar unter <https://www.bfs.admin.ch/bfs/de/home/statistiken/arbeit-erwerb/unbezahlte-arbeit.gnpdetail.2021-0410.html>, zuletzt aufgerufen am 30.08.2021.

Weitere Behandlungskosten

Neben den Arzneimittelkosten der betrachteten Krebstherapie fallen weitere Behandlungskosten, zum Beispiel aufgrund der ärztlichen Betreuung, der Behandlung unerwünschter Ereignisse oder Nebenwirkungen an. Berücksichtigt werden diese zusätzlichen Kosten als durchschnittliche Kosten pro Patienten über die gesamte Behandlungsdauer. Die weiteren Behandlungskosten, die von der OKP übernommen werden, setzen sich zusammen aus:

- andere medikamentöse Kosten
- Tarmed-Kosten (ärztliche und technische Leistungen)
- Kosten für Laboranalysen (Analyseliste)
- Kosten für stationäre Aufenthalte (DRG, ST Reha etc.)
- andere Kosten (Mittel- und Gegenstände, Paramedizin etc.)

Die Kosten können entweder anhand der Preise in den Tarifwerken und einer Erhebung der notwendigen Leistungen der oben genannten Kostenarten berechnet oder durch Experten abgeschätzt werden. Die Kosten zu ermitteln, ist aufwändig und zeitintensiv. Es empfiehlt sich die Einschätzung durch Experten. Dies kann durch Experten aus der Gesundheitsökonomie und/oder der Medizin geleistet werden.

Tabelle 6 Übersicht zu den Attributen

Kriterium		Attribut	Messgrösse
Patientennutzen	Behandlungserfolg	▪ Überleben	▪ Median OS oder PFS
		▪ Gesundheitsbezogene Lebensqualität	▪ SF-36, SF-12, EQ-5D, QLQ-C30 (priorisiert) ▪ Alternativen durch Experten zu prüfen
	Sicherheitsprofil	▪ Unerwünschte Ereignisse	▪ Anteil Patienten mit mind. einem unerwünschten Ereignis von Grad 3, 4 oder 5
		▪ Patientenkomfort	▪ Ort der Verabreichung ▪ Häufigkeit der Verabreichung
Gesellschaftlicher Nutzen	Krankheitslast und öffentliche Gesundheit	▪ Disability-Adjusted Life Years (DALY)	▪ DALY oder (Disability Weights * Prävalenz Indikation) ▪ Kennzahl: Anteil an totaler Anzahl DALY Krebs und Schweiz
	Kosten	▪ Indirekte Kosten	▪ Anzahl verlorene Arbeitsmonate (Patient und Umfeld) für gesamten Behandlungszeitraum ▪ Berechnung oder Experteneinschätzung
		▪ Weitere Behandlungskosten	▪ Kosten für typische begleitende Leistungen zur betrachteten Arzneimitteltherapie (pro Patient, gesamter Behandlungszeitraum) ▪ Berechnung oder Experteneinschätzung

Quelle: Eigene Darstellung, Polynomics.

3.3 Leistung messen

Schritt drei der Richtlinien beschreibt eigentlich die Bestimmung der Ausprägungen der ausgewählten Messgrößen für die zu bewertenden Alternativen. In unserem Fall entwickeln wir ein generisches Modell; wir kennen folglich noch nicht alle Anwendungen des Modells und können

deshalb die genaue Ausprägung der Attribute nicht von vornherein bestimmen. Nach der Festlegung unseres generischen MCDA-Modells testen wir dieses in Kapitel 4 an einem Anwendungsfall.

Deshalb definieren wir hier, welche Quellen als Input für das MCDA-Modell geeignet sind. Das MCDA-Modell soll einen Vergleich von zwei Arzneimitteln ermöglichen. Einzige Voraussetzung ist, dass für beide Alternativen die in Abschnitt 3.2.2 definierten Messgrößen vorhanden sind, oder, wo erlaubt, durch einen Experten abgeschätzt werden können. Folgende Grundsätze sind bei der Datenermittlung zu berücksichtigen:

- Daten aus begutachteten Artikeln (peer reviewed) sind vorzuziehen.
- Neuere Daten sind gegenüber älteren zu bevorzugen.
- Individuelle Einschätzungen von Experten sind nur dort einzusetzen, wo es das Modell vorsieht und wenn keine zuverlässigen und vertrauenswürdigen Daten bestehen. Es gibt zwei Arten:
 - Einsatz von Experten zur Bewertung von imperfekten Daten. Dazu zählt z. B. die Frage der Übertragbarkeit von Ergebnissen oder Daten aus dem Ausland, das Abwägen zwischen alternativen Datenquellen, jeweils sofern dies nicht offensichtlich ist. Diese Art der Expertenbefragung sollte relativ unproblematisch sein. Gegebenenfalls sind mehrere Einschätzungen abzuholen und entsprechende Sensitivitätsanalysen durchzuführen.
 - Expertenmeinungen: Hier handelt es sich um die Einschätzung von Experten, die nicht auf objektiver Basis getroffen wird. In der Regel wird darauf zurückgegriffen, wenn keine andere Quelle verfügbar ist. Diese Art der Expertenbefragung ist mit hoher Unsicherheit verbunden, weshalb idealerweise mehrere Experten zu befragen sind und das Spektrum der Antworten durch eine Sensitivitätsanalyse abzufangen ist.
- Die Daten der verglichenen Arzneimittel können aus unterschiedlichen Quellen stammen. Die Datenqualität und -herkunft ist jedoch zwingend transparent darzulegen und gegebenenfalls einer Sensitivitätsanalyse zu unterziehen.

Es kann bei allen Attributen vorkommen, dass die notwendigen Daten nicht vorhanden sind. Dies kann verschiedene Dimensionen betreffen, zum Beispiel liegt die Medianüberlebenszeit noch nicht vor, die Lebensqualität ist nicht in einer der akzeptierten PROM gemessen oder die Daten fehlen komplett. Im vornherein alle Möglichkeiten vorherzusehen und eine allgemeingültige Lösung vorzuschlagen, ist nicht möglich. Wir schlagen vor, bei fehlenden Daten in erster Priorität diese bei den Herstellern einzufordern. Ist die Nachlieferung nicht verhältnismässig oder schlichtweg nicht möglich (z. B. Medianüberlebenszeit noch nicht erreicht) sollte ein pragmatisches Vorgehen gewählt werden, um die Datenlücke schliessen zu können. Die dabei getroffenen Annahmen sollten transparent kommuniziert und in einer Sensitivitätsanalyse untersucht werden. Beobachtet man über die Zeit immer wieder die gleichen Datenlücken, muss das Modell im Sinne eines lernenden Systems angepasst und diese Datenlücken bei der Ausgestaltung zum Beispiel der Attribute berücksichtigt werden.

3.4 Bewertung der Attribute

In Schritt vier der Richtlinien geht es darum, zwischen den Attributen vergleichbare Wertefunktionen (Value Functions) zu erhalten. Es gilt, eine Wertefunktion zu schätzen, die zwischen 0 für den niedrigsten Wert und 100 für den höchsten (erzielbaren) Wert normalisiert werden kann. In diesem Schritt definieren wir für alle möglichen Ausprägungen der Messgrößen den Wert auf

der Nutzenskala für unseren generischen Wertebaum bei nichtkurativen Krebstherapien. Die Wertefunktionen bestimmen wir über eine Präferenzbefragung. Gemäss Marsh et al. (2016) muss in dieser Befragung den Befragten klar kommuniziert werden, welche Ausprägungen des Attributs mit den Enden der Werteskala korrespondieren. Diese minimalen und maximalen Referenzwerte beeinflussen die Interpretation der Wertefunktion und der Gewichte.

3.4.1 Bestimmung der Minima und Maxima

In einem ersten Schritt definieren wir die möglichen Spannweiten, d. h. die minimalen und maximalen Ausprägungen jedes Attributs für unseren Wertebaum im nichtkurativen Setting. Wichtig ist hier, dass wir unter Minimum und Maximum nicht unbedingt die logischen Grenzen der Messgrösse meinen, sondern die Werte, ab (bzw. bis zu) welchen keine grossen Nutzenwertänderungen mehr zu erwarten sind. Für den Fall einer nichtkurativ behandelten schweren Krebserkrankung, für die die Medianüberlebenszeit beispielsweise wenige Jahre ist, sollte der höchste Nutzenwert (100) deutlich unter der potenziell denkbaren Überlebenszeit von (vielleicht) 50 Jahren liegen: Man würde bereits eine Überlebenszeit mit neuer Therapie von 5 Jahren als Durchbruch bezeichnen. Hier ergibt sich ein gewisser Handlungsspielraum, der Fingerspitzengefühl bei der Wahl der Minima und Maxima erfordert. Für Werte jenseits der gesetzten Werte verwenden wir weiterhin 0 und 100 für das Nutzenminimum und –maximum des entsprechenden Attributs.⁹

Bei manchen Attributen stehen mehrere Messgrössen zur Auswahl. Dann muss sichergestellt werden, dass erstens die Gewichtung nicht von der ausgewählten Messgrösse abhängt und zweitens der minimale und maximale Nutzenwert jeweils identisch ist.¹⁰ Der zweite Punkt verlangt, dass vor der Bestimmung der zugehörigen Wertefunktionen nach einem kohärenten Set an Minima (Maxima) für alle Messgrössen gesucht werden muss, für die der Nutzenwert gleich null (100) gesetzt wird. Es ist dabei darauf zu achten, dass Werte jenseits der gewählten Minima und Maxima in der Regel nicht vorkommen.

Die minimalen und maximalen Ausprägungen jedes Attributs bestimmen wir mehrstufig:

1. Bestimmung der logischen Minima und Maxima
2. Plausibilisierung der logischen Minima und Maxima und Bestimmung der (für den Wertebaum) relevanten Minima und Maxima in der Fokusgruppe
3. Validierung anhand Daten (wo vorhanden)

Eine Übersicht bietet Tabelle 7 am Ende des Abschnitts.

Überleben

Aus der Diskussion in der Fokusgruppe ging hervor, dass die minimal akzeptierte Medianüberlebenszeit einer nichtkurativen Krebsbehandlung bei ca. 3 Monaten liegt. In diesem Setting nehmen wir zudem an, dass das Maximum bei 5 Jahren erreicht wird, d. h. ab diesem Wert sind keine grossen Nutzenwertänderungen zu erwarten. Wir messen die beobachtete Medianüberlebenszeit in Monaten, sodass der Wert zwischen 3 und 60 variiert.

⁹ Die Wertefunktionen und Gewichte werden nur zwischen Minimum und Maximum bestimmt. Eine Extrapolation ist zwar technisch leicht umzusetzen, ist jedoch auch mit Annahmen verbunden. So ändert sich beispielsweise das theoretische Minimum und Maximum des Gesamtmodells, das dann nicht mehr klar zwischen 0 und 100 Punkten liegt. In Extremfällen ist eine Sensitivitätsanalyse, die eine solche Extrapolation umsetzt, zu prüfen.

¹⁰ Wir haben in unserem Anwendungsbeispiel nur eine Messgrösse für die Bevölkerungsbefragung verwendet. Ob unsere Gewichtung sich verändern würde, wenn wir statt OS PFS verwenden, müsste näher untersucht werden.

Das Minimum und Maximum könnte auch aus der in der Schweiz beobachteten Verteilung der Überlebenszeiten aus dem nationalen Krebsregister (NICER) ermittelt werden. Dazu müssen zuerst die im nichtkurativen Setting relevanten Krebsarten identifiziert und danach die minimalen und maximalen Medianüberlebenszeiten von NICER bestellt werden. Dies war in diesem Projekt u. a. aus Zeitgründen nicht möglich.

Gesundheitsbezogene Lebensqualität

Das logische Minimum (0) und Maximum (100) ist bekannt. Relevant sind jedoch das realistische Minimum und Maximum in unserem nichtkurativen Setting mit Krebspatienten. Für die Schweiz sind Verteilungsmasse in der Bevölkerung leider nicht näher untersucht. In Deutschland beträgt der durchschnittlich angegebene Wert für die Lebensqualität 90.2.¹¹ Mehr als die Hälfte der deutschen Bevölkerung gibt an, keine nennenswerten Einschränkungen zu haben (Wert 100). Jede vierte Person hat jedoch einen Wert unter 78 (Szende et al., 2014).

Diese Werte beziehen sich auf die gesamte Bevölkerung und nicht auf Personen, die an einem nicht heilbaren Krebs erkrankt sind. Für unser Modell nehmen wir an, dass der minimale Wert bei 10 (gravierende Einschränkung) und der maximale Wert bei 80 (leichte Einschränkung) liegt.

Unerwünschte Ereignisse

Der Anteil unerwünschter Ereignisse erzielt den höchsten Nutzenwert bei 0 Prozent unerwünschter Ereignisse. Nach Einschätzung der in der Fokusgruppe konsultierten Experten ist das Minimum des Nutzenwertes auch im nichtkurativen Setting bei soliden Tumoren bei einem Anteil von 40 Prozent der Patienten mit schweren unerwünschten Ereignissen (Grad 3, 4 und theoretisch 5) erreicht. Bei nichtsoliden Tumoren kann dieser Anteil jedoch um einiges höher liegen. Die Konsultation der Literatur zeigt, dass in Realität oft höhere Werte als 40 Prozent erreicht werden. Wir legen das Minimum bei 5 und das Maximum bei 80 Prozent fest.

Patientenkomfort

Hierfür verwenden wir zwei Messgrößen: den Behandlungsort und die Behandlungshäufigkeit. Für den Behandlungsort definieren wir die zwei Ausprägungen «zu Hause» und «ambulant». Ambulant heisst, dass der Patient für die Behandlung eine Arztpraxis oder eine ambulante Station eines Spitals aufsuchen muss. Nach der Behandlung kann der Patient wieder nach Hause gehen, es ist kein Aufenthalt über Nacht notwendig. Wir definieren den Behandlungsort als Indikatorvariable, wobei «ambulant» den Wert 0 erhält und «zu Hause» den Wert 1.

Die Behandlungshäufigkeit messen wir in Tagen. Das Minimum der Häufigkeit der Verabreichung legen wir bei «täglich» (1), das Maximum bei «monatlich» (30) fest.

Krankheitslast (DALY)

Rund 18 Prozent der Krankheitslast in der Schweiz war 2019 auf Krebs zurückzuführen. Für eine mögliche Indikation einer nichtkurativen Krebserkrankung definieren wir das Minimum bei 1 Prozent der gesamten krebsbedingten Krankheitslast, was 0.18 Prozent der gesamten Krankheitslast in der Schweiz entsprechen. Das sind 46 DALY pro 100'000 Einwohner pro Jahr. Das Maximum legen wir bei 20 Prozent der krebsbedingten, respektive 3.6 Prozent der gesamten

¹¹ EQ-5D index value mit europäischem VAS value set, vgl. Janssen et al. (2019). Die Daten gehen zurück auf König et al. (2009). Neuere Daten scheinen nicht verfügbar.

Krankheitslast in der Schweiz fest. Das Maximum entspricht 920 DALY pro 100'000 Einwohner pro Jahr.

Indirekte Kosten

Die indirekten Kosten sind wie in Abschnitt 3.2.2 beschrieben mit dem Friktionskostenansatz zu bewerten. In Anlehnung an die internationale Literatur verwenden wir eine Friktionsperiode von sechs Monaten (vgl. Kigozi et al., 2016; Pike and Grosse, 2018). Dieser Wert entspricht dem Maximum dieser Messgrösse und findet auch Anwendung, wenn erkrankte Personen länger vom Arbeitsplatz fernbleiben. Daraus ergibt sich ein Maximum von sechs verlorenen Arbeitsmonaten. Das kontextrelevante Minimum setzen wir bei einem Arbeitsmonat fest.

Weitere Behandlungskosten

Das logische Minimum der weiteren Behandlungskosten ist bei null. Dies entspricht jedoch nicht dem realistischen (durchschnittlichen) Minimum für Patienten mit einer nichtkurativ behandelbaren Krebserkrankung. Neben der Krebstherapie nehmen diese Personen im Durchschnitt zusätzlich Gesundheitsdienstleistungen (z. B. beim Hausarzt) in Anspruch. Von Wyl et al. (2015) weisen in ihrer Untersuchung zu Kosten am Lebensende in der Schweiz kumulierte Gesundheitskosten zwischen rund 12'000 CHF und 85'000 CHF aus. Der grösste Kostenanteil entfällt auf Hospitalisierungen und Pflege, Medikamente machen nur einen geringen Kostenanteil aus. In Anlehnung an diese Studie setzen wir das Minimum für weitere Behandlungskosten (ohne die Kosten der Arzneimittelinnovation) bei 10'000 CHF und das Maximum bei 80'000 CHF pro Patienten über die gesamte Behandlungsdauer fest.

Tabelle 7 Minima und Maxima aller Attribute

Attribut	Einheit	Min	Max
1 Überleben (Median OS oder PFS)	Monate	3 (OS)	60 (OS)
2 Gesundheitsbezogene Lebensqualität	PROM utility score (0-100)	10	80
3 Unerwünschte Ereignisse	Prozent der Patienten	5	80
4 Behandlungshäufigkeit	Alle ... Tage	1	30
5 Behandlungsort	ambulant (0), zu Hause (1)	0	1
6 Krankheitslast in der Schweiz	DALY/100'000 Einwohner/Jahr	46	920
7 Indirekte Kosten (Arbeitsausfall)	Monate	1	6
8 Weitere Behandlungskosten	1'000 CHF	10	80

Quelle: Eigene Darstellung, Polynomics.

3.4.2 Lineare und nichtlineare Wertefunktionen

Nach der Festlegung von Minima und Maxima muss bestimmt werden, wie die Wertefunktion zwischen diesen zwei Referenzwerten verläuft. Die einfachste Möglichkeit ist die lineare Interpolation zwischen der minimalen und maximalen Ausprägung (vgl. Abbildung 1, linke Grafik). Je nach Vorzeichen ist das eine steigende oder fallende Gerade. Dem Minimum dieser Geraden wird der Wert 0 zugewiesen, dem Maximum der Wert 100. Die lineare Interpolation ist zulässig, wenn eine Veränderung in der Ausprägung eines Attributs unabhängig von dessen Ausgangswert immer zur selben Veränderung im Nutzen führt.

Ist der Nutzenzuwachs jedoch vom Ausgangswert abhängig, d. h., wenn beispielsweise eine Veränderung um eine Einheit bei einem hohen Ausgangswert zu einem kleineren Nutzenzuwachs als bei einem niedrigen Ausgangswert führt, muss eine nichtlineare Wertefunktion ermittelt werden (vgl. Abbildung 1, rechte Grafik). Dafür schlagen wir vor, neben Minimum und Maximum eine dritte Ausprägung bewerten zu lassen, sodass die Wertefunktion nun durch drei Koordinaten definiert ist. Anschliessend bestimmen wir eine mathematische Funktion, welche die Werte in dieses drei Punkten am besten wiedergibt (z. B. eine quadratische Funktion). Diese Funktion ist dann die nichtlineare Wertefunktion des entsprechenden Attributs. Wieder wird im relevanten Bereich der Attributausprägung dem niedrigsten Punkt der Kurve der Wert 0 und dem höchsten der Wert 100 zugewiesen. Es ist zu beachten, dass diese Extremwerte nicht mehr zwangsläufig beim Minimum oder Maximum der Ausprägung anfallen: Falls zum Beispiel eine Parabel ihren Wendepunkt innerhalb des betrachteten Wertebereichs hat, ist das Minimum bzw. Maximum entsprechend anzupassen.

Abbildung 1 Beispielhafte Darstellung von zwei Wertefunktionen

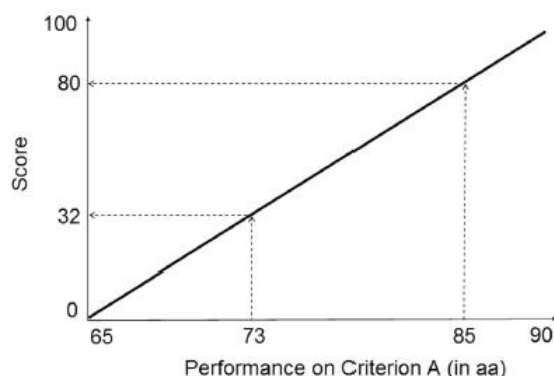


Fig. A1 – Linear partial value function for criterion A.

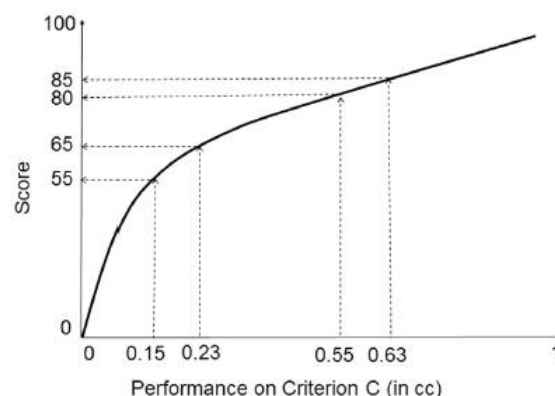


Fig. A3 – Nonlinear partial value function for criterion C.

Quelle: Marsh et al. (2016).

Ob eine Wertefunktionen eine lineare oder nichtlineare Form aufweist, kann entweder aufgrund theoretischer Überlegungen definiert oder in einer empirischen Präferenzenerhebung ermittelt werden. Dabei ist abzuwägen, wie einschränkend die A-priori-Definition der Form im Vergleich zur Komplexität einer entsprechenden Befragung ist. Um die Komplexität unseres Modells übersichtlich zu behalten, legen wir a priori fest, welche Attribute eine lineare respektive nichtlineare Form aufweisen. Die effektive Form der nichtlinearen Attribute ermitteln wir empirisch. Die Wertefunktion der linearen Attribute entspricht der linearen Interpolation zwischen der minimalen und maximalen Ausprägung (vgl. Abschnitt 3.4.1). Tabelle 8 weist jedem Attribut die Form seiner Wertefunktion zu. Die dargestellte Zuteilung wurde von den Experten der Fokusgruppe bestätigt.

Tabelle 8 Annahmen zur Form der Wertefunktionen

Attribute mit linearer Wertefunktion	Attribute mit nichtlinearer Wertefunktion
▪ Unerwünschte Ereignisse ▪ Patientenkomfort ▪ Indirekte Kosten ▪ Weitere Behandlungskosten	▪ Überleben ▪ Gesundheitsbezogene Lebensqualität ▪ Krankheitslast (DALY)

Quelle: Eigene Darstellung, Polynomics.

3.4.3 Massgebende Präferenzen für die Wertefunktionen

Welchen Nutzenwert eine spezifische Ausprägung eines Attributs schlussendlich aufweist, ist von den Präferenzen der befragten Person abhängig. Krebspatienten könnten beispielsweise niedrigen Ausprägungen des Attributs «Überleben» höhere Nutzenwerte zuweisen, als dies gesunde Versicherte tun. Damit aus dem Projekt ein transparenter und konsistenter Entscheidungsprozess resultiert, ist es essenziell, die für die Entscheidungssituation massgebenden Präferenzen zu erfassen.

Die OKP ist eine Sozialversicherung, welche die Solidarität, Qualität und Wirtschaftlichkeit der medizinischen Versorgung sicherstellen soll. In der Volksabstimmung vom 4. Dezember 1994 wurde das KVG und somit die Einführung der OKP vom Schweizer Stimmvolk angenommen. Die Ziele und die Ausgestaltung des KVG entsprechen demzufolge den Präferenzen der Mehrheit der Schweizer Stimmberechtigten. Deutliche Veränderungen der (durchschnittlichen) Präferenzen sollten zumindest mittelfristig zu Änderungen in der Gesetzgebung und damit auch zu Änderungen im KVG führen. Entsprechend sollten in unserem Modell die Präferenzen der Schweizer Stimmberechtigten abgebildet werden.

Es ist davon auszugehen, dass innerhalb der Bevölkerung unterschiedliche Präferenzen betreffend die Ziele des KVG bestehen, zum Beispiel zwischen Patienten oder Prämienzahlern, oder zwischen Experten wie Ärzten oder Patientenvertretern. Grundsätzlich sollte die Abhängigkeit von den Präferenzen über Sensitivitätsanalysen sichtbar gemacht werden können. Die Experten der Fokusgruppe haben bestätigt, dass für eine valide Umsetzung des MCDA-Modells eine repräsentative Bevölkerungsbefragung notwendig ist.

3.4.4 Methodenwahl

Die Erhebungsmethoden zur Bestimmung der Werte bzw. Wertefunktionen lassen sich grob in kompositionelle und dekompositionelle Verfahren einteilen. Kompositionelle Methoden bestimmen Wert und Gewicht für jedes Attribut separat und bilden anschliessend den Gesamtwert durch eine Aggregation der Einzelwerte. Dekompositionelle Methoden betrachten auf der anderen Seite den Gesamtwert der Alternativen als Ganzes. Aus diesem lassen sich Gewichte und Werte für die Attribute gemeinsam ableiten (vgl. für eine Besprechung der beiden Verfahren mit Beispielen und Literaturhinweisen Thokala et al., 2016).

Bisherige Anwendungen von MCDA für HTA haben die Schritte der Bewertung und Gewichtung in Expertengremien vorgenommen. Entsprechend wurden ausschliesslich kompositionelle Methoden mit Konsensbildung im Plenum angewandt, da dekompositionelle Methoden aufgrund der kleinen Stichprobe an ihre Grenzen stossen. Unsere Fokusgruppe war einstimmig der Meinung, dass die Präferenzen von Experten nicht ausreichen, sondern dass die Präferenzen der Bevölkerung erhoben werden müssen. Mit der damit verbundenen grösseren Stichprobe werden dekompositionelle Methoden attraktiv, weil mit diesen aufgrund der gemeinsamen Bestimmung von Werten und Gewichten das Problem der Aggregation von Einzelwerten reduziert werden kann. Eine der mächtigsten dekompositionellen Methoden sind die Discrete-Choice-Experimente (DCE), die wir auch für die Ermittlung der Werte und Gewichte der Attribute bei der Bevölkerung empfehlen.

Bei einem DCE steht ein Individuum vor der Entscheidung, zwischen alternativen Gütern und Leistungen zu wählen, die sich in ihren Eigenschaften unterscheiden. DCE eignen sich dabei besonders auch zur Bewertung von nicht handelbaren Gütern, wie sie im Gesundheitswesen häufig

anzutreffen sind (z. B. Lebensdauer, Lebensqualität). Diese Methode der Präferenzmessung beruht auf einer von Lancaster (1966) formulierten Nachfragetheorie sowie auf Ansätzen der Verhaltenspsychologie (Luce and Tukey, 1964). Im Gegensatz zur traditionellen Konsumtheorie setzt sich der Nutzen, den ein Individuum aus einem Gut oder einer Leistung zieht, aus den Teilnutzenwerten der einzelnen Eigenschaften (Attribute) zusammen. So können zwei Individuen aus einem Gut einen unterschiedlichen Nutzen ziehen, weil sie dessen Attribute unterschiedlich gewichten.

Im Rahmen eines DCE werden die Individuen vor verschiedene hypothetische Entscheidungssituationen gestellt. In diesen können jeweils Alternativen miteinander verglichen werden. Es wird unterstellt, dass die Befragten die Alternative wählen, die den höchsten Nutzen stiftet. Der statistischen Auswertung liegt ein entscheidungstheoretisches Modell zugrunde (McFadden, 2001). Diese Verankerung in der ökonomischen Theorie hat den Vorteil, dass ein Fundament zur Verfügung steht, das sich in der Vergangenheit auch in empirischen Anwendungen bewährt hat. Aus den getroffenen Entscheidungen lässt sich anschliessend mit geeigneten statistischen Methoden herausfiltern, wie die befragten Personen die verschiedenen Attribute gegeneinander abwägen. Dies erlaubt Rückschlüsse über die Wichtigkeit der Attribute, d. h. welche Attribute spielen bei der Entscheidung die grösste Rolle und welche sind eher unwichtig. Zusätzlich lassen sich Aussagen darüber machen, wie gross der Nutzengewinn für die Befragten im Durchschnitt ist, wenn bei gewissen Eigenschaften eine Verbesserung zu verzeichnen ist. Schliesslich kann auch der Nutzen aus dem gesamten Produkt oder der gesamten Leistung berechnet werden. Weiterführende Informationen zur Methode finden sich bei Ben-Akiva and Lerman (1985), Louviere et al. (2000), Telser (2002) und Train (2008).

Im Rahmen dieses Projektes haben wir ein DCE für ein Anwendungsbeispiel umgesetzt. Die konkrete Umsetzung und die Resultate sind in Kapitel 4 dokumentiert. In Anhang C ist der verwendete Onlinefragebogen einsehbar. Die notwendigen Schritte zur Durchführung eines DCE, deren theoretische Fundierung und statistischen Auswertungsmethoden sind ebenfalls im Anhang C: Discrete-Choice-Experiment zu finden.

3.5 Gewichtung der Attribute

Ist für jedes Attribut die Wertefunktion bekannt, müssen die einzelnen Nutzenwerte zu einem Gesamtnutzen aggregiert werden. Da für den Gesamtnutzen die einzelnen Attribute unterschiedlich relevant sein können, führen wir eine Gewichtung ein. Welchen Nutzenwert eine Ausprägung eines Attributs aufweist und welche Bedeutung es am Gesamtnutzen aufweist, unterscheidet sich je nach befragter Person. Entsprechend sind die Präferenzen derselben Personengruppe wie für die Bewertung der Attribute auch für die Gewichtung der Attribute zu erfassen. Ob die Präferenzmessung für die Bewertung und die Gewichtung der Attribute gemeinsam oder getrennt erfolgen muss, hängt von der gewählten Methode ab. Die von uns gewählte dekompositionelle Methode, das DCE, erlaubt eine simultane Erhebung der Präferenzen und Gewichte, was einer der grossen Vorteile der Methode ist.

Aus dem DCE resultieren also die Gewichte für jede Eigenschaft. Diese Gewichte werden benötigt, um im nächsten Schritt die Nutzenbewertung pro Eigenschaft zu einem Gesamtnutzen zusammenzufassen. Die konkrete Berechnung der Gewichte wird anhand des Anwendungsbeispiels in Abschnitt 4.1 dargelegt.

3.6 Zusammenfassung der Einzelnutzen und Kategorisierung

In diesem Schritt wird sichergestellt, dass die Wertefunktionen und Gewichte in Übereinstimmung mit den Präferenzen der Anspruchsgruppen zu einem Gesamtnutzen kombiniert werden. Das einfachste und weitest verbreitete Konzept ist die additive Nutzenfunktion. Diese trifft zwar starke Annahmen, reduziert aber auch die Komplexität der MCDA entscheidend.¹² Da die Akzeptanz und Transparenz unseres Modells von hoher Wichtigkeit ist, wählen wir die additive Nutzenfunktion als Aggregationsmethode. Die Nutzenwerte der einzelnen Attribute werden dementsprechend mit ihren Gewichten multipliziert und danach zum Gesamtnutzen aufaddiert, welcher als Nutzenindex interpretiert wird.

Aus unserem Modell resultiert damit der Gesamtnutzen eines Arzneimittels. Das Modell ist dazu konzipiert, eine Arzneimittelinnovation mit einem Komparator zu vergleichen. Die zu vergleichenden Arzneimittel werden getrennt mit dem Modell bewertet und danach deren Gesamtnutzen verglichen. Die Differenz der Gesamtnutzen kann als Zusatznutzen der Innovation interpretiert werden. Da die Nutzenbewertung auf subjektiven Werturteilen basiert, die stets eine gewisse Unschärfe implizieren, sollte der resultierende Nutzenwert in seiner absoluten Form nicht direkt interpretiert werden.

Wir schlagen die Bildung von Nutzenkategorien vor, um dieser Unschärfe entgegenzuwirken und die Interpretation der Ergebnisse zu vereinfachen. Die Grenzwerte sollten idealerweise empirisch aufgrund der Verteilung von Nutzenwerten ermittelt, die man aus einer Stichprobe von Vergleichen erhält. Wie sich dann die jeweils erzielte Nutzenverbesserung im Preis oder in der Priorisierung der Zulassung niederschlägt, muss in Verhandlungen mit dem Regulator (fallweise) definiert werden.

Grundsätzlich kann die Kategorisierung sich an der absoluten oder relativen (oder kombinierten) Differenz der Gesamtnutzen orientieren. Die Verwendung der absoluten Differenz hat zur Folge, dass bei Therapien mit einem hohen Nutzenwert des Komparators eine grosse Nutzenverbesserung schwieriger zu erreichen ist als bei Therapien mit einem niedrigen Nutzenwert. Bei einer relativen Kategorisierung, bei der also die Verbesserung abhängig vom Nutzenwert des Komparators definiert wird, ist bei einem hohen Nutzenwert des Komparators ein niedrigerer Nutzenzuwachs für eine «grosse Nutzenverbesserung» notwendig als bei einem niedrigen Ausgangsnutzenwert.

Da uns Informationen zur Verteilung von Nutzenwerten noch fehlen, können wir diese Kategorisierung hier nicht empirisch herleiten oder untermauern. Beispielhaft schlagen wir folgende Nutzenkategorien vor, die sich an den absoluten Differenzen der Gesamtnutzen ausrichten.

¹² Die additive Nutzenfunktion nimmt an, dass die Eigenschaften präferenzunabhängig sind, d. h. die Präferenz einer Eigenschaft hängt nicht von den Ausprägungen der anderen Eigenschaften ab (vgl. Abbas, 2018).

Tabelle 9 **Beispielhafte Darstellung von Nutzenkategorien**

Nutzenkategorie	Grenzwerte Nutzenverbesserung Innovation vs. Komparator	
	Untergrenze	Obergrenze
Keine Nutzenverbesserung		≤ 2
Geringe Nutzenverbesserung	> 2	≤ 5
Mittlere Nutzenverbesserung	> 5	≤ 10
Grosse Nutzenverbesserung	> 10	

Die Tabelle stellt beispielhaft die Kategorisierung der Nutzenverbesserung einer Innovation gegenüber ihrem Komparator dar. Hier verwenden wir nur die absolute Differenz zwischen den zwei Behandlungen, Grundsätzlich kann die Kategorisierung sich auch an der relativen Nutzenänderung orientieren. Das Minimum der Nutzenwerte ist 0, das Maximum liegt bei 100.

Die Grenzwerte sollten idealerweise empirisch aufgrund der Verteilung einer Stichprobe von Vergleichen ermittelt werden. Wie sich dann die jeweils erzielte Nutzenverbesserung im Preis oder in der Priorisierung der Zulassung niederschlägt, muss in Verhandlungen mit dem Regulator (fallweise) definiert werden.

Quelle: Eigene Darstellung, Polynomics.

3.7 Umgang mit Unsicherheit: Sensitivitätsanalyse

Die Validierung einer gegebenen Bewertung und Gewichtung ist schwierig, da sie immer subjektive Werturteile enthalten. Ein robustes Ergebnis erhält man deshalb nur, wenn es umfangreichen Sensitivitätsanalysen standhält. So kann sichergestellt werden, dass ein Ergebnis nicht von spezifischen Werturteilen oder Unsicherheit getrieben ist. Im vorliegenden Modell kann Unsicherheit aufgrund verschiedener Ursachen auftreten. Im Folgenden diskutieren wir die in Abschnitt 2.2 genannten Ursachen von Unsicherheit.

Die strukturelle Unsicherheit ist die wichtigste Art von Unsicherheit in unserem Modell. Strukturelle Unsicherheit tritt insbesondere bei der Kriterienauswahl auf, wo alle für die Frage relevanten Nutzenkomponenten erfasst werden müssen. Als Sensitivitätsanalyse können hier weitere Attribute respektive Messgrößen berücksichtigt oder auf bestehende verzichtet werden.¹³ Beim Attribut «Überleben» könnte zum Beispiel die objektive Ansprechrage (objective response rate, ORR) als zusätzliche Messgrösse berücksichtigt werden. Strukturelle Unsicherheit tritt jedoch auch im weiteren Verlauf der MCDA auf. Beispielsweise bei der Annahme der (Nicht-)Linearität der Wertefunktionen oder der Methodenauswahl zur Ermittlung der Wertefunktionen und Gewichte. Auch hier können in der Sensitivitätsanalyse zum Beispiel die Linearität der Wertefunktion variiert oder alternative Berechnungsmethoden getestet werden. Sobald mehrere Unsicherheiten auf verschiedenen Stufen im Prozess bestehen, empfiehlt es sich eine probabilistische Sensitivitätsanalyse durchzuführen, d. h. alle Unsicherheiten in einer gemeinsamen Analyse zu prüfen.

Stochastische Unsicherheit kann bei allen Attributen auftreten. In unserem Modell wäre sie im Schritt der Leistungsmessung anzusiedeln, d. h. bei Ermittlung der effektiven Ausprägungen der Attribute für Innovation und Komparator. Auf dieser Stufe ist ebenfalls die Unsicherheit über die klinische Evidenz eine Unsicherheitsquelle. Als Sensitivitätsanalyse können zum Beispiel die Konfidenzintervalle aus den klinischen Studien anstelle der Punktschätzer verwendet werden.

¹³ Der Verzicht auf Kriterien kann auch über eine Änderung des Gewichtungsschemas erreicht werden. Wird ein Gewicht von 0 gesetzt, entspricht dies einer Nichtberücksichtigung des entsprechenden Attributs.

Falls vorhanden können unterschiedliche Quellen für einen Wert berücksichtigt werden. Dies gilt insbesondere, wenn Werte anhand von Expertenmeinungen ermittelt wurden.

Parameterunsicherheit tritt bei der Berechnung der Wertefunktionen und der Gewichte anhand der empirisch ermittelten Daten auf. Mit passenden statistischen Methoden können hier wiederum Konfidenzintervalle berechnet werden, die im Rahmen einer Sensitivitätsanalyse verwendet werden können. Alternativ kann die Robustheit der Parameter auch mit verschiedenen Berechnungsmethoden und Modellspezifikationen getestet werden.

Heterogenität bezieht sich insbesondere auf die Präferenzen von Subgruppen der Bevölkerung. Heterogene Präferenzen lassen sich in der Auswertung des DCE berücksichtigen. Beispielsweise können die Gewichte nur für eine (homogene) Subgruppe der Befragten berechnet oder verschiedene Akteursgruppen unterschiedlich stark gewichtet werden.

4 Anwendungsbeispiel

In diesem Kapitel zeigen wir die Umsetzung unseres Modells an einem Beispiel. Abschnitte 4.1 und 4.2 beschreiben die Kalibrierung des Modells mittels eines Discrete-Choice-Experiments. In Abschnitt 4.3 untersuchen wir die Heterogenität der Präferenzen für verschiedene Bevölkerungsgruppen. Abschnitt 4.4 schliesslich zeigt die exemplarische Anwendung für die Bewertung einer Arzneimittelinnovation für fortgeschrittene Nierenzellkarzinome. Wir bewerten die Kombinationstherapie aus Nivolumab plus Ipilimumab als Innovation im Vergleich zu Sunitinib als Komparator (Standard of Care). Das Anwendungsbeispiel hat zum Ziel, die Umsetzung des oben vorgeschlagenen Modells zu veranschaulichen.

4.1 Stichprobenauswahl und Befragungsdesign

Als erstes geht es darum, das Modell zu kalibrieren, das heisst, die Parameter des Modells zu bestimmen. Dazu gehört insbesondere die Festlegung der Wertefunktionen sowie die Bestimmung der Gewichte der verschiedenen Attribute.

In Abschnitt 3.4.4 empfehlen wir, dafür ein Discrete-Choice-Experiment (DCE) durchzuführen, mit dem die Präferenzen der Schweizer Bevölkerung erfasst werden sollten. In einem ersten Schritt haben wir ein DCE mit Vertretern der Anspruchsgruppen Leistungserbringer, Finanzierer, Patienten, Industrie, Politik und Ökonomie durchgeführt. Insgesamt wurden 32 Experten eingeladen, sich an einer Onlineumfrage zu beteiligen. Die drei erstgenannten Gruppen waren in unserer Stichprobe übervertreten. Antworten haben wir von 19 Befragten erhalten, wobei 18 Antworten vollständig waren. Die Expertenbefragung diente uns als Machbarkeitsstudie und war die Grundlage für die repräsentative Befragung der Schweizer Bevölkerung. Die Resultate sind im Anhang D: Resultat der Expertenbefragung aufgeführt. Sie kam zu qualitativ vergleichbaren Gewichten, was darauf hindeutet, dass die geschätzte Gewichtung robust ist. Ein detaillierter Vergleich zwischen Experten- und Laien-Gewichtungen ist nicht möglich, weil die Teilnehmerzahl bzw. die Stichprobengrösse bei der Expertenumfrage zu gering war.

In einem zweiten Schritt wurde das DCE als repräsentative Befragung der Schweizer Bevölkerung in Zusammenarbeit mit dem Befragungsdienst gfs.bern durchgeführt. Dazu wurde zuerst ein qualitativer Pretest mit fünf Interviews durchgeführt, um das Verständnis und die Durchführbarkeit des Experiments sicherzustellen. Die Hauptbefragung fand im Februar 2023 statt. Die Befragung umfasste die Schweizer Bevölkerung ab 18 Jahren aus der deutsch- oder französischsprachigen Schweiz. Die Stichprobe wurde quotiert auf Alter, Geschlecht und Sprachregion. Insgesamt konnten 1'034 vollständige Interviews realisiert werden.

Die Behandlungsoptionen waren durch die acht in Tabelle 10 beschriebenen Attribute und ihre Ausprägungen charakterisiert. Das Design, d. h. die Charakterisierung der Behandlungsoptionen und die Auswahl der Entscheidungssituationen, wurde mit dem Modul *dcreate* der Statistiksoftware Stata (Version 17) erstellt.¹⁴ Insgesamt wurden 16 Entscheidungssituationen erstellt, die auf zwei Blöcke aufgeteilt wurden. Somit musste jeder Befragte acht Entscheidungen treffen. Abbildung 2 zeigt ein Beispiel einer Entscheidungssituation, wie sie den Befragten präsentiert wurde. Die Reihenfolge der Darstellung der Attribute zum Patientennutzen und gesellschaftlichen Nutzen wurden randomisiert, sodass bei einigen Befragte zuerst die Attribute zum Patientennutzen aufgeführt wurden und bei anderen zuerst die Attribute zum gesellschaftlichen Nutzen.¹⁵ Vor den

¹⁴ Vgl. <https://ideas.repec.org/c/boc/bocode/s458059.html>, zuletzt aufgerufen am 23.08.2021.

¹⁵ Durch die Randomisierung konnten wir testen, ob die Reihenfolge in der die Attribute präsentiert wurden einen

acht Entscheidungssituationen wurde den Befragten die Attribute mit ihren Ausprägungen erläutert. Nach den Entscheidungssituationen folgten Fragen zur Verständlichkeit sowie sozioökonomische Fragen. Der ganze Fragebogen ist im Anhang C: Discrete-Choice-Experiment enthalten.

Tabelle 10 **Verwendete Ausprägungen aller Attribute im DCE**

Attribut	Einheit	Wert1	Wert 2	Wert 3	Wertfunktion
1 Überleben (Median OS oder PFS)	Monate	3 (OS)	24 (OS)	60 (OS)	nichtlinear
2 Gesundheitsbezogene Lebensqualität	PROM utility score	10	45	80	nichtlinear
3 Unerwünschte Ereignisse	Prozent der Patienten	5	80		linear
4 Behandlungshäufigkeit	Alle ... Tage	1	30		linear
5 Behandlungsort: 0 (amb.) 1 (zu Hause)	Dummy	0	1		linear
6 Krankheitslast in der Schweiz	DALY/100k EW/Jahr	46	460	920	nichtlinear
7 Indirekte Kosten (Arbeitsausfall)	Monate	1	6		linear
8 Andere Behandlungskosten	1'000 CHF	10	80		linear

Die Behandlungsoptionen wurden im DCE anhand von acht Attributen beschrieben. Die abstrakteren Werte wurden für das Experiment qualitativ ausgedrückt und mit Beispielen erläutert. Die PROM utility scores 10, 45 und 80 beispielsweise hiessen «gravierende», «deutliche» und «leichte» Einschränkungen; die Krankheitslast wurde als «gering», «mittel» und «hoch» beschrieben, vgl. Tabelle 11. Für eine ausführliche Beschreibung siehe Anhang C: Discrete-Choice-Experiment.

Quelle: Eigene Darstellung, Polynomics.

Abbildung 2 Beispiel einer Entscheidungssituation

Wählen Sie die Option, die Sie persönlich besser finden.

Eigenschaft	Option 1	Option 2
Krankheitslast in der Gesellschaft [Ⓢ]	geringe Krankheitslast [Ⓢ]	hohe Krankheitslast [Ⓢ]
Verlorene Arbeitszeit je Behandlung [Ⓢ]	6 verlorene Arbeitsmonate [Ⓢ]	1 verlorener Arbeitsmonat [Ⓢ]
Weitere Kosten je Behandlung [Ⓢ]	10'000 CHF [Ⓢ]	80'000 CHF [Ⓢ]
Überlebenszeit [Ⓢ]	2 Jahre [Ⓢ]	3 Monate [Ⓢ]
Langfristige Lebensqualität [Ⓢ]	mittlere Einschränkungen [Ⓢ]	leichte Einschränkungen [Ⓢ]
Schwere Nebenwirkungen [Ⓢ]	80% der Patienten [Ⓢ]	5% der Patienten [Ⓢ]
Behandlungshäufigkeit [Ⓢ]	täglich [Ⓢ]	täglich [Ⓢ]
Behandlungsort [Ⓢ]	Arztpraxis/Klinik [Ⓢ]	zu Hause [Ⓢ]

Option 1

Option 2

<<

>>

Jede befragte Person hatte acht unterschiedliche Entscheidungssituation, bei welcher sie jeweils entscheiden musste, welche Option sie persönlich besser findet. Die Ausprägungen der Attribute beider Optionen variieren dabei zwischen den Entscheidungssituationen.

Quelle: Darstellung gfs.bern.

Einfluss auf die Entscheidungen hat. Unsere Auswertungen haben ergeben, dass dies nicht der Fall ist. Die Teilnehmer haben sich also diesbezüglich nicht strategisch verhalten und beispielsweise nur die obersten drei Attribute miteinander verglichen.

4.2 Statistische Auswertung und Resultate DCE

Die Auswertung der Umfrage erfolgte mit der Statistiksoftware Stata 17. Wir schätzen ein Mixed-Logit-Modell, mit der Entscheidung als abhängige und den acht kategorischen Variablen der Attribute als unabhängige Variablen. Kategorische Variable heisst hier, dass die oberen Ausprägungen (Max bzw. zweiter Wert und Max) jeweils in Relation zur untersten Ausprägung (Min) und nicht als kontinuierliche Funktion geschätzt werden. Der Koeffizient für die unterste Ausprägung nimmt so definitionsgemäss immer den Wert 0 an. Die ökonometrischen Ergebnisse finden sich in Tabelle 11.

Tabelle 11 Koeffizienten aus dem DCE

Attribut	Ausprägung	Koeffizient	Standardfehler	p-Wert
Überlebenszeit nach Therapiebeginn (Median)	3 Monate	0	.	.
	2 Jahre	1.625 ***	0.114	0.000
	5 Jahre	2.537 ***	0.160	0.000
Langfristige Lebensqualität	gravierende Einschränkungen	0	.	.
	mittlere Einschränkungen	1.013 ***	0.055	0.000
	leichte Einschränkungen	1.496 ***	0.075	0.000
Schwere unerwünschte Ereignisse	5%	0	.	.
	80%	-0.792 ***	0.049	0.000
Behandlungshäufigkeit	täglich	0	.	.
	monatlich	0.194 ***	0.032	0.000
Behandlungsort	ambulant	0	.	.
	zu Hause	0.305 ***	0.035	0.000
Krankheitslast in der Gesellschaft	geringe Krankheitslast	0	.	.
	mittlere Krankheitslast	-0.023	0.031	0.473
	hohe Krankheitslast	-0.138 ***	0.029	0.000
Indirekte Kosten für die Gesellschaft	1 verlorener Arbeitsmonat	0	.	.
	6 verlorene Arbeitsmonate	-0.029	0.023	0.193
Weitere Behandlungskosten	10'000 CHF	0	.	.
	80'000 CHF	-0.242 ***	0.027	0.000

***: signifikant auf dem 0.1-Prozent-Niveau, **: 1-Prozent-Niveau, *: 5-Prozent-Niveau.

Geschätzt wurde ein cmxtmixlogit-Modell ohne Konstante mit einem Entscheidungs-Dummy als abhängiger Variable und den Attributen als kategorische Variablen als unabhängige Variablen. Sonstige Kovariaten etwa zu den Eigenschaften der Befragten gab es nicht, da das Sample repräsentativ für die Schweizer Bevölkerung ist, deren Präferenzen wir erheben wollen.

Statistisch signifikant von null verschieden sind alle Ausprägungen, mit Ausnahme der mittleren Krankheitslast (gegenüber einer geringen Krankheitslast) und den indirekten Kosten für die Gesellschaft. Alle Vorzeichen sowie die Höhe der Koeffizienten sind nachvollziehbar.

Quelle: Eigene Berechnungen aus DCE-Experiment, Polynomics.

Alle Koeffizienten haben plausible Vorzeichen. Statistische Signifikanz wird für alle Attribute ausser den indirekten Kosten für die Gesellschaft erreicht. Die Krankheitslast in der Gesellschaft ist nur für die Ausprägung «hoch» signifikant negativ. Die Schweizer Bevölkerung möchte also eher Krankheiten mit geringer oder mittlerer Krankheitslast behandeln als welche mit hoher Krankheitslast für die Gesellschaft. Die Ausprägung «2 Jahre» beim Attribut Überlebenszeit weist einen Koeffizienten von 1.625 auf. Eine Veränderung von 3 Monaten (Basisausprägung) auf 2 Jahre Überlebenszeit wird somit wie erwartet positiv bewertet. Eine Veränderung von 3 Monaten auf 5 Jahre Überlebenszeit wird mit einem Koeffizienten von 2.537 ebenfalls positiv bewertet, die Veränderung von 2 auf 5 Jahre ergibt jedoch nur einen zusätzlichen Wert von 0.9 (2.537-1.625). Die Interpretation für die Koeffizienten der übrigen Attribute erfolgt analog. Die Werte der Koeffizienten können nicht absolut interpretiert werden, sondern nur relativ zueinander.

Bestimmung der Wertefunktion

Aus den geschätzten Koeffizienten lassen sich nun die Wertefunktionen und die Gewichte der Attribute herleiten. Bei den Wertefunktionen unterscheiden wir zwischen linearen und nichtlinearen Wertefunktionen. Die linearen Wertefunktionen ergeben sich als lineare Interpolation zwischen Minimum und Maximum der jeweiligen Attribute. Zusätzlich muss bestimmt werden, ob es sich um eine steigende oder fallende (inverse) Funktion handelt. Beides wurde bereits in Abschnitt 3.4.1 festgelegt. Die Schätzergebnisse des Experiments können wir für die Überprüfung der Steigung verwenden: ist das Vorzeichen der Koeffizienten positiv, ergibt sich eine steigende Wertefunktion, wenn er negativ ist, eine fallende. Die Punktschätzer bestätigen die Form der linearen Wertefunktionen, die wir nochmal in Tabelle 12 zusammengefasst haben.

Tabelle 12 Bestimmung der linearen Wertefunktionen

Attribut	Min	Max	Steigung
Schwere unerwünschte Ereignisse	5	80	Fallend (100 bei 5, 0 bei 80)
Behandlungshäufigkeit	1	30	Steigend (0 bei 1, 100 bei 30)
Behandlungsort	0	1	Steigend (0 bei 0, 100 bei 1)
Indirekte Kosten für die Gesellschaft	30	180	Fallend (0 bei 180, 100 bei 30)
Weitere Behandlungskosten	10'000	80'000	Fallend (0 bei 80k, 100 bei 10k)

Diese drei Angaben genügen, um die linearen Wertefunktionen einzupassen und zu skalieren. Das Experiment ist dafür nicht nötig. Das braucht es anschliessend für die Bestimmung der Gewichte.

Quelle: Eigene Darstellung, Polynomics.

Bei nichtlinearen Wertefunktionen ermitteln wir die Parameter (a, b, c) , die die quadratische Funktion $y_i = a(x_i)^2 + bx_i + c$ durch die drei Punkte (x_i^1, y_i^1) , (x_i^2, y_i^2) und (x_i^3, y_i^3) legt. Die x-Werte kommen dabei aus Tabelle 10, die y-Werte sind die Koeffizienten aus Tabelle 11. Für Überleben beispielsweise ergeben sich damit die Punkte (3,0), (24,1.625) und (60,2.537). Auch die nichtlinearen Wertefunktionen müssen nun entsprechend von 0 für ihren niedrigsten Wert bis 100 für ihren höchsten Wert skaliert werden. Die Werte der Parameter (a, b, c) sind in Tabelle 13 zusammengefasst.

Tabelle 13 Parameter für die Ermittlung der nichtlinearen Wertefunktionen

Attribut	a	b	c
Überlebenszeit nach Therapiebeginn (Median)	$-9.13 \cdot 10^{-4}$	0.102	-0.298
Langfristige Lebensqualität	$-2.16 \cdot 10^{-4}$	0.041	-0.387
Krankheitslast in der Gesellschaft	$-2.24 \cdot 10^{-7}$	$5.86 \cdot 10^{-5}$	$-2.22 \cdot 10^{-3}$

Die Parameter (a, b, c) passen die Funktion $y_i = a(x_i)^2 + bx_i + c$ für nichtlineare Wertefunktionen in die durch die Koeffizienten vorgegebenen Punkte ein. Anschliessend muss zwischen niedrigstem und höchstem Wert skaliert werden, sodass der niedrigste Wert 0 entspricht und der höchste 100.

Quelle: Eigene Berechnung, Polynomics.

Berechnung der Gewichte

Da wir für alle Attribute die gesamte Spannweite plausibler Werte abdecken, kann die Distanz der jeweiligen Schätzer zwischen der niedrigsten und der höchsten Ausprägung als Gewicht für das gegebene Attribut interpretiert werden (vgl. Tervonen et al., 2017, p. 2). Es ist zu beachten, dass wir für die nichtlinearen Wertefunktionen jeweils überprüfen müssen, ob im relevanten Wertebereich die gefittete Parabel ein Minimum oder Maximum hat. In diesem Fall ist statt der Distanz, die sich aus den Koeffizienten ergibt, je nachdem die Distanz zwischen diesem jeweils «tatsächlichen» Minimum oder Maximum und dem entsprechenden Koeffizienten zu bestimmen. In unserem Fall betrifft dies die Attribute Überlebenszeit und Krankheitslast. Das Maximum der Überlebenszeit ist (kurz) vor dem von uns berücksichtigten höchsten Wert von 60 Monaten (Max bei $y = 2.552$ für $x = 55.9$). Für die Krankheitslast ergibt sich eine Parabel in U-Form, die zunächst fällt (bis zu ihrem Minimum bei $y = 1.62 \cdot 10^{-4}$ für $x = 131.1$) und anschliessend wieder ansteigt. Demnach ergibt sich die in Tabelle 14 dargestellte Gewichtung.

Tabelle 14 Gewichtung der Attribute aus dem DCE

Attribut	Distanz der Koeffizienten	Gewicht
Überlebenszeit nach Therapiebeginn (Median)	2.553	44.4%
Langfristige Lebensqualität	1.496	26.0%
Schwere unerwünschte Ereignisse	0.792	13.8%
Behandlungshäufigkeit	0.194	3.4%
Behandlungsort	0.305	5.3%
Krankheitslast in der Gesellschaft	0.138	2.4%
Indirekte Kosten für die Gesellschaft	0.029	0.5%
Weitere Behandlungskosten	0.242	4.2%
Summe	5.749	100.0%

Die Distanz der Koeffizienten ergibt sich für Attribut i aus $D_i = |\max(\beta_i) - \min(\beta_i)|$. Das Gewicht folgt als $G_i = D_i / \sum_i D_i$. Die Distanz der Koeffizienten für Überlebenszeit und Krankheitslast wurde, wie im Haupttext erklärt, aufgrund der nichtlinearen Form der Wertefunktion angepasst.

Quelle: Eigene Berechnungen, Polynomics.

Für unsere Bevölkerungsbefragung ergibt sich folgendes Bild betreffend die Gewichtung der Attribute. Das mit Abstand wichtigste Attribut ist mit einem Wert von 44 Prozent die Überlebenszeit

nach Therapiebeginn. Die eher langfristige gesundheitsbezogene Lebensqualität und die Häufigkeit der eher kurzfristig während der intensiven Phase der Behandlung auftretenden schweren Nebenwirkungen sind zusammen weniger wichtig als die Überlebenszeit: Ihre Gewichte addieren sich zu rund 40 Prozent, wobei der langfristigen Lebensqualität mit einem Gewicht von 26 Prozent eine fast doppelt so hohe Bedeutung beigemessen wird, wie der Häufigkeit der Nebenwirkungen mit 14 Prozent. Die verbleibenden patientenbezogenen Attribute der Behandlungshäufigkeit und des Behandlungsorts kommen gemeinsam auf weitere rund 9 Prozent.

Damit bleiben nur rund 7 Prozent für die gesellschaftlichen Folgen der Behandlung einer Krankheit. Das wichtigste Attribut sind dabei die weiteren Behandlungskosten mit 4 Prozent. Praktisch keine Rolle spielen die indirekten Kosten für die Gesellschaft aufgrund von Produktivitätsausfällen.

Die Ergebnisse sind mit denjenigen der Expertenbefragung vergleichbar, vgl. Tabelle 21 im Anhang. Allerdings wurde dort die Überlebenszeit mit gut 50 Prozent etwas stärker gewichtet, dafür war das Gewicht für Lebensqualität und Nebenwirkungen mit gut 20 Prozent etwas geringer. Zusammen macht bei der Expertenbefragung wie bei der Bevölkerungsbefragung der Behandlungserfolg 70 Prozent aus. Die gesellschaftlichen Folgen der Behandlung erhielt bei der Expertenbefragung eine Gewichtung von gut 8 Prozent, wobei die Krankheitslast am wichtigsten war, gefolgt von den indirekten Kosten.

4.3 Präferenzheterogenität

Um die unterschiedlichen Präferenzen in der Schweizer Bevölkerung zu erfassen, haben wir das Modell für einzelne Bevölkerungsgruppen separat geschätzt und die Gewichte berechnet. Insgesamt unterscheidet sich die Gewichtung nur leicht. Die französischsprachige Bevölkerung gibt der Überlebenszeit etwas mehr Gewicht als die deutschsprachige, umgekehrt ist das Gewicht für die Lebensqualität bei der deutschsprachigen Bevölkerung etwas höher als bei der französischsprachigen. Befragte über 64 Jahre geben der Überlebenszeit ebenfalls etwas weniger Gewicht als andere Bevölkerungsgruppen. Die Gewichte sind in Tabelle 15 zusammengefasst.

Tabelle 15 Gewichtung der Attribute aus dem DCE für einzelne Gruppen

Attribut	Deutschsprachig	Franz.-sprachig	Mann	Frau	bis 64 Jahre	> 64 Jahre
Überlebenszeit nach Therapiebeginn (Median)	42.3%	46.1%	44.6%	41.3%	43.7%	40.8%
Langfristige Lebensqualität	28.1%	24.7%	27.6%	26.5%	27.4%	27.7%
Schwere unerwünschte Ereignisse	14.2%	13.3%	13.1%	14.7%	13.3%	16.8%
Behandlungshäufigkeit	2.9%	4.7%	3.1%	3.7%	4.0%	1.0%
Behandlungsort	5.7%	4.6%	4.5%	6.2%	5.4%	5.1%
Krankheitslast in der Gesellschaft	2.3%	3.0%	2.7%	3.4%	2.4%	2.8%
Indirekte Kosten für die Gesellschaft	0.0%	0.0%	0.0%	0.0%	0.0%	0.0%
Weitere Behandlungskosten	4.4%	3.8%	4.3%	4.2%	3.9%	5.8%

Die Tabelle zeigt die Gewichtung der einzelnen Attribute für verschiedene Bevölkerungsgruppen. Für die Berechnung wurde das Modell auf Befragte der entsprechenden Bevölkerungsgruppen eingeschränkt, z. B. nur deutschsprachige Befragte. Die Schätzung wurde mit den kontinuierlichen Werten aus Tabelle 10 vorgenommen. (Für die Hauptauswertung wurden kategorische Variablen verwendet, um die jeweiligen Ausprägungen zu kennzeichnen.) Mit dieser Variante können für die nichtlinearen Funktionen direkt die Koeffizienten als Parameter verwendet werden. Diese alternative Vorgehensweise kann zu kleineren Abweichungen gegenüber den Resultaten in Tabelle 14 führen, so ist z. B. das Gewicht für die indirekten Kosten kleiner als 0.5%. Weiter ist anzumerken, dass durch die Korrektur von «tatsächlichen» Minimum oder Maximum bei den nichtlinearen Wertefunktionen das durchschnittliche Gewicht über die Bevölkerungsgruppen vom Gewicht der Gesamtbevölkerung leicht abweichen kann. Die Unterschiede zwischen den Gruppen sind basierend auf unserer Stichprobengrösse statistisch nicht signifikant, zeigen jedoch gewisse Tendenzen auf.

Quelle: Eigene Berechnungen, Polynomics.

4.4 Nutzenbewertung: Fortgeschrittenes Nierenzellkarzinom

Unser Anwendungsfall bezieht sich auf die Therapie des fortgeschrittenen Nierenzellkarzinoms, welches nichtkurativ behandelt wird. Als Arzneimittelinnovation betrachten wir die Kombinations-therapie aus den Wirkstoffen Nivolumab plus Ipilimumab. Als Komparator dient die Therapie mit Sunitinib. Wir verwenden die in Tabelle 16 dargestellten Datenquellen.

Tabelle 16 **Verwendete Datenquellen und deren Qualität – Anwendungsfall**

Datenquelle	Bewertung Datenqualität
Motzer R. J. et al. (2018). Nivolumab plus Ipilimumab versus Sunitinib in advanced renal-cell carcinoma. <i>New England Journal of Medicine</i> .	▪ begutachtet (peer reviewed) ▪ Phase-III-Studie (Zulassungsstudie) ▪ sehr gut
Albiges L. et al. (2020). Nivolumab plus Ipilimumab versus Sunitinib for first-line treatment of advanced renal cell carcinoma: extended 4-year follow-up of the phase III CheckMate 214 trial. <i>ESMO Open</i> .	▪ begutachtet (peer reviewed) ▪ Phase-III-Studie ▪ sehr gut
Cella D. et al. (2019). Patient-reported outcomes of patients with advanced renal cell carcinoma treated with Nivolumab plus Ipilimumab versus Sunitinib (CheckMate 214): A randomized, phase 3 trial. <i>The Lancet Oncology</i> .	▪ begutachtet (peer reviewed) ▪ Phase-III-Studie ▪ sehr gut
Oniangue-Ndza C. et al. (2019). Cost effectiveness analysis of Nivolumab in combination with Ipilimumab for the first-line treatment of advanced/metastatic renal cell carcinoma in Switzerland. <i>ISPOR Europe Conference 2019</i> .	▪ Konferenz-review ▪ aufbauend auf Phase-III-Studie ▪ gut
Experteneinschätzung	▪ situativ zu beurteilen

Quelle: Eigene Darstellung, Polynomics.

Aus den Datenquellen lesen wir die in Tabelle 17 zusammengefassten Ausprägungen der Attribute unseres Wertebaums. Dabei gilt es einige Besonderheiten zu beachten. Die eigentliche Zulassungsstudie von Motzer et al. (2018) verfolgt die Patienten nicht lange genug, um die Medianüberlebenszeit für die Innovation auszuweisen. Als untere Grenze lässt sich nur sagen, dass sie grösser als 32 Monate ist. Für den Komparator werden 26.0 Monate angegeben. Wir greifen deshalb auf eine erweiterte Auswertung von Albiges et al. (2020) zurück, welche die Studienpopulation nach vier Jahren erneut ausgewertet hat. Die Medianüberlebenszeit für Nivolumab und Ipilimumab beträgt 48.1 Monate im Vergleich zu Sunitinib mit neu 26.6 Monate. Wir verwenden die Werte derselben Studienpopulation wie in der Zulassungsstudie von Motzer et al. (2018), d. h. Patienten mit hohem und mittlerem Risiko («intermediate- and poor-risk population»).

Auch die Angaben zur langfristigen Lebensqualität lassen sich nicht direkt aus der Zulassungsstudie ermitteln. Die dort angegebenen FKSI-19-Werte gehören nicht zu den präferierten PROM, weshalb wir auf die Auswertungen in Cella et al. (2019) zurückgreifen. Hier werden EQ-5D-3L Nutzenwerte für Grossbritannien (UK utilities) veröffentlicht. Wir verwenden die Patienten-Gruppe mit hohem und mittlerem Risiko («participants with intermediate or poor risk»). Eigentlich ist vorgesehen, die durchschnittliche Lebensqualität in dem Zeitraum zwischen der halben und der ganzen Medianüberlebenszeit zu verwenden (vgl. Abschnitt 3.2.2). Dies würde hier bedeuten, dass wir die durchschnittliche gesundheitsbezogene Lebensqualität in dem Zeitraum von 24 bis 48 Monaten bestimmen. Für diesen Zeitraum gibt es jedoch keine Daten. In allen uns bekannten Studien stammen die letzten Beobachtungen von Woche 103 (Monat 24) nach dem Beginn der Behandlung. Wir schreiben deshalb die Entwicklung der letzten Beobachtungen fort, wobei wir dies (etwas unscharf, da «zu früh») ab der Woche 91, also ab Monat 21 machen. Dies ermöglicht es uns jedoch, noch von grösseren Studienpopulationen und mehr als einer Beobachtung zu profitieren. So kommen wir auf Werte von 82 für die Innovation und 75 für den Komparator (siehe Cella et al., 2019 Table 2 und Figure 3). Die damit einhergehende Unsicherheit kann im Rahmen einer Sensitivitätsanalyse bestimmt werden. Für die Bestimmung der Untergrenze der Nutzenverbesserung würde man unterstellen, dass sich die PROM-Werte angleichen; für die Obergrenze der Nutzenverbesserung, dass sie weiter auseinandergehen.

Tabelle 17 Ausprägungen der Attribute für Innovation und Komparator des Anwendungsfalls

Attribut	Messgrösse	Innovation (Nivolumab + Ipilimumab)	Komparator (Sunitinib)	Quelle
Überleben	Median OS	48.1 Monate	26.6 Monate	Albiges et al. (2020)
Lebensqualität	EQ-5D 3L UK	82	75	Cella et al. (2019)
Unerwünschte Ereignisse	Anteil unerwünschte Ereignisse	48%	64%	Albiges et al. (2020)
Patientenkomfort	Ort der Behandlung	ambulant (Praxis oder Klinik)	zu Hause	Motzer et al. (2018)
	Häufigkeit der Behandlung	18.7 Tage	1.5 Tage	Motzer et al. (2018)
Krankheitslast	DALY (Rate per 100k Einwohner pro Jahr)	ca. 31, verwendet: Minimum von 46	wie Innovation	GBD ¹⁶ , Cella et al. (2019)
Kosten	Indirekte Kosten (Arbeitsausfall in Tagen)	1	1	-
	Zusätzliche OKP-Kosten	72'100	73'335	Oniangue-Ndza et al. (2019)

Die Tabelle fasst die Ausprägungen der Attribute unseres Wertebaums für die Innovation und den Komparator zusammen. Die Werte wurden aus den Datenquellen hergeleitet, Details zur Herleitung finden sich im Haupttext, die Datenquellen sind in Tabelle 16 bewertet.

Quelle: Eigene Darstellung, Polynomics.

Für den Anteil Patienten mit schweren unerwünschten Ereignissen weisen sowohl Motzer et al. (2018) als auch Albiges et al. (2020) Werte aus. Da wir beim Überleben bereits mit den Daten von Albiges et al. (2020) arbeiten, verwenden wir auch hier diese Quelle. Für die Kombinationstherapie sind dies 48 Prozent für die Kontrolltherapie 64 Prozent. Motzer et al. 2018 weist ähnliche Werte von 46 und 63 Prozent aus.

Für die Häufigkeit und den Ort der Verabreichung greifen wir auf die Zulassungsstudie zurück. Demnach benötigt die Kombinationstherapie typischerweise 4 Dosen im Abstand von je 3 Wochen, danach alle 2 Wochen intravenöse Verabreichung. Die Behandlungsdauer betrug im Median 7.9 Monate. So kommen wir auf eine durchschnittliche Behandlungsfrequenz von 14.5 Tagen. Der Komparator erfordert einmal täglich eine orale Einnahme über 4 Wochen innerhalb jedes 6-Wochen-Zyklus, was sich zu einer Behandlungsfrequenz von 1.5 Tagen übersetzt (unabhängig der Behandlungsdauer). Behandlungsort ist ambulant (Innovation) bzw. zu Hause (Komparator).

Die Krankheitslast approximieren wir mithilfe von Informationen aus der Datenbank der «Global Burden of Disease» und den Angaben zur Inzidenz von fortgeschrittenem Nierenzellkarzinom in Cella et al. (2019). Demnach gibt es in der Schweiz 2019 104 verlorene DALY je 100'000 Einwohner. Von denen sind rund 30 Prozent in einem fortgeschrittenen Stadium, sodass der Wert der Krankheitslast für Innovation und Komparator bei rund 31 liegt. Da der kleinste abgefragte Wert bei 46 liegt, verwenden wir diesen für unser Anwendungsbeispiel.

¹⁶ Vgl. <http://ghdx.healthdata.org/gbd-results-tool?params=gbd-api-2019-permalink/f9b803aba2c9801e67718a318a1aed2c>, zuletzt aufgerufen am 23.8.2021.

Angaben zu den indirekten Kosten konnten wir leider keine finden. Eine Abschätzung durch Experten für die exemplarische Umsetzung ist nicht verhältnismässig. Wir nehmen an, dass sich dieses Attribut zwischen Innovation und Komparator nicht unterscheidet und verwenden die kleinste Ausprägung von einem Monat Arbeitsausfall. Die indirekten Kosten werden in der Sensitivitätsanalyse nochmals aufgenommen. Bleiben die weiteren Behandlungskosten. Gemäss unserer Definition handelt es sich dabei um Behandlungskosten, die neben den Kosten der betrachteten Arzneimittel typischerweise anfallen. Gemäss einer Schätzung liegen diese mit 72'100 und 73'335 CHF in etwa gleichauf auf hohem Niveau (Oniangue-Ndza et al., 2019).

Gibt man die oben beschriebenen Werte in unser Modell, erhält man für die Innovation 80 und für den Komparator 68 Punkte (von einem theoretischen Maximum von 100 Punkten). Welches Attribut wie viel zu der Gesamtbewertung beiträgt, kann der Zusammensetzung aus den Einzelnutzen in Tabelle 18 entnommen werden.

Auch wenn wir mit der Bevölkerungsbefragung ausreichend und repräsentativen Daten erhoben haben, die eine genaue Schätzung der Gewichte zulässt, gibt die Berechnung der Einzelnutzen und der Gesamtbewertung eine arithmetische Genauigkeit vor, die nicht gerechtfertigt ist. Die direkte Interpretation der Nutzendifferenz ist nicht zu empfehlen. Wir schlagen eine Nutzenkategorisierung vor, die eine Interpretation der Nutzenverbesserung zulässt (vgl. Abschnitt 3.6).

Es bleiben noch folgende Punkte, die illustrieren, wo wir nicht mit der erforderlichen Präzision vorgehen konnten:

- Die vorgesehenen Wertebereiche mussten teilweise geringfügig verlassen werden. Im Modell wurden stattdessen die nächsten Grenzwerte eingesetzt, was die Ergebnisse verfälschen kann, wenn die Grenzwerte nicht richtig gesetzt wurden. Dies betrifft die Bewertung der Innovation. Mit einer langfristigen Lebensqualität von 82 ist das zweitwichtigste Attribut oberhalb des definierten Maximums von 80 (vgl. dazu die Sensitivitätsanalyse im folgenden Abschnitt). Das zweite Attribut, das ausserhalb des kalibrierten Bereichs liegt, ist die Krankheitslast, die eher bei 31 als bei 46 ist. Hier sind aber Innovation und Komparator gleichermassen betroffen, weshalb es erst zu einer Verzerrung käme, wenn der Komparator eine andere Krankheitslast hätte.
- Schliesslich ist die Datenqualität selbst nicht optimal, was zu stochastischer Unsicherheit führt. So kann die Lebensqualität nur zwei Jahre beobachtet werden; idealerweise hätten wir hier einen Beobachtungszeitraum bis zum Medianüberleben.

Tabelle 18 Nutzenbewertung des Anwendungsfalls

Eigenschaft	Einheit	Gewicht	Wertefunktion	Min/Max	Nivolumab + Ipilimumab		Sunitinib (SoC)	
					Eingabe	Wert	Eingabe	Wert
1 Überlebenszeit	Monate	44.4%	quadratisch	3/60	48.1	43.4	26.6	30.8
2 Langfristige Lebensqualität	Utilities gemäss Liste	26.0%	quadratisch	10/80	80.0	26.0	75.0	25.4
3 Schwere unerwünschte Ereignisse	Anteil Patienten	13.8%	linear	5/80	48.0	5.9	64.0	2.9
4 Behandlungshäufigkeit	in Tagen	3.4%	linear	1/30	14.5	1.6	1.5	0.1
5 Behandlung zu Hause	1: ja, 0: nein	5.3%	linear	0/1	0.0	0.0	1.0	5.3
6 Krankheitslast in der Gesellschaft	DALY/100k EW/Jahr	2.4%	quadratisch	46/920	46.0	2.4	46.0	2.4
7 Indirekte Kosten Gesellschaft	Monate	0.5%	linear	1/6	1.0	0.5	1.0	0.5
8 Weitere Behandlungskosten	kCHF	4.2%	linear	10/80	72.1	0.5	73.3	0.4
Summe		100%				80		68

Dargestellt ist die Bewertung von zwei Behandlungsoptionen für fortgeschrittene Nierenzellkarzinome. Nivolumab plus Ipilimumab ist eine innovative Behandlungsmethode, die Behandlung mit Sunitinib entspricht dem Standard of Care und dient als Komparator. Die Zeilen enthalten die berücksichtigten Attribute, in welchen Einheiten wir diese messen, wieviel Gewicht sie in unserem Modell erhalten, die Angabe, ob wir eine lineare oder quadratische Wertefunktion unterstellen und für welchen Eingabebereich wir das Modell konfiguriert haben (Min und Max).

Für beide Behandlungsoptionen ist angegeben, welchen Wert wir jeweils eingegeben haben. Daneben erscheint der zugehörige Wert der Einzelnutzen. (Die Eingabe von 48.1 Monaten Medianüberlebenszeit etwa übersetzen sich in 43.4 gewichtete Nutzenwerte.) Die unterste Zeile stellt die Summe der Einzelnutzen dar. Demnach kommt die Innovation auf 80 und der Komparator auf 68 Punkte. Theoretisches Maximum sind 100 Punkte.

Caveat: Das Modell ist nur so gut wie sein Input. Zwar verwenden wir eine Gewichtung die den Präferenzen der Schweizer Bevölkerung entsprechen. Andere Aspekte bleiben aber weniger gut abgesichert: Teilweise lagen die Eingabewerte ausserhalb der definierten Wertebereiche (Innovation mit Lebensqualität 82, Krankheitslast eher bei 31 als bei 46) oder die Datenqualität selbst ist nicht optimal (Lebensqualität nicht lange genug beobachtbar). Unser Vorschlag ist also noch nicht an allen Stellen bis ins letzte Detail ausgearbeitet. Deshalb sollte die absolute Höhe der Werte nicht direkt interpretiert werden – wir empfehlen stattdessen eher die Betrachtung der Nutzendifferenz in Kategorien.

Quelle: Eigene Berechnungen gemäss oben beschriebenem Modell und seiner Kalibrierung, Polynomics.

Sensitivitätsanalyse

Ein Teil der gerade genannten Einschränkungen lässt sich mittels Sensitivitätsanalysen darstellen oder abfangen. So könnte beispielsweise die Unsicherheit über die Entwicklung der langfristigen Lebensqualität, für die eine zu kurze Datenreihe vorliegt, durch (aus Sicht der Innovation) optimistische oder pessimistische Szenarien variiert werden. Dies wäre ein Beispiel für stochastische

Unsicherheit (vgl. Kapitel 3.7). Wird das Modell mit diesen Werten erneut befüllt und eine Gesamtbewertung erstellt, kann so abgeschätzt werden, welche Rolle die mit der Bestimmung der langfristigen Lebensqualität zusammenhängende Unsicherheit spielt. In einem «optimistischen Szenario» etwa könnte unterstellt werden, dass der Wert für den Komparator 70 (statt 75) ist, in einem «pessimistischen Szenario» könnte sich der Wert des Komparators dem der Innovation angleichen (auf 80). Die Auswertung ergibt, dass der Zusatznutzen der Innovation zwischen 13 Punkten («optimistisches Szenario»: 80 vs. 67) und 12 Punkten («pessimistisches Szenario»: 80 vs. 68) beträgt.

Eine ähnliche Sensitivität kann auch mithilfe der nachträglichen Erweiterung des zulässigen Wertebereichs für die langfristige Lebensqualität ermittelt werden. Dafür ist einerseits die ermittelte Wertefunktion fortzuschreiben und gleichzeitig die Funktion neu zu skalieren, um die ursprüngliche Gewichtung mit 26 Prozent beizubehalten. Erhöht man den maximal zulässigen Wert dieses Attributs zum Beispiel von 80 auf 90, können wir für die Behandlung mit Nivolumab und Ipilimumab neu den tatsächlichen Wert von 82 eingeben. (Der Wert von 90 entspricht dem durchschnittlichen Wert der Lebensqualität in der Gesamtbevölkerung, vgl. Abschnitt 3.4.1.) Dies verschiebt die Grenze, ab der der volle Nutzenwert von 26 Punkten vergeben wird. Beide Behandlungsmethoden erhalten also einen niedrigeren Wert. Die Behandlung mit Nivolumab und Ipilimumab erhält noch 25.5 statt 26 Punkte, die mit Sunitinib 24.7 statt 25.4. Der Abschlag für die Innovation ist nun geringer, weil ihr Wert nicht mehr auf 80 limitiert ist und sie ihren tatsächlichen Nutzen voll abbilden kann, sie also nicht mehr von einem Ceiling-Effekt betroffen ist. Insgesamt erhöht sich die Nutzendifferenz um einen Punkt von 12 auf 13.

Als Beispiel für strukturelle Unsicherheit, rechnen wir eine Sensitivitätsanalyse unter Verwendung alternativer Spezifikationen für die nichtlinearen Wertefunktionen. Nehmen wir statt der quadratischen Form $y_i = a(x_i)^2 + bx_i + c$ eine Wurzelfunktion mit der Form $y_i = a\sqrt{x_i} + bx_i + c$ an, führt dies zu einer leichten Verschiebung der Gewichte (da sich Minima und Maxima innerhalb der jeweiligen Wertebereiche leicht ändern) und zu einer deutlichen Veränderung der durch die drei Punktpaare gelegten Wertefunktionen. Implementiert man diese Form, kommen für den hier betrachteten Fall des Nierenzellkarzinoms leicht niedrigere Werte heraus: Die Behandlung mit Nivolumab plus Ipilimumab erhält noch 79 Punkte (-1), die mit Sunitinib unverändert 68. Der Nutzengewinn verkleinert sich also minimal um 1 Punkt. An der Bewertung in Nutzenkategorien ändert sich nichts.

Ein weiteres Thema, welches Teil einer Sensitivitätsanalyse sein kann, ist die Untersuchung von fehlenden oder unvollständigen Inputdaten. In unserem Beispiel haben wir keine Werte zu den indirekten Kosten der beiden Arzneimittel. Eine Möglichkeit wäre hier, ein minimales und ein maximales Szenario zu rechnen. Das minimale Szenario (aus Sicht der Innovation) entspräche einem Wert von einem Monat Arbeitsausfall bei der Komparatortherapie und sechs Monaten Arbeitsausfall bei der Innovation. Das maximale Szenario entspräche gerade dem Gegenteil. Da die indirekten Kosten nur ein Gewicht von 0.5 Prozent aufweisen, sind die Auswirkungen eher gering: Das minimale Szenario resultiert in einer Nutzendifferenz von 13 Punkten, das maximale in 12 Punkten.

Kategorisierung

Abgesehen von den oben genannten, hauptsächlich dem Erhebungsdesign respektive der Stichprobenwahl geschuldeten Einschränkungen muss stets bedacht werden, dass die vorliegende Nutzbewertung auf subjektiven Werturteilen basiert, die immer eine gewisse Unschärfe implizier-

ren. Die resultierenden Nutzenwerte sollten deshalb nicht in ihrer absoluten Form direkt interpretiert werden. In Kapitel 3.6 schlagen wir eine grobe Kategorisierung der Nutzenverbesserung vor. Gemäss unserem Vorschlag in Tabelle 9 würde die hier bewertete Innovation als «grosse Nutzenverbesserung» kategorisiert. Die Nutzenverbesserung der Sensitivitätsanalyse resultierte ebenfalls immer in dieser Kategorie.

5 Fazit

Im vorliegenden Projekt konnten wir aufzeigen, dass ein Instrument zur standardisierten Nutzenbewertung von Onkologika mittels einer Multi-Criteria Decision Analysis (MCDA) für die Schweiz umsetzbar ist. Ein einziges generisches Modell für alle Krebsmedikamente ist zwar aufgrund der unterschiedlichen Ziele und damit verbundenen Nutzenbewertungen kaum möglich. Generische Modelle für wenige Kategorien wie adjuvante und potenziell kurative Therapien, nichtkurative Therapien sowie Therapien gegen seltene Krankheiten sollten jedoch umsetzbar sein. Im Fall der nichtkurativen Therapien haben wir dies mit einem konkreten Vorschlag veranschaulicht und für den Fall einer spezifischen Innovation die Umsetzung exemplarisch durchgeführt.

Die Resonanz zum hier vorgestellten Nutzenbewertungssystem war bei den konsultierten Experten grossmehrheitlich positiv. Den Nutzenbegriff breiter zu fassen und insbesondere auch den sozialen Nutzen in die Bewertung einfließen zu lassen, wurde als äusserst relevant erachtet. Auch wenn im Einzelfall die Meinungen der Experten in Details nicht übereinstimmten, konnten alle wichtigen Stellschrauben des Modells in einer Fokusgruppe mit Experten der wichtigsten Anspruchsgruppen im Konsens bestimmt werden.

Das vorgeschlagene Modell ist als Hilfsmittel zur Bestimmung der Kostenübernahme durch die OKP gedacht, ist jedoch nicht zur spezifischen Preissetzung konzipiert. Grundsätzlich lässt sich das Modell aber überall verwenden, wo Nutzenbewertungen von Arzneimitteln im regulatorischen Kontext gebraucht werden, z. B. beim therapeutischen Quervergleich (TQV), bei Priorisierungen, bei wertebasierten Vergütungsmodellen wie Value Based Pricing. Das Konzept ist gut anpassbar und erlaubt für solche Anwendungsformen den einfachen Einbezug weiterer Nutzendimensionen.

Unser Vorschlag ist ein Machbarkeitsnachweis. Naturgemäss ist er noch nicht an allen Stellen bis ins letzte Detail ausgearbeitet. Für die Umsetzung in der Praxis sind trotz inzwischen erfolgter Bevölkerungsbefragung immer noch einige dieser Detailarbeiten zu erledigen. So empfehlen wir, das Modell für die Praxis final auszuarbeiten um danach Praxiserfahrung zu sammeln. Denkbar ist auch eine Quervalidation des Modells mit vergangenen Vergütungsentscheiden des BAG oder anderen Bewertungsrahmen (z. B. ESMO MCBS). Das vorgeschlagene Modell wurde zwar in Expertengesprächen und einer Experten-Fokusgruppe besprochen und validiert, dennoch lässt sich nicht ausschliessen, dass sich gewisse Aspekte in der Zukunft ändern und Anpassungen am Modell notwendig werden. Solche Anpassungen am Modell sollen im Sinne eines lernenden Systems möglich sein. Aufgrund der Erfahrungen des Modells bei nichtkurativen Krebstherapien sollen dann in einem weiteren Schritt die zwei Modelle zu kurativen Krebstherapien und Therapien gegen seltene Krebserkrankungen ausgearbeitet werden.

6 Bibliografie

- Abbas, A.E., 2018. Foundations of Multiattribute Utility, 1st ed. Cambridge University Press.
<https://doi.org/10.1017/9781316596739>
- Abdullah, A.H., Holtorf, A.-P., Al-Hussaini, M., Lemay, J., Alowayesh, M., Kaló, Z., 2019. Stakeholder driven development of a multi-criteria decision analysis tool for purchasing off-patent pharmaceuticals in Kuwait. *J. Pharm. Policy Pract.* 12, 9.
<https://doi.org/10.1186/s40545-019-0171-4>
- Albiges, L., Tannir, N.M., Burotto, M., McDermott, D., Plimack, E.R., Barthélémy, P., Porta, C., Powles, T., Donskov, F., George, S., Kollmannsberger, C.K., Gurney, H., Grimm, M.-O., Tomita, Y., Castellano, D., Rini, B.I., Choueiri, T.K., Saggi, S.S., McHenry, M.B., Motzer, R.J., 2020. Nivolumab plus ipilimumab versus sunitinib for first-line treatment of advanced renal cell carcinoma: extended 4-year follow-up of the phase III CheckMate 214 trial. *ESMO Open* 5, e001079. <https://doi.org/10.1136/esmoopen-2020-001079>
- Angelis, A., Kanavos, P., 2017. Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The Advance Value Framework. *Soc. Sci. Med.* 188, 137–156. <https://doi.org/10.1016/j.socscimed.2017.06.024>
- Angelis, A., Linch, M., Montibeller, G., Molina-Lopez, T., Zawada, A., Orzel, K., Arickx, F., Espin, J., Kanavos, P., 2020. Multiple Criteria Decision Analysis for HTA across four EU Member States: Piloting the Advance Value Framework. *Soc. Sci. Med.* 246, 112595.
<https://doi.org/10.1016/j.socscimed.2019.112595>
- Baltussen, R., Marsh, K., Thokala, P., Diaby, V., Castro, H., Cleemput, I., Garau, M., Iskrov, G., Olyaeemanesh, A., Mirelman, A., Mobinizadeh, M., Morton, A., Tringali, M., van Til, J., Valentim, J., Wagner, M., Youngkong, S., Zah, V., Toll, A., Jansen, M., Bijlmakers, L., Oortwijn, W., Broekhuizen, H., 2019. Multicriteria Decision Analysis to Support Health Technology Assessment Agencies: Benefits, Limitations, and the Way Forward. *Value Health* 22, 1283–1288. <https://doi.org/10.1016/j.jval.2019.06.014>
- Bana e Costa, C.A., Carnero, M.C., Oliveira, M.D., 2012. A multi-criteria model for auditing a Predictive Maintenance Programme. *Eur. J. Oper. Res.* 217, 381–393.
<https://doi.org/10.1016/j.ejor.2011.09.019>
- Baran-Kooiker, A., Czech, M., Kooiker, C., 2018. Multi-Criteria Decision Analysis (MCDA) Models in Health Technology Assessment of Orphan Drugs—a Systematic Literature Review. Next Steps in Methodology Development? *Front. Public Health* 6.
<https://doi.org/10.3389/fpubh.2018.00287>
- Becker, K., 2006. Flexibilisierungsmöglichkeiten in der Krankenversicherung. Verlag Dr. Kovač, Hamburg.
- Belton, V., Stewart, T.J., 2002. Multiple Criteria Decision Analysis. Springer US, Boston, MA.
<https://doi.org/10.1007/978-1-4615-1495-4>
- Ben-Akiva, M., Lerman, S.R., 1985. Discrete Choice Analysis. MIT Press, Cambridge, MA.
- Bill, M., Kuhlmeier, F., Telser, H., 2020. Onkologika: Innovation kostet, stiftet aber auch Nutzen. Polynomics, Olten.

- Briggs, A.H., Weinstein, M.C., Fenwick, E.A.L., Karnon, J., Sculpher, M.J., Paltiel, A.D., 2012. Model Parameter Estimation and Uncertainty Analysis: A Report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group–6. *Med. Decis. Making* 32, 722–732. <https://doi.org/10.1177/0272989X12458348>
- Brixner, D., Maniadakis, N., Kaló, Z., Hu, S., Shen, J., Wijaya, K., 2017. Applying Multi-Criteria Decision Analysis (MCDA) Simple Scoring as an Evidence-based HTA Methodology for Evaluating Off-Patent Pharmaceuticals (OPPs) in Emerging Markets. *Value Health Reg. Issues* 13, 1–6. <https://doi.org/10.1016/j.vhri.2017.02.001>
- Cella, D., Grünwald, V., Escudier, B., Hammers, H.J., George, S., Nathan, P., Grimm, M.-O., Rini, B.I., Doan, J., Ivanescu, C., Paty, J., Mekan, S., Motzer, R.J., 2019. Patient-reported outcomes of patients with advanced renal cell carcinoma treated with nivolumab plus ipilimumab versus sunitinib (CheckMate 214): a randomised, phase 3 trial. *Lancet Oncol.* 20, 297–310. [https://doi.org/10.1016/S1470-2045\(18\)30778-2](https://doi.org/10.1016/S1470-2045(18)30778-2)
- Ezeife, D.A., Dionne, F., Fares, A.F., Cusano, E.L.R., Fazelzad, R., Ng, W., Husereau, D., Ali, F., Sit, C., Stein, B., Ellis, P.M., Berry, S., Peacock, S., Mitton, C., Earle, C.C., Chan, K.K.W., Leighl, N.B., 2020. Value assessment of oncology drugs using a weighted criterion-based approach. *Cancer* 126, 1530–1540. <https://doi.org/10.1002/cncr.32639>
- Franco, L.A., Montibeller, G., 2011. Problem Structuring for Multicriteria Decision Analysis Interventions, in: *Wiley Encyclopedia of Operations Research and Management Science*. John Wiley & Sons, Inc., Hoboken, NJ, USA, p. eorms0683. <https://doi.org/10.1002/9780470400531.eorms0683>
- Fries, R., Hummel, Y., Kipfer, B., Schafroth, U., Seiler, B., Zollikofer, J., 2018. OLUtool als neues Einheitsmodell der Vertrauensärzte. *Schweiz. Ärztsztg.* 99, 1571–1572. <https://doi.org/10.4414/saez.2018.17190>
- Garau, M., 2018. Applying a Multicriteria Decision Analysis (MCDA) Approach to Elicit Stakeholders' Preferences in Italy: The Case of Obinutuzumab for Rituximab-Refractory Indolent Non-Hodgkin Lymphoma (iNHL). *PharmacoEconomics Open* 153–163. <https://doi.org/10.1007/s41669-017-0048-x>
- Ghabri, S., Josselin, J.-M., Le Maux, B., 2019. Could or Should We Use MCDA in the French HTA Process? *PharmacoEconomics* 37, 1417–1419. <https://doi.org/10.1007/s40273-019-00846-w>
- Hsu, J.C., Lin, J.-Y., Lin, P.-C., Lee, Y.-C., 2019. Comprehensive value assessment of drugs using a multi-criteria decision analysis: An example of targeted therapies for metastatic colorectal cancer treatment. *PLOS ONE* 14, e0225938. <https://doi.org/10.1371/journal.pone.0225938>
- Inotai, A., Brixner, D., Maniadakis, N., Dwiprahasto, I., Kristin, E., Prabowo, A., Yasmina, A., Priohutomo, S., Németh, B., Wijaya, K., Kalo, Z., 2018. Development of multi-criteria decision analysis (MCDA) framework for off-patent pharmaceuticals – an application on improving tender decision making in Indonesia. *BMC Health Serv. Res.* 18, 1003. <https://doi.org/10.1186/s12913-018-3805-3>
- Janssen, M.F., Szende, A., Cabases, J., Ramos-Goñi, J.M., Vilagut, G., König, H.H., 2019. Population norms for the EQ-5D-3L: a cross-country analysis of population surveys for 20 countries. *Eur. J. Health Econ.* 20, 205–216. <https://doi.org/10.1007/s10198-018-0955-5>

- Johnson, F.R., Desvousges, W.H., 1997. Estimating Stated Preferences with Rated-Pair Data: Environmental, Health, and Employment Effects of Energy Programs. *J. Environ. Econ. Manag.* 34, 79–99.
- Kägi, W., Frey, M., Möhr, T., Bollag, Y., Brugger, C., 2020. Evaluation der Vergütung von Arzneimitteln im Einzelfall nach den Art. 71a–71d KVV (Evaluation im Auftrag des Bundesamts für Gesundheit). BSS Volkswirtschaftliche Beratung und asim, Basel.
- Kigozi, J., Jowett, S., Lewis, M., Barton, P., Coast, J., 2016. Estimating productivity costs using the friction cost approach in practice: a systematic review. *Eur. J. Health Econ.* 17, 31–44.
- König, H.-H., Bernert, S., Angermeyer, M.C., Matschinger, H., Martinez, M., Vilagut, G., Haro, J.M., de Girolamo, G., de Graaf, R., Kovess, V., Alonso, J., Investigators, the Esem. 2000, 2009. Comparison of Population Health Status in Six European Countries: Results of a Representative Survey Using the EQ-5D Questionnaire. *Med. Care* 47, 255–261. <https://doi.org/10.1097/MLR.0b013e318184759e>
- Kwon, S.-H., 2017. Eliciting societal preferences of reimbursement decision criteria for anti-cancer drugs in South Korea. *Expert Rev. Pharmacoecon. Outcomes Res.* 17, 411–419. <https://doi.org/10.1080/14737167.2017.1277144>
- Laba, T.-L., Jiwani, B., Crossland, R., Mitton, C., 2020. Can multi-criteria decision analysis (MCDA) be implemented into real-world drug decision-making processes? A Canadian provincial experience. *Int. J. Technol. Assess. Health Care* 1–6. <https://doi.org/10.1017/S0266462320000525>
- Lancaster, K.J., 1966. A New Approach to Consumer Theory. *J. Polit. Econ.* 74, 132–157.
- Louviere, J.J., Hensher, D., Swait, J., 2000. Stated choice methods: analysis and applications. Cambridge University Press, Cambridge.
- Luce, R.D., Tukey, J.W., 1964. Simultaneous conjoint measurement: A new type of fundamental measurement. *J. Math. Psychol.* 1, 1–27. [https://doi.org/10.1016/0022-2496\(64\)90015-X](https://doi.org/10.1016/0022-2496(64)90015-X)
- Marsh, K., IJzerman, M., Thokala, P., Baltussen, R., Boysen, M., Kaló, Z., Lönnngren, T., Mus-sen, F., Peacock, S., Watkins, J., Devlin, N., 2016. Multiple Criteria Decision Analysis for Health Care Decision Making—Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health* 19, 125–137. <https://doi.org/10.1016/j.jval.2015.12.016>
- Marsh, K.D., Sculpher, M., Caro, J.J., Tervonen, T., 2018. The Use of MCDA in HTA: Great Potential, but More Effort Needed. *Value Health* 21, 394–397. <https://doi.org/10.1016/j.jval.2017.10.001>
- Mas-Colell, A., Whinston, M.D., Green, J.R., 1995. Microeconomic theory. Oxford university press New York.
- McFadden, D., 2001. Economic Choices. *Am. Econ. Rev.* 91, 351–378.
- McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior, in: Zarembka, P. (Ed.), *Frontiers in Econometrics*. Academic Press, New York, pp. 105–142.
- Motzer, R.J., Tannir, N.M., McDermott, D.F., Arén Frontera, O., Melichar, B., Choueiri, T.K., Plimack, E.R., Barthélémy, P., Porta, C., George, S., Powles, T., Donskov, F., Neiman, V., Kollmannsberger, C.K., Salman, P., Gurney, H., Hawkins, R., Ravaud, A., Grimm, M.-O.,

- Bracarda, S., Barrios, C.H., Tomita, Y., Castellano, D., Rini, B.I., Chen, A.C., Mekan, S., McHenry, M.B., Wind-Rotolo, M., Doan, J., Sharma, P., Hammers, H.J., Escudier, B., 2018. Nivolumab plus Ipilimumab versus Sunitinib in Advanced Renal-Cell Carcinoma. *N. Engl. J. Med.* 378, 1277–1290. <https://doi.org/10.1056/NEJMoa1712126>
- Oliveira, M.D., Mataloto, I., Kanavos, P., 2019. Multi-criteria decision analysis for health technology assessment: addressing methodological challenges to improve the state of the art. *Eur. J. Health Econ.* 20, 891–918. <https://doi.org/10.1007/s10198-019-01052-3>
- Oniangue-Ndza, C., Schneider, R., Malcolm, B., May, J., Gooden, K.M., Klijn, S., Cakar, E., 2019. Cost-effectiveness of Nivolumab in combination with Ipilimumab compared to Sunitinib for the first-line treatment of advanced/metastatic intermediate- or poor-risk renal carcinoma in Switzerland. *Value Health* 22, S462. <https://doi.org/10.1016/j.jval.2019.09.336>
- Parlamentarische Verwaltungskontrolle, 2013. Evaluation der Zulassung und Überprüfung von Medikamenten in der obligatorischen Krankenpflegeversicherung (Evaluationsbericht zuhanden der Geschäftsprüfungskommission des Ständerats). Parlamentarische Verwaltungskontrolle (PVK), Bern.
- Pike, J., Grosse, S.D., 2018. Friction Cost Estimates of Productivity Costs in Cost-of-Illness Studies in Comparison with Human Capital Estimates: A Review. *Appl. Health Econ. Health Policy.* 16, 765–778. <https://doi.org/10.1007/s40258-018-0416-4>
- Richardson, J., Iezzi, A., Khan, M.A., 2015. Why do multi-attribute utility instruments produce different utilities: the relative importance of the descriptive systems, scale and ‘micro-utility’ effects. *Qual. Life Res.* 24, 2045–2053. <https://doi.org/10.1007/s11136-015-0926-6>
- Schey, C., Postma, M.J., Krabbe, P.F.M., Topachevskyi, O., Volovyk, A., Connolly, M., 2020. Assessing the Preferences for Criteria in Multi-Criteria Decision Analysis in Treatments for Rare Diseases. *Front. Public Health* 8, 162. <https://doi.org/10.3389/fpubh.2020.00162>
- Szende, A., Janssen, B., Cabases, J. (Eds.), 2014. Self-Reported Population Health: An International Perspective based on EQ-5D. Springer Netherlands. <https://doi.org/10.1007/978-94-007-7596-1>
- Telser, H., 2002. Nutzenmessung im Gesundheitswesen: Die Methode der Discrete-Choice-Experimente. Verlag Dr. Kovač, Hamburg.
- Tervonen, T., Gelhorn, H., Bhashyam, S.S., Poon, J.-L., Gries, K.S., Rentz, A., Marsh, K., 2017. MCDA swing weighting and discrete choice experiments for elicitation of patient benefit-risk preferences: a critical assessment. *Pharmacoepidemiol. Drug Saf.* 26, 1483–1491. <https://doi.org/10.1002/pds.4255>
- Thokala, P., Devlin, N., Marsh, K., Baltussen, R., Boysen, M., Kalo, Z., Longrenn, T., Mussen, F., Peacock, S., Watkins, J., Ijzerman, M., 2016. Multiple Criteria Decision Analysis for Health Care Decision Making—An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health* 19, 1–13. <https://doi.org/10.1016/j.jval.2015.12.003>
- Train, K., 2008. Discrete choice methods with simulation, repr. ed. Cambridge Univ. Press, Cambridge.

- von Wyl, V., Telser, H., Weber, A., Fischer, B., Beck, K., 2015. Cost trajectories from the final life year reveal intensity of end-of-life care and can help to guide palliative care interventions. *BMJ Support. Palliat. Care* 8, 325–334.
- Wagner, M., 2017. Appraising the holistic value of Lenvatinib for radio-iodine refractory differentiated thyroid cancer: A multi-country study applying pragmatic MCDA. *BMC Cancer* 17, 12. <https://doi.org/10.1186/s12885-017-3258-9>

Anhang A: Beteiligte Akteure

Bei der Erstellung unseres Bewertungsmodells waren wir neben dem Literaturstudium und unserer Expertise auf externes Fachwissen angewiesen. Nur so konnten wir sicherstellen, ein ausgewogenes, belastbares und gleichzeitig realisierbares Modell zu entwickeln. Für die Erstellung des Wertebaums, also für die Bestimmung der Nutzenkomponenten, haben wir bilaterale Expertengespräche mit folgenden Personen geführt:

- PD Dr. med. Antonia Maria Müller
- Dr. Heiner Sandmeier
- Prof. Dr. Matthias Schwenkglenks
- Prof. Dr. Michael Schlander
- Prof. Dr. Roger von Moos

Für die Validierung unseres finalen Wertebaums und für die Spezifizierungen der nachfolgenden Schritte arbeiteten wir mit einer multidisziplinären Fokusgruppe zusammen, die sich folgendermassen konstituierte:

- lic. iur. Andreas Faller
- Dr. med. Robert Fries
- David Haerry
- PD Dr. med. Antonia Maria Müller
- Dr. Andreas Schiesser
- Prof. Dr. Matthias Schwenkglenks
- Prof. Dr. Michael Schlander

Anhang B: Methoden zur Präferenzmessung

Tabelle 19 Methoden zur Präferenzmessung

	Methode	Beschreibung	Vorteile	Nachteile
Dekompositionelle Methoden	Discrete-Choice-Experiment (DCE)	Wahlbasiert. Die Befragten müssen in mehreren Entscheidungssets jeweils die bessere von zwei gegenübergestellten Alternativen auswählen.	Wertung und Gewichtung können im gleichen Framework ermittelt werden	- Nur begrenzte Anzahl Attribute abfragbar (kognitive Überforderung) - Relativ grosse Stichprobe notwendig
	DCE mit Best-Worst Scaling Technique	Analog zu DCE. Die Befragten müssen zusätzlich bei jeder Alternative sagen, was sie daran am besten und am schlechtesten finden.	- Zusätzliche Einsichten im Vergleich zu herkömmlichem DCE. Einfluss der Attribute - Potenziell einfacher für Befragte	- Nur begrenzte Anzahl Attribute abfragbar (kognitive Überforderung) - Relativ grosse Stichprobe notwendig
Kompositionelle Methoden	Direct Rating – Scaling	Die Wichtigkeit jedes Attributs wird auf einer vordefinierten Skala (etwa zwischen 1 und 5) unabhängig der anderen Attribute bewertet und danach in standardisierte Gewichte umgerechnet.	Einfach für die Befragten	- Kein Abwägen von Veränderungen zwischen Kriterien (Trade-offs) - Verteilung der Gewichte sehr flach
	Analytical Hierarchy Process (AHP)	Paarweiser Vergleich von Attributen oder Kriterien und derer «Intensität der Wichtigkeit» auf einer Verhältnisskala von 1 (gleich wichtig) bis 9 (eindeutig viel wichtiger). Die Resultate werden in einer Matrix zusammengefasst und mithilfe der Eigenvektor-Methode kondensiert.	- In kleinen Gruppen durchführbar - Kognitive Belastung für Befragte vertretbar - Keine Transitivitätsannahme, dafür kann die Rangfolge bei Einführung neuer Attribute ändern	- Muss für Bewertung und Gewichtung einzeln durchgeführt werden - Kein Abwägen von Veränderungen zwischen Kriterien (Trade-offs)
	MACBETH	Paarweiser Vergleich von Attributen oder Kriterien und Bewertung derer relativen Wichtigkeit auf einer semantischen Skala (7 Stufen)	- Ermöglicht Übertragung einer qualitativen Bewertung auf eine Kardinalskala - In kleinen Gruppen durchführbar - Kognitive Belastung für Befragte vertretbar	- Muss für Bewertung und Gewichtung einzeln durchgeführt werden - Kein Abwägen von Veränderungen zwischen Kriterien (Trade-offs)
	Simple Multi-Attribute Rating Technique (SMAR)T	Die Attribute werden zuerst nach ihrer Wichtigkeit geordnet. Danach wird entweder dem wichtigsten Attribut die Zahl 100 oder dem am wenigsten wichtigen Attribut die Zahl 10 zugewiesen. Die Wichtigkeit der anderen Attribute wird nun relativ zum Referenzattribut bewertet.		Nur für Gewichtung anwendbar

Methode	Beschreibung	Vorteile	Nachteile
Swing Weighting (z. B. SMARTS)	Swing Weighting ist eine Weiterentwicklung von SMART. Hier wird zweistufig vorgegangen: Zuerst wird eine Rangfolge der Attribute vom Wichtigsten zum Unwichtigsten erstellt und im zweiten Schritt werden die Gewichte selbst ermittelt. Begonnen wird bei SMART mit der Alternative, in welcher alle Attribute auf dem schlechtesten Wert sind. Danach wird dasjenige Attribut ausgewählt, für welches ein Wechsel zur maximalen Ausprägung den grössten Nutzenzuwachs bedeutet. Diesem wird ein Nutzenwert von 100 gegeben. Von diesem Zustand ausgehend, wird nun aus den noch nicht veränderten Attributen dasjenige ausgewählt, für welches ein Wechsel auf die beste Ausprägung den grössten (verbleibenden) Nutzengewinn bedeutet. Für diese Veränderung wird anschliessend gefragt, wie gross der Nutzenzugewinn relativ zum ersten Attribut (mit Gewicht 100) ist. So «Schwingen» nach und nach alle Attribute gemäss ihrem Nutzenpotenzial vom schlechtesten zum besten Wert. Anschliessend werden die Nutzenwerte in Gewichte normalisiert.	- Trade-off-Entscheidungen - Stellt sicher, dass die Gewichte die Bedeutung und die Spannweite der Attribute berücksichtigen	- Nur für Gewichtung anwendbar - Restriktive Annahmen: (1) Monotonie in der Bewertung der Attribute («mehr ist immer besser oder immer schlechter») und (2) «bedingte Monotonie», womit die Unabhängigkeit der Attribute gemeint ist (die Bewertung von X hängt nicht von der Ausprägung von Y ab)
Scoring Functions	Das Maximum und Minimum des Möglichkeitsraums werden als 0 und 100 angenommen. 1: Zweiteilung: Für welchen Wert ist der Abstand vom Maximum und vom Minimum identisch? Das heisst, welcher Wert im Möglichkeitsraum entspricht 50? Danach kann in den Unterabschnitten der gleiche Prozess nochmals durchgeführt werden. 2: Differenz: Welcher Wert auf der Skala von 0 bis 100 entspricht dem mittleren Wert des Möglichkeitsraums?	- Einfach umsetzbar und verständlich für Befragte - Anpassbar an geforderte Präzision - Anpassbar an Ausprägungen des Möglichkeitsraums	- Nur für Bewertung - Kein Abwägen von Veränderungen zwischen Kriterien (Trade-offs).
Kepner-Tregoe Analysis	Zuerst wird das wichtigste Attribut identifiziert und der Wert 10 zugewiesen. Alle anderen Attribute werden relativ zum wichtigsten auf einer Skala von 10 (gleich wichtig) bis 1 (nicht sehr wichtig) bewertet. Die Skala kann danach zwischen 0 und 1 normalisiert werden.		Nur für Gewichtung anwendbar

Quelle: Eigene Darstellung in Anlehnung an Marsh et al. (2016).

Anhang C: Discrete-Choice-Experimente

Grundprinzip von Discrete-Choice-Experimenten

In einem Discrete-Choice-Experiment werden Produkte gemäss der Nachfragetheorie von Lancaster (1966) durch verschiedene Attribute beschrieben, die jeweils unterschiedliche Ausprägungen annehmen können. Im Gegensatz zur klassischen Konsumtheorie, die davon ausgeht, dass ein Gut als Ganzes bewertet wird und der Nutzen eines Gutes umso höher ist, je mehr Einheiten davon konsumiert werden, ziehen die Konsumenten in der alternativen Nachfragetheorie von Lancaster einen Nutzen aus den einzelnen Attributen eines Produktes. Bei der Konsumententscheidung bewertet ein Konsument die einzelnen Attribute und vergleicht sie mit den Attributsausprägungen alternativer Produkte. Dabei kommt es zu Abwägungen zwischen den verschiedenen Attributen, die individuell unterschiedlich sind. Ein Konsument wird sich für ein Produkt entscheiden, wenn der Gesamtnutzen aus den Attributen höher ist als bei den Alternativen. Dieser Ansatz erlaubt es, aus den in einem kontrollierten Experiment getroffenen Entscheidungen abzuleiten, wie gross der Nutzen der einzelnen Attribute ist und welche Austauschverhältnisse die Konsumenten zwischen den Attributen machen, das heisst, wie viel sie von einem Attribut aufzugeben bereit sind, wenn sie von einer anderen mehr erhalten. Werden die Produkte im Experiment jeweils mit einem Preis versehen, lässt sich der Nutzen monetär in Form von Zahlungsbereitschaften bewerten und vergleichen.

Mit einem DCE lassen sich insbesondere die Präferenzen für Produkte oder Leistungen bewerten, die nicht – oder noch nicht – am Markt gehandelt werden. Folglich eignet sich die Methode im Marketingbereich bei der Produkt- und Preisgestaltung, aber auch zur Bewertung von öffentlichen Gütern. Aufgrund der theoretischen Fundierung (vgl. nächster Abschnitt) lassen sich DCE sehr flexibel gestalten und ermöglichen so umfangreiche und detaillierte Analysen der Bestimmungsgründe von Entscheidungen sowie die Bewertung von einzelnen Attributen und Gesamtprodukten.

Theoretische Fundierung und statistische Auswertung von Discrete-Choice-Experimenten

Der statistischen Auswertung von DCE liegt ein entscheidungstheoretisches, mikroökonomisch fundiertes Modell, das sogenannte «Random-Utility-Modell», zugrunde (McFadden, 2001, 1974). Das Modell unterstellt, dass die Entscheidungen eines Individuums zwar deterministisch sind, der Untersuchende aber nicht alle Aspekte, die eine Entscheidung beeinflussen, vollständig beobachten kann. Mit anderen Worten entscheiden sich die Individuen jeweils für die Alternative, die ihnen den höchsten Nutzen garantiert, der Beobachter kennt diesen Nutzen aber nicht mit Sicherheit. Das Modell enthält folglich eine stochastische Komponente und es kann lediglich eine Wahrscheinlichkeit (P_{im}) angegeben werden, mit der eine befragte Person i sich für die Alternative m gegenüber j entscheidet:

$$P_{im} = \text{Prob}[v(p_m, z_m, s_i, \varepsilon_{im}) \geq v(p_j, z_j, s_i, \varepsilon_{ij})].$$

Der Nutzen der Alternativen $v_m(\cdot)$ beziehungsweise $v_j(\cdot)$ ist dabei eine Funktion des Preises p , der übrigen Attribute des Produkts, dargestellt im Vektor z , und den sozioökonomischen Eigenschaften s_i des Individuums. Die Zufallsvariable, welche die unbeobachtbaren Faktoren enthält, ist durch den Vektor ε abgebildet (Störterm). Eine Standardannahme im Random-Utility-Modell

ist, dass die bedingte Nutzenfunktion als Summe von beobachtbaren (deterministischen) und unbeobachtbaren (stochastischen) Komponenten des Gesamtnutzens dargestellt werden kann (vgl. Ben-Akiva and Lerman, 1985, Kap. 3). Somit lässt sich obige Gleichung umformen:

$$\begin{aligned} P_{im} &= \text{Prob}[w(p_m, z_m, s_i) + \varepsilon_{im} \geq w(p_j, z_j, s_i) + \varepsilon_{ij}, \forall j \neq m] \\ &= \text{Prob}[\varepsilon_{ij} - \varepsilon_{im} \leq w(p_m, z_m, s_i) - w(p_j, z_j, s_i), \forall j \neq m]. \end{aligned}$$

Um das Modell empirisch schätzen zu können, ist zum einen eine Annahme über die Verteilung der Störterme ε zu treffen. Hier werden typischerweise Normal- oder Extremwertverteilungen unterstellt. Zum anderen ist eine Annahme über die funktionale Form der deterministischen Komponente $w(\cdot)$ der Nutzenfunktion erforderlich. In den meisten praktischen Anwendungen werden lineare Approximationen der Nutzenfunktion verwendet (vgl. z. B. Johnson and Desvousges, 1997), da diese einfach berechenbar sind. Louviere et al. (2000) konnten zudem nachweisen, dass mit einer linearen Funktion gute Voraussagen im mittleren Bereich der Nutzenfunktion erzielt werden. Bei weit entfernten Werten ist jedoch damit zu rechnen, dass die Modellvoraussagen ungenauer werden. Eine lineare Nutzenfunktion kann beispielsweise folgende Form aufweisen:

$$w(p_m, z_m, s_i) = \beta_0 + \sum_{k=1}^K \beta_k z_{mk} + \beta_p p_m + \beta_s s_i.$$

Die Parameter β werden ökonometrisch geschätzt und bezeichnen den Grenznutzen der einzelnen Produktattribute z_k beziehungsweise des Preises p . Bei einer Linearitätsannahme der Nutzenfunktion ist der Grenznutzen der Attribute konstant, das heisst unabhängig von der Höhe der Ausprägung eines Attributs gilt der gleiche marginale Nutzen. Diese Annahme ist in vielen Fällen zu restriktiv, so dass auch quadratische Nutzenfunktionen oder funktionale Formen mit Interaktionstermen zwischen Produktattributen (vgl. z. B. Becker, 2006) eingesetzt werden. In diesem Fall entspricht der Grenznutzen allgemein der marginalen Ableitung der Nutzenfunktion nach dem entsprechenden Attribut.

$$\partial v(p_m, z_m, s_i, \varepsilon_{im}) / \partial z_k$$

Integriert man den Grenznutzen bis zu einer spezifischen Attributsausprägung, resultiert der Teilnutzenwert für diese Ausprägung. Wiederholt man das für alle Attribute und aggregiert die Teilnutzenwerte gemäss der gewählten Nutzenfunktion auf, lässt sich der Gesamtnutzen für ein Produkt mit beliebig spezifizierten Attributsausprägungen ermitteln.

Aus den getroffenen Entscheidungen lassen sich ebenfalls die marginalen Grenzzraten der Substitution (GRS) zwischen den Produkteigenschaften berechnen. Dies entspricht dem Austauschverhältnis, in welchem die Personen bereit sind, zwei Attribute gegeneinander auszutauschen. Die Berechnung der GRS zwischen einer Produkteigenschaft z_k und einem Attribut z_l lässt sich als das Verhältnis zwischen den partiellen Ableitungen der Nutzenfunktionen nach z_k und z_l ausdrücken (vgl. z. B. Mas-Colell et al., 1995, Kap. 3):

$$GRS = - \frac{\partial v(p_m, z_m, s_i, \varepsilon_{im}) / \partial z_k}{\partial v(p_m, z_m, s_i, \varepsilon_{im}) / \partial z_l}.$$

Handelt es sich bei der zweiten Eigenschaft um den Preis des Produktes (p), kann der Nenner obiger Gleichung als Grenznutzen des Einkommens interpretiert werden. Die GRS gibt in diesem Fall an, um wie viel Einheiten Einkommen man zu verzichten bereit ist, um eine Einheit mehr des Attributs z_k zu erhalten. Das ist aber nichts anderes als die marginale Zahlungsbereitschaft für die Produkteigenschaft z_k .

Durchführung eines Discrete-Choice-Experiments

Zur Durchführung eines DCE sind verschiedene Schritte erforderlich, die im Folgenden allgemein skizziert werden:¹⁷

- Definition der Attribute und Ausprägungen für das zu untersuchende Produkt
- Optimierung des Befragungsdesigns
- Auswahl der Stichprobe und der Befragungsform
- Statistische Auswertung

Definition von Attributen und Ausprägungen

Als erstes müssen für das zu untersuchende Produkt oder die Leistung die wesentlichen, nutzenstiftenden Attribute definiert und die dazugehörigen Ausprägungen festgelegt werden, die im Rahmen des Entscheidungsexperiments variiert werden sollen. Über die Anzahl der Attribute und Ausprägungen herrscht in der wissenschaftlichen Literatur kein Konsens. Je nach Untersuchungsgegenstand kann ein Produkt durch mehr oder weniger Attribute beschrieben werden. Wichtig ist es, das Entscheidungsexperiment nicht zu komplex auszugestalten und möglichst realistische Szenarien abzubilden, so dass robuste Wahlentscheidungen getroffen werden können. Bei einem zu hohen Komplexitätsgrad besteht die Gefahr, dass die Befragten heuristisch entscheiden (z. B. nur aufgrund einer einzelnen Eigenschaft) oder schneller ermüden und es zu Fehlern bei den Entscheidungen kommt.

Bezüglich der möglichen Produkteigenschaften gibt es kaum Einschränkungen. Grundsätzlich lassen sich quantitative und qualitative Attribute unterscheiden. Quantitative Attribute bewegen sich auf einer klar definierten, messbaren Skala (z. B. Wartezeit, Netzabdeckung), während qualitative Attribute nicht direkt messbar sind (z. B. Handhabbarkeit oder Farbe eines Produkts). Alle untersuchten Attribute sollten für die Entscheidung relevant, realistisch und plausibel sein, was im Rahmen eines Prätests überprüft werden sollte.

Optimierung des Befragungsdesigns

Ist einmal festgelegt, mit welchen Attributen das zu untersuchende Produkt beschrieben wird und welche Ausprägungen diese Attribute annehmen können, muss in einer zweiten Stufe entschieden werden, welche Alternativen (Kombinationen von Attributsausprägungen) im Experiment verwendet werden sollen. Diese Frage stellt sich, da die Anzahl möglicher Alternativen mit der Zahl der Attribute und deren Ausprägungen exponentiell steigt. Hat man zum Beispiel vier Attribute mit je drei Ausprägungen, ergeben sich bereits 81 mögliche Alternativen ($3^4 = 81$). Um aus der Vielzahl von Kombinationsmöglichkeiten eine Anzahl an Alternativen zu bestimmen, die sich im Rahmen eines Entscheidungsexperiments von den Befragten bewältigen lässt, werden sogenannte Designoptimierungsverfahren eingesetzt.

Die Grundidee dahinter besteht darin, eine Teilmenge von Alternativen zu finden, die das vollständige Design möglichst gut repräsentiert. Da bei einer solchen Reduzierung zwangsläufig statistische Informationen verloren gehen, stellen die Designoptimierungsverfahren sicher, dass alle interessierenden Effekte auch mit dem reduzierten Design ökonometrisch geschätzt werden können.

¹⁷ Die folgenden Ausführungen basieren hauptsächlich auf Telser (2002) und Becker (2006).

Im Rahmen der Befragung müssen sich die Befragten dann mehrmals zwischen verschiedenen hypothetischen Gütern oder Leistungen entscheiden. Bei Vorlage einer Karte mit typischerweise zwei Alternativen muss sich die befragte Person entscheiden, welche Alternative sie bevorzugt. Um sicherzustellen, dass alle Befragungsteilnehmer das gleiche unter den Attributen verstehen, wird eine kurze Beschreibung jedes Attributs zur Verfügung gestellt. Für das Entscheidungsexperiment selbst werden die Personen dann gebeten, alle anderen, nicht im Experiment abgebildeten Eigenschaften als konstant anzusehen.

Auswahl der Stichprobe und der Befragungsform

Die für die Befragung zu ziehende Stichprobe orientiert sich an der Fragestellung der Untersuchung. Je nachdem, wessen Präferenzen ermittelt werden sollen, gibt es andere Anforderungen an die Stichprobe. Prinzipiell kann zwischen einfachen Zufallsstichproben und mittels Quotenverfahren exogen stratifizierten Zufallsstichproben unterschieden werden. Im ersten Fall hat jedes Individuum der Grundgesamtheit die gleiche Wahrscheinlichkeit, für die Stichprobe ausgewählt zu werden. Beim Quotenverfahren soll hingegen durch die Verwendung von A-priori-Informationen eine die Grundgesamtheit repräsentierende Teilmenge gewonnen werden. Um repräsentative Resultate für die Gesamtbevölkerung zu erhalten, wird beispielsweise häufig nach Alter und Geschlecht quotiert.

Will man separate Auswertungen für Teilgruppen der Bevölkerung vornehmen, muss neben der richtigen Quotierung der Stichprobe zusätzlich sichergestellt werden, dass für all diese Teilgruppen eine genügend grosse Unterstichprobe vorhanden ist, damit überhaupt statistisch gesicherte Aussagen für diese Gruppen gemacht werden können.

Neben der Stichprobe ist zu entscheiden, welche Form der Befragung für das DCE eingesetzt werden soll. Grundsätzlich sind sowohl persönliche Interviews, computergestützte Telefoninterviews, schriftliche Befragungen oder auch Online-Befragungen über das Internet möglich. Jedes der Verfahren hat Vor- und Nachteile, so sind zum Beispiel schriftliche Befragungen relativ billig durchzuführen, haben aber den Nachteil, dass der Rücklauf oft ungenügend ist und die erhaltenen Daten häufig Selektionseffekte beinhalten. Es gilt, die Vor- und Nachteile gegeneinander abzuwägen und sich basierend darauf für eine Befragungsmethode zu entscheiden.

Statistische Auswertung

Die statistische Auswertung der Daten erfolgt mittels ökonometrischer Methoden, die auf dem ökonomischen Entscheidungsmodell beruhen, welches oben skizziert wurde. Da es sich bei den getroffenen Entscheidungen um binäre Variablen (Ja-Nein-Antworten) handelt, kommen für die Auswertung verschiedene Logit- und Probit-Modelle in Frage. Logit-Modelle haben gegenüber Probit-Modellen den grossen Vorteil, dass sie eine analytische Lösung besitzen, was die ökonometrische Schätzung deutlich erleichtert. Mit diesen Modellen lässt sich aus den Entscheiden der Befragten bestimmen, welche impliziten Abwägungen diese zwischen den Produkteigenschaften gemacht haben, welche Grenznutzen bestehen, welche Teilnutzen und Gesamtnutzen für spezifische Ausprägungen der Attribute bestehen und was für eine Zahlungsbereitschaft sie für eine Verbesserung der Attribute haben. Je nach Ausgestaltung des Modells lassen sich durchschnittliche oder individuelle Zahlungsbereitschaften ermitteln, welche zwischen verschiedenen Gruppen verglichen werden können.

Fragebogen des durchgeführten DCE

Testinterview - at Sprachwahl ⚙

Bitte wählen Sie Ihre Sprache.

Choisissez votre langue s.v.p.

Deutsch

Français

<<

>>

Testinterview - at einl ⚙

Was ist der Nutzen neuer Behandlungsmöglichkeiten?

Wir möchten den Nutzen von Krebstherapien messen, die nicht auf die Heilung vom Krebs, sondern auf die Stabilisierung oder eine Verbesserung des Krankheitszustands ausgelegt sind. Um diese Nutzenmessung für die Schweizer Bevölkerung richtig einzustellen, brauchen wir Ihre Mithilfe.

Mit Ihrer Teilnahme unterstützen Sie ein Forschungsprojekt über die Meinungen der Bevölkerung zur Gesundheitsversorgung. Die Ergebnisse des Forschungsprojektes werden eingesetzt, um Empfehlungen für die Gesundheitspolitik in der Schweiz zu erstellen.

Die Angaben zu Ihrer Person behandeln wir streng vertraulich: Die Ergebnisse der Studie werden keine Rückschlüsse auf Ihre Person zulassen. Die Beantwortung der Fragen wird ca. 15 Minuten in Anspruch nehmen.

Wir danken für Ihre wertvolle Mitarbeit!

<<

>>

Testinterview - at F1 ⚙

Starten wir mit ein paar Fragen zu Ihrer Person.

Wie alt sind Sie?

Alter: Jahre

<<

>>

Testinterview - at F2 ⚙

Geschlecht:

weiblich

männlich

divers

<<

>>

Testinterview - at F3

Welche Postleitzahl hat Ihr Wohnort?

☐ keine Antwort

Testinterview - at dispBew

Im Folgenden werden Sie 8 Entscheidungssituationen rund um eine Krebstherapie sehen. Die Therapie hat nicht die Heilung zum Ziel, sondern die Stabilisierung oder Verbesserung des Krankheitszustandes. Sie finden bei diesen Entscheidungssituationen immer zwei Behandlungsoptionen.

Die jeweiligen Behandlungsoptionen unterscheiden sich in acht Eigenschaften, die sowohl den **Nutzen für die behandelte Person** als auch den **gesellschaftlichen Nutzen** der Behandlung beschreiben. Wir stellen Ihnen diese Eigenschaften auf den nächsten Seiten genau vor.

Weiteres Vorgehen

- Vorstellung der Eigenschaften: ca. 5 min
- Acht Entscheidungssituationen: ca. 6 min
- Allgemeine Fragen zu Ihrer Person: ca. 2 min

Testinterview - at dispGes6

Befassen wir uns zuerst mit den Folgen für die ganze Gesellschaft.

Krankheitslast in der Gesellschaft

Die Krankheitslast misst die Belastung der Gesellschaft durch die Krankheit, die behandelt wird.

Sie entspricht der «Anzahl verlorener Lebensjahre». Vorzeitige Todesfälle sind dabei nur ein Teil der Krankheitslast. Sie berücksichtigt nämlich auch, wieviele Lebensjahre aufgrund reduzierter Lebensqualität «verloren» gehen. Die Treiber einer hohen Krankheitslast sind somit:

- Hohe Sterblichkeit aufgrund der Krankheit
- Niedrige Lebensqualität mit der Krankheit
- Hohe Verbreitung der Krankheit

Krebsbedingte Krankheitslast

Gemessen an der gesamten Krankheitslast für die Schweiz entfallen rund 18% auf Krebs.



Drei Ausprägungen der Krankheitslast

geringe Krankheitslast	Die betrachtete Krankheit verursacht 1% der krebbsbedingten Krankheitslast in der Schweiz. 
mittlere Krankheitslast	Die betrachtete Krankheit verursacht 10% der krebbsbedingten Krankheitslast in der Schweiz. 
hohe Krankheitslast	Die betrachtete Krankheit verursacht 20% der krebbsbedingten Krankheitslast in der Schweiz. 



Testinterview - at dispGes7 ⚙

Verlorene Arbeitszeit je Behandlung

Durch die Erkrankung und die Behandlung geht der Wirtschaft produktive Arbeitszeit verloren. Dies bei der behandelten Person selbst, wenn diese bspw. nicht oder nur eingeschränkt arbeiten kann. Aber auch im Umfeld der behandelten Person kann Arbeitszeit verloren gehen, etwa durch Pflege von Angehörigen.

Zwei Ausprägungen der verlorenen Arbeitszeit je Behandlung

1 verlorener Arbeitsmonat	Jede behandelte Person und das pflegende Umfeld verlieren durch die Behandlung produktive Zeit in der Höhe von durchschnittlich 1 Arbeitsmonat.
6 verlorene Arbeitsmonate	Jede behandelte Person und das pflegende Umfeld verlieren durch die Behandlung produktive Zeit in der Höhe von durchschnittlich 6 Arbeitsmonaten.

<<

>>

Testinterview - at dispGes8 ⚙

Kosten je Behandlung

Für unser Experiment gehen wir davon aus, dass die Arzneimittelkosten selbst immer gleich hoch sind. Neben diesen Arzneikosten entstehen aber je nach Behandlung noch weitere Kosten, z. B. aufgrund der ärztlichen Betreuung oder der Behandlung von Nebenwirkungen. Diese Zusatzkosten werden grundsätzlich von der obligatorischen Krankenversicherung bezahlt.

Zwei Ausprägungen der weiteren Kosten je Behandlung

10'000 CHF	Neben den Arzneimittelkosten entstehen je Behandlung weitere Kosten in Höhe von 10'000 CHF.
80'000 CHF	Neben den Arzneimittelkosten entstehen je Behandlung weitere Kosten in Höhe von 80'000 CHF.

<<

>>

Testinterview - at dispPat1 ⚙

Befassen wir uns nun mit den Folgen für die einzelne behandelte Person.

Überlebenszeit nach Therapiebeginn

Darunter verstehen wir die Überlebenszeit **deit dem Beginn** der Behandlung. Die Hälfte der behandelten Personen überlebt nach dem Therapiebeginn länger, die andere Hälfte weniger lang als die genannte Zeit.

Drei Ausprägungen der Überlebenszeit nach Therapiebeginn

3 Monate	Die Hälfte der behandelten Personen lebt nach Therapiebeginn noch mehr als 3 Monate.
2 Jahre	Die Hälfte der behandelten Personen lebt nach Therapiebeginn noch mehr als 2 Jahre (24 Monate).
5 Jahre	Die Hälfte der behandelten Personen lebt nach Therapiebeginn noch mehr als 5 Jahre (60 Monate).

<<

>>

Langfristige Lebensqualität

Die langfristige Lebensqualität beschreibt den Gesundheitszustand nachdem die intensive Behandlung abgeschlossen ist.

Fünf Lebensbereiche fließen in die Beurteilung der gesundheitsbezogenen Lebensqualität ein:

- Beweglichkeit, Mobilität
- Die Fähigkeit, für sich selbst zu sorgen
- Alltägliche Tätigkeiten (z. B. Arbeit, Familie, Freizeit)
- Schmerzen, körperliche Beschwerden
- Angst, Niedergeschlagenheit

Drei Ausprägungen der langfristigen Lebensqualität

leichte Einschränkungen	Gesundheitszustand mit Problemen in einem der fünf genannten Lebensbereiche und keinen Problemen in allen anderen.
mittlere Einschränkungen	Gesundheitszustand mit Problemen in mehreren der fünf genannten Lebensbereichen oder extremen Problemen in einem der fünf Bereiche.
starke Einschränkungen	Gesundheitszustand mit extremen Problemen in den meisten der fünf genannten Lebensbereichen.

**Schwere Nebenwirkungen**

Darunter verstehen wir den Anteil behandelter Personen mit mindestens einer schweren Nebenwirkung **während der Behandlung**. Durch die schwere Nebenwirkung braucht es (unter Umständen: lebensrettende) Massnahmen im Spital. Selten stirbt man auch an der schweren Nebenwirkung. Normalerweise haben die schweren Nebenwirkungen aber keine langfristigen gesundheitlichen Schäden zur Folge.

Zwei Ausprägungen schwerer Nebenwirkungen

5% der behandelten Personen	5 von 100 behandelten Personen haben eine schwere Nebenwirkung.
80% der behandelten Personen	80 von 100 behandelten Personen haben eine schwere Nebenwirkung.

Vor dem Beginn einer Behandlung muss jede behandelte Personen entscheiden, welches Risiko in Bezug auf schwere Nebenwirkung eingegangen werden soll. Dafür kann die Behandlung bei Erfolg die langfristige Lebensqualität erhöhen. Wie ist das bei Ihnen, welche der folgende drei Aussagen trifft für Sie am meisten zu?

Selten auftretende schwere Nebenwirkungen sind bei meiner Entscheidung für oder gegen eine Behandlung wichtiger als eine hohe langfristige Lebensqualität.

Eine hohe langfristige Lebensqualität ist bei meiner Entscheidung für oder gegen eine Behandlung wichtiger als selten auftretende schwere Nebenwirkungen.

Ich kann mich nicht zwischen langfristiger Lebensqualität und seltenen schweren Nebenwirkungen entscheiden.



Testinterview - at dispPat4 **Behandlungshäufigkeit**

Darunter verstehen wir die Häufigkeit, in der einzelne Behandlungsschritte (Sitzungen) während der Hauptbehandlung durchgeführt werden. Die Länge der Sitzungen und die Dauer der Behandlung ist davon unabhängig.

Zwei Ausprägungen der Behandlungshäufigkeit

Täglich	Die Behandlung erfolgt täglich.
Monatlich	Die Behandlung erfolgt einmal pro Monat.

Testinterview - at dispPat5 **Behandlungsort**

Darunter verstehen wir den Behandlungsort für die Durchführung der Behandlung.

Zwei Ausprägungen des Behandlungsorts

Arztpraxis/Klinik	Für die Behandlung muss die behandelte Person in eine Arztpraxis oder eine Klinik gehen. Es ist kein stationärer Aufenthalt (über Nacht) notwendig.
zu Hause	Die behandelte Person kann die Behandlung zu Hause durchführen.

Testinterview - at dispCJ 

Nach diesem allgemeinen Teil bitten wir Sie im Folgenden acht Entscheidungen zu treffen. Wir zeigen Ihnen gleichzeitig immer zwei Optionen für die Behandlung von Krebserkrankungen. Jede Option ist durch die soeben vorgestellten acht Eigenschaften charakterisiert.

Gehen Sie davon aus, dass alle übrigen hier nicht genannten Aspekte einer Krankheit bzw. ihrer Behandlung bei allen Entscheidungen genau gleich sind und deshalb bei Ihrer Entscheidung nicht berücksichtigt werden müssen.

Wählen Sie jeweils die **Option, die Sie persönlich besser finden**. Manchmal kann Ihnen die Entscheidung leichtfallen. Es wird aber auch vorkommen, dass Sie sich schwer entscheiden können. Wichtig ist, dass Sie sich auch in diesen Situationen zu einer Entscheidung durchringen. Eine Option «Unentschieden» gibt es nicht.

Sie können sich bei jeder Entscheidung genauere Informationen zu den Eigenschaften mit Hilfe des Info-Buttons  anzeigen lassen.



Testinterview - at cj101r 

Wählen Sie die Option, die Sie persönlich besser finden.

Eigenschaft	Option 1	Option 2
Krankheitslast in der Gesellschaft [ⓘ]	geringe Krankheitslast [ⓘ]	hohe Krankheitslast [ⓘ]
Verlorene Arbeitszeit je Behandlung [ⓘ]	6 verlorene Arbeitsmonate [ⓘ]	1 verlorener Arbeitsmonat [ⓘ]
Weitere Kosten je Behandlung [ⓘ]	10'000 CHF [ⓘ]	80'000 CHF [ⓘ]
Überlebenszeit [ⓘ]	2 Jahre [ⓘ]	3 Monate [ⓘ]
Langfristige Lebensqualität [ⓘ]	mittlere Einschränkungen [ⓘ]	leichte Einschränkungen [ⓘ]
Schwere Nebenwirkungen [ⓘ]	80% der Patienten [ⓘ]	5% der Patienten [ⓘ]
Behandlungshäufigkeit [ⓘ]	täglich [ⓘ]	täglich [ⓘ]
Behandlungsort [ⓘ]	Arztpraxis/Klinik [ⓘ]	zu Hause [ⓘ]

Option 1

Option 2

<<

>>

Testinterview - at cj103r 

Wählen Sie die Option, die Sie persönlich besser finden.

Eigenschaft	Option 1	Option 2
Krankheitslast in der Gesellschaft [ⓘ]	geringe Krankheitslast [ⓘ]	mittlere Krankheitslast [ⓘ]
Verlorene Arbeitszeit je Behandlung [ⓘ]	6 verlorene Arbeitsmonate [ⓘ]	1 verlorener Arbeitsmonat [ⓘ]
Weitere Kosten je Behandlung [ⓘ]	80'000 CHF [ⓘ]	10'000 CHF [ⓘ]
Überlebenszeit [ⓘ]	3 Monate [ⓘ]	2 Jahre [ⓘ]
Langfristige Lebensqualität [ⓘ]	mittlere Einschränkungen [ⓘ]	starke Einschränkungen [ⓘ]
Schwere Nebenwirkungen [ⓘ]	5% der Patienten [ⓘ]	80% der Patienten [ⓘ]
Behandlungshäufigkeit [ⓘ]	monatlich [ⓘ]	monatlich [ⓘ]
Behandlungsort [ⓘ]	zu Hause [ⓘ]	Arztpraxis/Klinik [ⓘ]

Option 1

Option 2

<<

>>

Testinterview - at cj106r 

Wählen Sie die Option, die Sie persönlich besser finden.

Eigenschaft	Option 1	Option 2
Krankheitslast in der Gesellschaft [ⓘ]	hohe Krankheitslast [ⓘ]	geringe Krankheitslast [ⓘ]
Verlorene Arbeitszeit je Behandlung [ⓘ]	1 verlorener Arbeitsmonat [ⓘ]	6 verlorene Arbeitsmonate [ⓘ]
Weitere Kosten je Behandlung [ⓘ]	80'000 CHF [ⓘ]	10'000 CHF [ⓘ]
Überlebenszeit [ⓘ]	5 Jahre [ⓘ]	2 Jahre [ⓘ]
Langfristige Lebensqualität [ⓘ]	starke Einschränkungen [ⓘ]	leichte Einschränkungen [ⓘ]
Schwere Nebenwirkungen [ⓘ]	5% der Patienten [ⓘ]	80% der Patienten [ⓘ]
Behandlungshäufigkeit [ⓘ]	täglich [ⓘ]	monatlich [ⓘ]
Behandlungsort [ⓘ]	Arztpraxis/Klinik [ⓘ]	Arztpraxis/Klinik [ⓘ]

Option 1

Option 2

<<

>>

Testinterview - at cj105r 

Wählen Sie die Option, die Sie persönlich besser finden.

Eigenschaft	Option 1	Option 2
Krankheitslast in der Gesellschaft [ⓘ]	mittlere Krankheitslast [ⓘ]	geringe Krankheitslast [ⓘ]
Verlorene Arbeitszeit je Behandlung [ⓘ]	6 verlorene Arbeitsmonate [ⓘ]	1 verlorener Arbeitsmonat [ⓘ]
Weitere Kosten je Behandlung [ⓘ]	10'000 CHF [ⓘ]	80'000 CHF [ⓘ]
Überlebenszeit [ⓘ]	5 Jahre [ⓘ]	2 Jahre [ⓘ]
Langfristige Lebensqualität [ⓘ]	mittlere Einschränkungen [ⓘ]	leichte Einschränkungen [ⓘ]
Schwere Nebenwirkungen [ⓘ]	80% der Patienten [ⓘ]	5% der Patienten [ⓘ]
Behandlungshäufigkeit [ⓘ]	monatlich [ⓘ]	täglich [ⓘ]
Behandlungsort [ⓘ]	zu Hause [ⓘ]	Arztpraxis/Klinik [ⓘ]

Option 1

Option 2

<<

>>

Testinterview - at cj102r 

Wählen Sie die Option, die Sie persönlich besser finden.

Eigenschaft	Option 1	Option 2
Krankheitslast in der Gesellschaft [ⓘ]	mittlere Krankheitslast [ⓘ]	geringe Krankheitslast [ⓘ]
Verlorene Arbeitszeit je Behandlung [ⓘ]	1 verlorener Arbeitsmonat [ⓘ]	6 verlorene Arbeitsmonate [ⓘ]
Weitere Kosten je Behandlung [ⓘ]	80'000 CHF [ⓘ]	10'000 CHF [ⓘ]
Überlebenszeit [ⓘ]	2 Jahre [ⓘ]	5 Jahre [ⓘ]
Langfristige Lebensqualität [ⓘ]	leichte Einschränkungen [ⓘ]	starke Einschränkungen [ⓘ]
Schwere Nebenwirkungen [ⓘ]	80% der Patienten [ⓘ]	5% der Patienten [ⓘ]
Behandlungshäufigkeit [ⓘ]	monatlich [ⓘ]	täglich [ⓘ]
Behandlungsort [ⓘ]	zu Hause [ⓘ]	zu Hause [ⓘ]

Option 1

Option 2

<<

>>

Testinterview - at cj104r 

Wählen Sie die Option, die Sie persönlich besser finden.

Eigenschaft	Option 1	Option 2
Krankheitslast in der Gesellschaft [ⓘ]	hohe Krankheitslast [ⓘ]	mittlere Krankheitslast [ⓘ]
Verlorene Arbeitszeit je Behandlung [ⓘ]	1 verlorener Arbeitsmonat [ⓘ]	6 verlorene Arbeitsmonate [ⓘ]
Weitere Kosten je Behandlung [ⓘ]	10'000 CHF [ⓘ]	80'000 CHF [ⓘ]
Überlebenszeit [ⓘ]	3 Monate [ⓘ]	2 Jahre [ⓘ]
Langfristige Lebensqualität [ⓘ]	leichte Einschränkungen [ⓘ]	starke Einschränkungen [ⓘ]
Schwere Nebenwirkungen [ⓘ]	5% der Patienten [ⓘ]	80% der Patienten [ⓘ]
Behandlungshäufigkeit [ⓘ]	monatlich [ⓘ]	täglich [ⓘ]
Behandlungsort [ⓘ]	Arztpraxis/Klinik [ⓘ]	zu Hause [ⓘ]

Option 1

Option 2

<<

>>

Testinterview - at cj107r 

Wählen Sie die Option, die Sie persönlich besser finden.

Eigenschaft	Option 1	Option 2
Krankheitslast in der Gesellschaft [ⓘ]	hohe Krankheitslast [ⓘ]	mittlere Krankheitslast [ⓘ]
Verlorene Arbeitszeit je Behandlung [ⓘ]	1 verlorener Arbeitsmonat [ⓘ]	6 verlorene Arbeitsmonate [ⓘ]
Weitere Kosten je Behandlung [ⓘ]	10'000 CHF [ⓘ]	80'000 CHF [ⓘ]
Überlebenszeit [ⓘ]	2 Jahre [ⓘ]	5 Jahre [ⓘ]
Langfristige Lebensqualität [ⓘ]	mittlere Einschränkungen [ⓘ]	leichte Einschränkungen [ⓘ]
Schwere Nebenwirkungen [ⓘ]	5% der Patienten [ⓘ]	80% der Patienten [ⓘ]
Behandlungshäufigkeit [ⓘ]	monatlich [ⓘ]	monatlich [ⓘ]
Behandlungsort [ⓘ]	Arztpraxis/Klinik [ⓘ]	zu Hause [ⓘ]

Option 1

Option 2

<<

>>

Testinterview - at cj108r 

Wählen Sie die Option, die Sie persönlich besser finden.

Eigenschaft	Option 1	Option 2
Krankheitslast in der Gesellschaft [ⓘ]	mittlere Krankheitslast [ⓘ]	hohe Krankheitslast [ⓘ]
Verlorene Arbeitszeit je Behandlung [ⓘ]	1 verlorener Arbeitsmonat [ⓘ]	6 verlorene Arbeitsmonate [ⓘ]
Weitere Kosten je Behandlung [ⓘ]	10'000 CHF [ⓘ]	80'000 CHF [ⓘ]
Überlebenszeit [ⓘ]	3 Monate [ⓘ]	2 Jahre [ⓘ]
Langfristige Lebensqualität [ⓘ]	mittlere Einschränkungen [ⓘ]	starke Einschränkungen [ⓘ]
Schwere Nebenwirkungen [ⓘ]	5% der Patienten [ⓘ]	80% der Patienten [ⓘ]
Behandlungshäufigkeit [ⓘ]	täglich [ⓘ]	täglich [ⓘ]
Behandlungsort [ⓘ]	Arztpraxis/Klinik [ⓘ]	zu Hause [ⓘ]

Option 1

Option 2

<<

>>

Testinterview - at gF14

Geschafft!

Das waren alle Entscheidungssituationen.

Wie leicht oder wie schwer fiel es Ihnen, jeweils eine Entscheidung zu treffen?

☐	sehr leicht
☐	eher leicht
☐	eher schwer
☐	sehr schwer
☐	keine Antwort

Stellen Sie sich vor, Sie würden nächste Woche denselben Fragebogen noch einmal ausfüllen. Wie sicher sind Sie, dass Sie dieselben Entscheidungen treffen würden wie heute?

☐	Sehr sicher. Ich würde in den allermeisten Fällen gleich entscheiden.
☐	Eher sicher. Ich würde in den meisten Fällen gleich entscheiden.
☐	Eher unsicher. Ich würde aber wohl immer noch in mehr als der Hälfte der Fälle gleich entscheiden.
☐	Unsicher.
☐	keine Antwort

>>

Testinterview - at gF15

Bei den folgenden Fragen geht es um die Gesundheit von Ihnen und Ihnen nahestehenden Personen. Ausserdem sind wir daran interessiert, wie Sie bestimmte gesundheitsbezogene Entscheidungen getroffen haben.

Wie ist Ihr Gesundheitszustand im Allgemeinen?

☐	sehr gut
☐	gut
☐	mittelmässig
☐	schlecht
☐	sehr schlecht
☐	keine Antwort

<<

>>

Testinterview - at F16 ⚙

Kennen Sie eine Person in Ihrem näheren Umfeld, die an Krebs erkrankt ist?

Ja, in den letzten fünf Jahren
Ja, vor mehr als fünf Jahren
Nein
keine Antwort

<<

>>

Testinterview - at F17 ⚙

Haben Sie sich schon einmal Gedanken darüber gemacht, Organspender:in zu werden?

Ja, ich bin Organspender:in oder möchte eine:r werden
Ja, aber ich habe mich noch nicht entschieden
Ja, ich habe mich gegen die Organspende entschieden
Nein, ich habe mir dazu noch keine Gedanken gemacht
keine Antwort

<<

>>

Testinterview - at F18 ⚙

Haben Sie sich schon einmal Gedanken über eine Patientenverfügung gemacht, in der geregelt ist, welche Massnahmen ergriffen werden sollen, wenn Sie in eine persönliche medizinische Notsituation geraten würden?

Ja, ich habe eine Patientenverfügung
Ja, ich will eine Patientenverfügung, habe aber noch keine
Ja, ich habe mir Gedanken gemacht und möchte keine Patientenverfügung
Nein, ich habe mir noch keine Gedanken gemacht
keine Antwort

<<

>>

Testinterview - at F19

Wie hoch ist Ihre Franchise für die obligatorische Grundversicherung, also der jährliche Beitrag, den Sie selbst bezahlen müssen, bevor die Krankenkasse Leistungen vergütet?

Die Mindestfranchise für Erwachsene liegt bei 300 Franken.

Hinweis: Sie finden diese Angaben auf Ihrer Versicherungspolice.

☐	300 Franken
☐	500 Franken
☐	1'000 Franken
☐	1'500 Franken
☐	2'000 Franken
☐	2'500 Franken
☐	keine Antwort



Testinterview - at gF20

Ganz zum Schluss haben wir nochmal ein paar Fragen zu Ihrer Person und Ihrem Haushalt.

Was ist Ihre höchste abgeschlossene Ausbildung?

Primar-, Real-, Sekundarschule
Berufslehre, Berufsschule, KV, Gewerbeschule
Maturitätsschule, Gymnasium, LehrerInnenseminar, Berufsmaturität
Höhere Fach- oder Berufsausbildung, Kunstgewerbeschule
Höhere Fachschule z.B. HTL, HWV
Fachhochschule (FHS) und Pädagogische Hochschule
Technische Hochschulen (ETH), Universität
Doktorat, Habilitation
Andere, und zwar:
keine Antwort

<<

>>

Testinterview - at F21

Befinden Sie sich ...

in bezahlter Erwerbstätigkeit, Vollzeit (30 Std. und mehr pro Woche)
in bezahlter Erwerbstätigkeit, Teilzeit (1-29 Std. pro Woche)
in der Ausbildung
in einer Tätigkeit im eigenen Haushalt (Hausfrau/-mann)
oder sind Sie ...
pensioniert, RentnerIn
arbeitslos, auf Stellensuche
nicht erwerbstätig
keine Antwort

<<

>>

Testinterview - at F22

Wie viele Personen (Sie eingeschlossen) leben in Ihrem Haushalt?

Anzahl Erwachsene:

Anzahl Kinder unter 18 Jahren:

☐ keine Antwort

Testinterview - at F23

Bitte geben Sie das **Netto**-Monatseinkommen Ihres Haushalts an.

Denken Sie dabei an Ihr eigenes Einkommen oder Ihre Rente, aber auch an das Einkommen oder die Rente von allfällig weiteren Personen in ihrem Haushalt.

Sie können sicher sein, dass die Angabe streng vertraulich und anonym behandelt wird.

unter Fr. 3'000

zwischen Fr. 3'000 und 5'000

zwischen Fr. 5'000 und 7'000

zwischen Fr. 7'000 und 9'000

zwischen Fr. 9'000 und 11'000

zwischen Fr. 11'000 und 13'000

zwischen Fr. 13'000 und 15'000

mehr als Fr. 15'000

keine Antwort

Testinterview - at g24

Die Befragung ist nun abgeschlossen. Möchten Sie uns noch etwas mitteilen?

☐ keine Antwort

Mit einem Klick auf den Pfeil schliessen Sie die Umfrage ab.



Testinterview - at ende ⚙

Sie sind nun fertig, Ihre Antworten wurden übermittelt.

Vielen Dank für Ihre Teilnahme!



Anhang D: Resultat der Expertenbefragung

Die Expertenbefragung basiert auf 18 Personen mit jeweils zwölf Entscheidungssituationen, in welchen aus zwei möglichen Behandlungsoptionen die jeweils präferierte zu wählen war. Das DCE wurde als Onlineumfrage im Mai und Juni 2021 durchgeführt. Die in Tabelle 20 enthaltenen Koeffizienten dienten als Ausgangswerte für das effiziente Design der im Haupttext beschriebenen Bevölkerungsbefragung. Die Gewichte, die (statistischer Signifikanz unbeachtet) daraus resultieren sind in Tabelle 21 enthalten.

Tabelle 20 Koeffizienten aus dem DCE: Expertenbefragung

Attribut	Ausprägung	Koeffizient	Standardfehler	p-Wert
Überlebenszeit nach Therapiebeginn (Median)	3 Monate	0	.	.
	2 Jahre	2.490 ***	0.619	0.000
	5 Jahre	3.957 ***	1.121	0.000
Langfristige Lebensqualität	gravierende Einschränkungen	0	.	.
	mittlere Einschränkungen	0.990	0.680	0.145
	leichte Einschränkungen	1.654	1.179	0.161
Schwere unerwünschte Ereignisse	5%	0	.	.
	80%	-0.839 *	0.358	0.019
Behandlungshäufigkeit	täglich	0	.	.
	monatlich	0.230	0.397	0.563
Behandlungsort	ambulant	0	.	.
	zu Hause	0.454	0.377	0.230
Krankheitslast in der Gesellschaft	geringe Krankheitslast	0	.	.
	mittlere Krankheitslast	-0.292	0.267	0.274
	hohe Krankheitslast	-0.159	0.227	0.483
Indirekte Kosten für die Gesellschaft	1 verlorener Arbeitsmonat	0	.	.
	6 verlorene Arbeitsmonate	-0.188	0.311	0.546
Weitere Behandlungskosten	10'000 CHF	0	.	.
	80'000 CHF	-0.151	0.330	0.646

***: signifikant auf dem 0.1-Prozent-Niveau, **: 1-Prozent-Niveau, *: 5-Prozent-Niveau.

Quelle: Eigene Berechnungen aus DCE-Experiment mit Experten, Darstellung Polynomics.

Tabelle 21 Gewichtung der Attribute aus dem DCE: Expertenbefragung

Attribut	Distanz der Koeffizienten	Gewicht aus Expertenbefragung
Überlebenszeit nach Therapiebeginn (Median)	3.970	51.0%
Langfristige Lebensqualität	1.654	21.2%
Schwere unerwünschte Ereignisse	0.839	10.8%
Behandlungshäufigkeit	0.230	2.9%
Behandlungsort	0.454	5.8%
Krankheitslast in der Gesellschaft	0.304	3.9%
Indirekte Kosten für die Gesellschaft	0.188	2.4%
Weitere Behandlungskosten	0.151	1.9%
Summe	7.789	100.0%

Quelle: Eigene Berechnungen aus DCE-Experiment der Experten, Darstellung Polynomics.

Polynomics AG
Baslerstrasse 44
CH-4600 Olten

www.polynomics.ch
polynomics@polynomics.ch

Telefon +41 62 205 15 70
Fax +41 62 205 15 80