

**Bewertung von Lebensqualität im Kontext
gesundheitsökonomischer Evaluationen
– Ermittlung der gesellschaftlichen Präferenzwerte
für die deutsche Version des EQ-5D-5L**

Kristina Ludwig (geb. Klose), BSc, MSc in Public Health

Kumulative Dissertationsschrift zur Erlangung des Grades „Doctor of Public Health“

Universität Bielefeld

Fakultät für Gesundheitswissenschaften

Gutachter

1. Prof. Dr. Wolfgang Greiner
2. Prof. Dr. J.-Matthias Graf von der Schulenburg

Bielefeld, den 13. Juni 2018

Inhaltsverzeichnis

| | Seite |
|--|-------|
| Abkürzungsverzeichnis | III |
| Abbildungs- und Tabellenverzeichnis | IV |
| Zusammenfassung | V |
| | |
| 1. Von einer Morbiditäts- zu einer Lebensqualitätsorientierung im Gesundheitswesen | 1 |
| 1.1 Problemstellung und Public Health-Relevanz | 1 |
| 1.2 Konzept und Fragestellungen der Dissertation | 3 |
| 2. Nutzentheoretische Lebensqualitätsmessung im Rahmen gesundheitsökonomischer Evaluationen | 6 |
| 2.1 Theoretische Fundierung und Grundannahmen ökonomischer Analysen | 6 |
| 2.2 Lebensqualität als Outcomeparameter in gesundheitsökonomischen Evaluationen | 9 |
| 2.3 Vom Fragebogen zum Nutzwert mit dem Methoden-Protokoll der EuroQol-Gruppe | 11 |
| 3. Methodisches Vorgehen der empirischen Studien | 15 |
| 4. Ergebnissynthese: Ermittlung gesellschaftlicher Präferenzwerte für die deutsche Version des EQ-5D-5L | 18 |
| 4.1 Methodische Entscheidungen und Eignung des Methoden-Protokolls für den deutschen Kontext | 18 |
| 4.2 Einfluss der Präferenzmethoden auf die Lebensqualitätsabschläge | 21 |
| 4.3 Value Set für die deutsche Version des EQ-5D-5L | 23 |
| 4.4 Einsatz des Methoden-Protokolls in künftigen Valuierungsstudien..... | 24 |
| 5. Diskussion | 28 |
| 5.1 Ergebnisorientierte Diskussion | 28 |
| 5.2 Methodenorientierte Diskussion | 31 |
| 6. Implikationen für Forschung, Politik und Praxis | 36 |
| | |
| Referenzen | 41 |
| Anhang | 48 |
| Module der kumulativen Dissertation | 52 |

Abkürzungsverzeichnis

| | |
|------------|---|
| CADTH | Canadian Agency for Drugs and Technologies in Health |
| CAPI | Computer Assisted Personal Interviewing |
| cTTO | composite Time-Trade-Off |
| DCE | Discrete Choice Experiment |
| EQ-VT | EuroQol-Valuation Technique |
| EQ-5D-Y | EQ-5D-Youth |
| EQ-5D-Y-5L | fünf-Level-Version des EQ-5D-Youth |
| EQ-5D-3L | drei-Level-Version des EQ-5D |
| EQ-5D-5L | fünf-Level-Version des EQ-5D |
| HUI | Health Utility Index |
| IQWiG | Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen |
| LT-TTO | lead-time Time-Trade-Off |
| MSAC | Medical Services Advisory Committee |
| NICE | National Institute of Health and Care Excellence |
| PRO | patient-reported outcomes |
| QALY | quality adjusted life year |
| SF-6D | Short Form-6D |
| SG | Standard Gamble |
| SGB | Sozialgesetzbuch |
| TTO | Time-Trade-Off |
| VAS | visuelle Analogskala |

Abbildungsverzeichnis

| | Seite |
|---|-------|
| Abbildung 1: Konzept der Dissertation | 4 |

Tabellenverzeichnis

| | Seite |
|--|-------|
| Tabelle 1: Methodik der empirischen Studien der Dissertation | 16 |
| Tabelle 2: Checkliste für die Planung und Durchführung einer Face-to-Face-Studie zur Valuierung eines Lebensqualitätsinstrumentes | 19 |
| Tabelle 3: Lessons learned bei der Anwendung des Methoden-Protokolls der EuroQol-Gruppe | 25 |

Zusammenfassung

Hintergrund: Lebensqualitätseffekte von Gesundheitsleistungen haben zunehmend Eingang in gesundheitsökonomische Evaluationen gefunden, um eine effektive und zugleich effiziente Versorgung für Patienten sicherzustellen. Der Einbezug der Lebensqualität bei Kosten-Nutzwert-Analysen erfordert nutzentheoretische Indexinstrumente. Derzeit besteht jedoch ein Mangel an geeigneten Value Sets für jene Instrumente in Deutschland. Die Zielsetzung der Dissertation liegt darin, ein präferenzbasiertes Value Set für die deutsche fünf-Level-Version des EQ-5D (EQ-5D-5L) unter Anwendung des Methoden-Protokolls der EuroQol-Gruppe zu entwickeln.

Methode: Anhand von theoretischen Vorüberlegungen sowie drei empirischen Studien mit der deutschen Allgemeinbevölkerung, werden methodische und inhaltliche Aspekte in Bezug auf die Bewertung der gesundheitlichen Einschränkungen für den EQ-5D-5L in Deutschland erforscht. Es werden zwei explorative Vorstudien durchgeführt, um die methodische Vorgehensweise zu testen. In der anschließenden *Valuierungsstudie* steht die Ermittlung der gesellschaftlichen Präferenzwerte für den EQ-5D-5L für Deutschland im Fokus. Die Erkenntnisse der Dissertation münden in die Ableitung von Implikationen für Forschung, Politik und Praxis.

Ergebnisse: Die kumulative Dissertation umfasst sechs Publikationen. Die *Module 1 und 2* beinhalten theoretische und methodische Vorüberlegungen für die Valuierung eines Lebensqualitätsinstrumentes. *Modul 3* zeigt, dass ein Value Set ausschließlich für das valuierte Messinstrument und nicht für etwaige Kinderversionen ungeprüft anzuwenden ist. *Modul 4* hat einen Beitrag zu der Weiterentwicklung des internationalen Methoden-Protokolls geliefert und zeigte die Eignung des Protokolls für den deutschen Kontext. In der Valuierungsstudie (*Modul 5*) konnte ein konsistentes Value Set für den EQ-5D-5L auf Basis der gesellschaftlichen Präferenzwerte für Deutschland ermittelt werden. In dem *Modul 6* werden die zentralen Erkenntnisse für den Einsatz des Methoden-Protokolls in künftigen Studien zusammengeführt.

Schlussfolgerungen: Das entwickelte EQ-5D-5L Value Set stellt ein Novum für die deutsche Lebensqualitätsforschung dar und bietet eine aktuelle und robuste Berechnungsgrundlage, um einen präferenzbasierten EQ-5D-5L Index für Deutschland zu aggregieren. Diese gesellschaftlichen Präferenzwerte haben Implikationen für die Bewertung von Therapiemaßnahmen und ermöglichen den Einsatz des EQ-5D-5L in verschiedenen Anwendungsbereichen. So kann der EQ-5D-5L künftig als Operationalisierung des patientenrelevanten Endpunktes Lebensqualität zur Unterstützung von Allokationsentscheidungen im deutschen Gesundheitswesen eingesetzt werden.

1. Von einer Morbiditäts- zu einer Lebensqualitätsorientierung im Gesundheitswesen

1.1 Problemstellung und Public Health-Relevanz

Durch den Wandel des Krankheitspanoramas von akut-somatischen hin zu chronisch-degenerativen Erkrankungen stoßen die traditionellen Endpunkte bzw. Outcomeparameter (wie z. B. Überlebenszeit und Laborwerte) bei der Evaluation von Therapiemaßnahmen an Grenzen. Neue Bewertungsmethoden, wie der Einbezug der Patientensicht¹, gewinnen folglich an Relevanz [1–3]. In der medizinischen Forschung sowie der Versorgungspraxis werden sogenannte patientenberichtete Endpunkte (patient-reported outcomes, PRO) zunehmend berücksichtigt [2]. Die Messung der gesundheitsbezogenen Lebensqualität² (nachfolgend als Lebensqualität bezeichnet) zählt zu den am häufigsten eingesetzten PRO [5, 6] und soll gemäß der deutschen Sozialgesetzgebung neben Morbiditäts- und Mortalitätsparametern zur Bewertung von Arzneimitteln als patientenrelevanter Nutzen berücksichtigt werden (§ 35b Sozialgesetzbuch [SGB] V) [7].

Neben der klinisch-medizinischen Bedeutung besitzt die Lebensqualitätsforschung für den gesundheitspolitischen Bereich eine hohe Public Health-Relevanz. Die Implementierung ökonomischer Prinzipien und Handlungsweisen ist im Gesundheitswesen unumgänglich. Neben der Beurteilung der klinischen Wirksamkeit (Effektivität) von Behandlungsmaßnahmen gilt es auch, deren Wirtschaftlichkeit (Effizienz) zu begutachten, um mit den verfügbaren Ressourcen eine effektive und zugleich effiziente Versorgung für den Patienten sicherzustellen [8, 9]. Es bedarf folglich geeigneter gesundheitsökonomischer Studien, um die Kosten und den Nutzen medizinischer Maßnahmen zu erfassen [9]. Vor allem bei chronischen Krankheiten wirken sich Behandlungsmaßnahmen immer weniger auf die Lebenserwartung, sondern vielmehr auf intangible Parameter, wie die Veränderung der Lebensqualität, aus. Insbesondere für chronisch erkrankte Patienten ist es daher bedeutsam, die Verbesserung der Lebensqualität als Nutzen einer medizinischen Intervention (korrekt) zu erfassen. Neben der individuellen Bedeutung für den Patienten nimmt die Ermittlung der intangiblen Effekte auch für deren Sachverwalter, wie z. B. Krankenkassen, einen hohen Stellenwert ein, um bei begrenzten Ressourcen einen größtmöglichen Nutzen zu realisieren. Vor diesem Hintergrund haben Lebensqualitätseffekte von

¹ An dieser Stelle sei darauf verwiesen, dass im Rahmen dieser Synopse die Verwendung der männlichen Form grundsätzlich die weibliche Form mit einschließt.

² Die gesundheitsbezogene Lebensqualität umfasst verschiedene Dimensionen, die den Gesundheitszustand eines Menschen direkt beeinflussen. Der interessierte Leser sei für umfassende definitorische und konzeptuelle Grundlagen sowie Begriffsabgrenzungen u. a. an Radoschewski [4] verwiesen.

Gesundheitsleistungen zunehmend Eingang in gesundheitsökonomische Evaluationen, u. a. im Rahmen von Kosten-Nutzwert-Analysen, gefunden [9, 10]. Letztere ermöglichen eine systematische Evaluation von medizinischen Interventionen unter Effizienz Gesichtspunkten, einem zentralen gesundheitswissenschaftlichen Erkenntnisinteresse, und können somit eine wichtige Grundlage für Allokationsentscheidungen im Gesundheitswesen darstellen [1, 9].

Der aus dem epidemiologischen Wandel der Morbiditätsstrukturen resultierende Bedarf an adäquaten Messinstrumenten zur Einschätzung des subjektiven Gesundheitszustandes und seiner Beeinflussung durch Therapiemaßnahmen stellt den Ausgangspunkt für die Entwicklung von Lebensqualitätsinstrumenten dar [4]. Die Einbeziehung von Lebensqualität im Rahmen von Kosten-Nutzwert-Analysen erfordert nutzentheoretische (präferenzbasierte) Indexinstrumente, wie z. B. der EuroQol-Fragebogen (EQ-5D). Jene Instrumente ermöglichen die Aggregation der Lebensqualitätswerte (in der Regel) des Patienten auf Grundlage von Präferenzhebungen bzw. den resultierenden Berechnungsvorschriften („Tarif“, „Value Set“) in Nutzwerte [2, 11]. Derzeit besteht jedoch ein Mangel an geeigneten präferenzbasierten Tarifen für Instrumente zur Messung der Lebensqualität in Deutschland [4, 12]. Dies stellt eine Limitation für die Einbeziehung der Lebensqualität in Kosten-Nutzwert-Analysen zur Evaluation von Gesundheitsleistungen im deutschen Kontext dar.

Die drei-Level-Version des EQ-5D (EQ-5D-3L)³ [13] zählt zu den weitverbreitetsten generischen Lebensqualitätsinstrumenten in Deutschland und in anderen Ländern [11, 14]. Dennoch ist der EQ-5D-3L Gegenstand kritischer Diskussionen, da vielfach *Ceiling*-Effekte bemängelt werden, die die diskriminative Fähigkeit des Instrumentes einschränken. Im Jahr 2009 wurde die Anzahl der Antwortkategorien von drei auf fünf erhöht, um die Sensitivität und Diskriminierbarkeit des Fragebogens zu verbessern: die fünf-Level-Version des EQ-5D (EQ-5D-5L)⁴ [15]. Für den EQ-5D-5L ist bisher kein von der EuroQol-Gruppe empfohlenes Set an gesellschaftlichen Präferenzwerten für den deutschen Kontext verfügbar, um die Antworten des deskriptiven Systems in einen singulären Index zu transformieren. Die Entwicklung eines solchen Value Sets ist jedoch essentiell, um das Messinstrument in Kosten-Nutzwert-Analysen und somit zur Unterstützung von Allokationsentscheidungen einsetzen zu können. Für die Ermittlung der nationalen EQ-5D-5L Value Sets hat die EuroQol-Gruppe ein standardisiertes Methoden-Protokoll unter Anwendung

³ Die deutsche Version des EQ-5D-3L findet sich im Anhang 1 dieser Synopse.

⁴ Die deutsche Version des EQ-5D-5L findet sich im Anhang 2 dieser Synopse.

einer innovativen Methodenkombination aus composite Time-Trade-Off (cTTO) und Discrete Choice Experimenten (DCE) generiert [14, 16]. Die Anwendung und Erprobung des Methoden-Protokolls ist für den deutschen Kontext bisher allerdings noch ausstehend. Der aufgeführte Bedarf an geeigneten präferenzbasierten Tarifen für Lebensqualitätsinstrumente in Deutschland unterstreicht die Notwendigkeit der Ermittlung eines solchen Value Sets für die deutsche Version des erweiterten EQ-5D-5L inklusive der Testung des Methoden-Protokolls in Deutschland.

1.2 Konzept und Fragestellungen der Dissertation

Die übergeordnete Zielsetzung der vorliegenden Dissertation besteht darin, ein präferenzbasiertes Value Set für die deutsche Version des EQ-5D-5L für die deutsche Gesellschaft zu entwickeln. Die kumulative Dissertation setzt sich dabei aus folgenden sechs Publikationen zusammen:

- [Modul 1] **Klose, K.**, Kreimeier, S., Tangermann, U., Aumann, I., Damm, K. (2016): Patient- and person-reports on healthcare: preferences, outcomes, experiences, and satisfaction. *Health Economics Review*, 6(18) (DOI: 10.1186/s13561-016-0094-6).
- [Modul 2] Greiner, W., **Klose, K.** (2014): Lebensqualitätsbewertung und Utilities in der Gesundheitsökonomie. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 108(2-3): 120-125.
- [Modul 3] Kind, P., **Klose, K.**, Gusi, N., Olivares, P.R., Greiner, W. (2015): Can adult weights be used to value child health states? Testing the influence of perspective in valuing EQ-5D-Y. *Quality of Life Research*, 24(10): 2519-2539 (*empirische Studie 1*).
- [Modul 4] **Ludwig, K.**, Schulenburg, J-M. Graf v. d., Greiner, W. (2017): Valuation of the EQ-5D-5L with composite time trade-off for the German population - an exploratory study. *Health and Quality of Life Outcomes*, 15(39) (DOI: 10.1186/s12955-017-0617-9) (*empirische Studie 2*).
- [Modul 5] **Ludwig, K.**, Schulenburg, J-M. Graf v. d., Greiner, W. (2018): German Value Set for the EQ-5D-5L. *PharmacoEconomics*, 36(6): 663-674 (*empirische Studie 3*).
- [Modul 6] Stolk, E., **Ludwig, K.**, Rand, K., van Hout, B., Ramos Goñi, JM.: Overview, update and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. *Value in Health*. Akzeptiert zur Publikation.

Das Konzept der kumulativen Dissertation ist in Abbildung 1 visualisiert. Um die Zielsetzung der Dissertation zu erreichen, werden vier aufeinander aufbauende Hauptschritte verfolgt:

- I. Im ersten Schritt erfolgen theoretische und methodische Vorüberlegungen für die Durchführung einer sogenannten Valuierungsstudie zur Ermittlung eines Value Sets, wobei die Einordnung des Themas in den Teilbereich des Selbstberichts im Gesundheitswesen (*Modul 1*) sowie in den Anwendungsbereich der Gesundheitsökonomie erfolgt (*Modul 2*).
- II. Darauf aufbauend werden im zweiten Schritt zwei explorative empirische Vorstudien (*Module 3 und 4*) durchgeführt, um die methodische Vorgehensweise im deutschen Kontext zu testen und somit eine eigene empirische Basis für die folgende Hauptstudie zu generieren.
- III. In der anschließenden Valuierungsstudie für die deutsche Version des EQ-5D-5L steht die Ermittlung der gesellschaftlichen Präferenzwerte bzw. des Value Sets für den EQ-5D-5L für Deutschland im Fokus (*Modul 5*).
- IV. Auf Grundlage der generierten Erkenntnisse werden abschließend Implikationen für Forschung, Politik und Praxis abgeleitet (*Modul 6*).

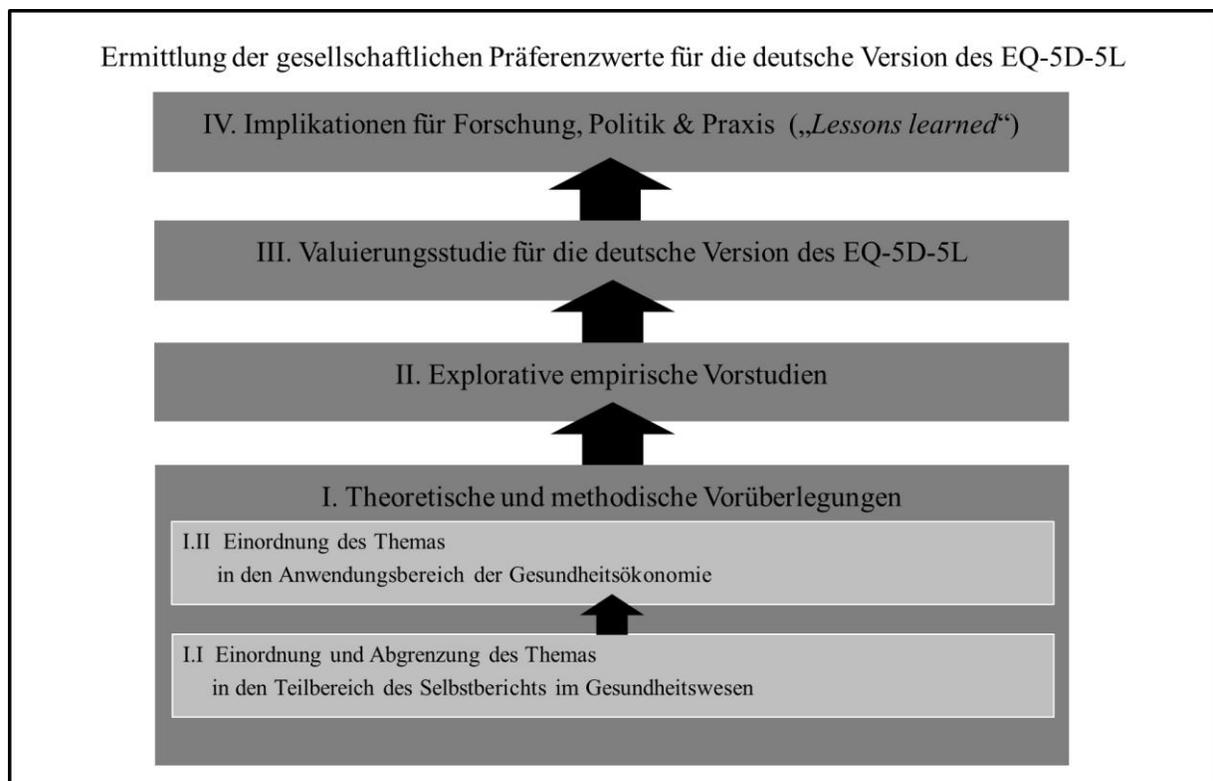


Abbildung 1: Konzept der Dissertation (Quelle: eigene Darstellung)

Der Fokus dieser Dissertation liegt darin, sowohl methodische als auch inhaltliche Aspekte in Bezug auf die Ermittlung der gesellschaftlichen Präferenzen zu gesundheitlichen Einschränkungen im deutschen Kontext zu erforschen. Als Personengruppe wird ausschließlich die deutsche Allgemeinbevölkerung einbezogen. Die Präferenzen von Patienten oder anderen Stakeholdern, wie z. B. Entscheidungsträgern, sind nicht Gegenstand dieser Dissertation. Zudem sei angemerkt, dass die Bewertung von gesundheitlichen Einschränkungen dabei in den Kontext der gesundheitsökonomischen Evaluationen als Anwendungsgebiet von Lebensqualitätsverfahren eingebettet wird.

Die zugrunde liegenden übergeordneten Fragestellungen der Dissertation lauten wie folgt:

1. Welche methodischen Entscheidungen sind bei einer Valuierungsstudie zu treffen und eignet sich das internationale Methoden-Protokoll für die Schätzung des EQ-5D-5L Index für die deutsche Bevölkerung?
2. Welchen Einfluss haben die Präferenzmethoden cTTO und DCE auf die zu ermittelnden Lebensqualitätsabschläge bei dem EQ-5D-5L?
3. Wie gewichtet die deutsche Bevölkerung die einzelnen Lebensqualitätsabschläge bei dem EQ-5D-5L? (*Leitfrage der Dissertation*)
4. Wie sollte das Methoden-Protokoll der EuroQol-Gruppe unter Berücksichtigung der verfügbaren Evidenz bestmöglich eingesetzt werden?

Der Aufbau der Synopse spiegelt den Gang der Dissertation wider, indem zunächst eine theoretische Basis zur Einordnung des Forschungsvorhabens gegeben wird und daran anschließend die Darstellung der empirischen Studien erfolgt. In dem nachfolgenden Kapitel 2 wird ein kurzer literaturbasierter Überblick über die Grundlagen der nutzentheoretischen Lebensqualitätsmessung im Rahmen gesundheitsökonomischer Evaluationen gegeben. In Kapitel 3 wird die methodische Vorgehensweise der empirischen Studien dargestellt, die dieser Dissertation zugrunde liegen. Anschließend erfolgt die Synthese der zentralen Erkenntnisse in Bezug auf die aufgeführten Fragestellungen dieser Dissertation (Kapitel 4). In Kapitel 5 erfolgt eine übergreifende ergebnis- sowie methodenorientierte Diskussion. Die Erkenntnisse der Dissertation werden dabei in den gegenwärtigen Stand der Forschung eingeordnet, die methodische Vorgehensweise wird kritisch reflektiert und mögliche Limitationen werden aufgezeigt. Die vorliegende Synopse schließt mit einer Darstellung der Implikationen für Forschung, Politik und Praxis, welche die Haupteckdaten dieser Dissertation zusammenführt (Kapitel 6).

2. Nutzentheoretische Lebensqualitätsmessung im Rahmen gesundheitsökonomischer Evaluationen

2.1 Theoretische Fundierung und Grundannahmen ökonomischer Analysen

Einer Bewertung von Kosten und Nutzen bei gesundheitsökonomischen Evaluationen liegen insbesondere folgende Grundannahmen einer ökonomischen Analyse zugrunde:

- (1) Den Ausgangspunkt der ökonomischen Grundannahmen bildet die (traditionelle) Wohlfahrtstheorie. Entsprechend des wohlfahrtsökonomischen Ansatzes ergibt die Summe des individuellen Nutzens (oder Wohlfahrt) aller Mitglieder einer Gesellschaft den gesellschaftlichen Nutzen (gesellschaftliche Wohlfahrt), den es zu maximieren gilt. Die Wohlfahrtsfunktion ergibt sich demnach aus der Summe der Nutzen aller Individuen einer Gesellschaft.
- (2) Eine weitere Grundannahme liegt in der kardinalen Messbarkeit des individuellen Nutzens als elementare Voraussetzung bei der Gegenüberstellung von Kosten und Nutzen in einem Quotienten. Dies impliziert gewisse Anforderungen an geeignete Messmethoden, insbesondere zur Nutzenquantifizierung (kardinale Nutzenmessung).
- (3) Eine relevante Grundannahme liegt ferner darin, dass eine Knappheit an Ressourcen vorliegt (Knappheitsprinzip). Aufgrund dessen sind ökonomische Analysen erforderlich, da es nicht möglich ist, alle Ziele gleichzeitig zu erreichen bzw. die Befriedigung aller Bedürfnisse zu realisieren. Zudem entstehen bei der Allokation der knappen Ressourcen Opportunitätskosten, d. h. die Kosten einer Handlungsalternative entsprechen dem Nutzen derjenigen Alternative, die aus Knappheitsgründen nicht gewählt werden kann (Opportunitäts- oder Alternativkostenkonzept) [17–19].

Da gesundheitsökonomische Evaluationen auf der ökonomischen Wohlfahrtstheorie basieren, soll auf diese nachfolgend vertiefend eingegangen werden. Das Ziel der Wohlfahrtstheorie liegt in der allokativen und technischen Effizienz des Einsatzes knapper Ressourcen. Effizienz liegt vor, wenn ein Optimum an gesellschaftlicher Wohlfahrt bei gegebenen Ressourcen erreicht ist. Die Wohlfahrtstheorie beschreibt, wie diese Wohlfahrt durch die Veränderung der Ressourcenallokation verändert werden kann. Dabei liegt der bekannteste Grundsatz in dem *ökonomischen Prinzip*. Um eine technische (d. h. auf die Produktion bezogene) Effizienz zu erreichen, sollte eine bestimmte nach Qualität und Umfang definierte Leistung mit möglichst geringem Ressourceneinsatz erstellt werden (Minimalprinzip). Es wird von Kosten-Effektivität

gesprochen, wenn die Preise für die erforderlichen Ressourcen einbezogen werden. Andererseits kann gemäß des ökonomischen Prinzips auch angestrebt werden, bei vorgegebenem Budget ein maximales Ergebnis zu erzielen (Maximalprinzip), wie z. B. eine maximal gute Gesundheit im Gesundheitswesen. Bei der *allokativen Effizienz* erfolgt eine Ergänzung von Verteilungsvorstellungen. Es wird zusätzlich berücksichtigt, mit welcher Effizienz die Güter und Leistungen zur Maximierung der Wohlfahrt in der Gesellschaft verteilt werden. Gesundheitsökonomische Evaluationen fokussieren insbesondere die *technische Effizienz*. Letztere liegt vor, wenn Güter, Dienstleistungen oder Gesundheitsprogramme bei einem gegebenen Budget, wie z. B. das festgelegte Arzneimittelbudget einzelner Arztpraxen, ein Höchstmaß an gesundheitlichem Nutzen erzeugen. Falls es Alternativen gibt, die dieses Ergebnis mit einem geringeren Ressourceneinsatz erzielen, werden diese als effizienter eingestuft und daher vorgezogen [20–22].

Für die Quantifizierung der Wohlfahrt bzw. eines Wohlfahrtsgewinns wird ein Maß benötigt. In Anlehnung an die Nutzentheorie nach von Neumann und Morgenstern [23] kann der Nutzen einer Alternative aus Sicht eines Individuums aus seinen Wahlentscheidungen für Alternativen abgeleitet werden. Es wird dabei angenommen, dass Individuen rational handeln und sich für diejenige Alternative entscheiden, die ihnen voraussichtlich den größtmöglichen Nutzen stiftet. Die Messung kann zum einen durch die Beobachtung und Analyse von Präferenzen erfolgen, welche z. B. durch getätigte (Kauf-) Entscheidungen aufgedeckt werden: d. h., wenn ein Individuum sich für A anstelle von B entscheidet, wird davon ausgegangen, dass seine Wohlfahrt bei A höher ist als bei B (*Revealed-Preference-Methoden*). Zum anderen kann die Messung von Präferenzen durch eine Bewertung von hypothetischen Alternativen (z. B. Gesundheitsleistungen) durch Individuen erfolgen, die durch verschiedene Merkmale und Merkmalsausprägungen dargestellt werden (*Stated-Preference-Methoden*) [24–27]. Da letztere im Fokus dieser Dissertation stehen, werden jene Methoden der Präferenzmessung im Zuge einer eigens entwickelten Klassifikation von Patienten- und Personenberichten im Gesundheitswesen in *Modul 1* dargestellt.

Gesundheitsökonomische Evaluationen werden als praktische Anwendung der Wohlfahrtsökonomie eingestuft. Zur Einordnung der vergleichenden Studienformen der gesundheitsökonomischen Evaluation kann die Unterscheidung zwischen dem Welfarismus und dem Extra-Welfarimus hinzugezogen werden. Bei dem *Welfarismus* (basierend auf der traditionellen Wohlfahrtsökonomie) geht jedes Individuum mit seinem individuellen Nutzen in die

kollektive Entscheidung ein. Das Ziel des Systems liegt in einem größtmöglichen Beitrag zur gesamtgesellschaftlichen Wohlfahrt. Es wird dabei gefordert, den gesamten Nutzen einer Person als Grundlage zu verwenden. Der Nutzen steht im Zentrum des Entscheidungsträgers, z. B. die Ermittlung von *Willingness-to-pay* im Rahmen von Kosten-Nutzen-Analysen [18, 22, 28]. Gemäß des welfaristischen Ansatzes werden die Individuen selbst als einzige Berechtigte angesehen, um ihre eigene Wohlfahrt zu quantifizieren (Konsumentensouveränität). Nur sie können entscheiden, ob ihre Wohlfahrt sich verbessert hat oder nicht, und treffen Entscheidungen basierend auf ihren Präferenzen, um ihre eigene Wohlfahrt zu verbessern. Es wird davon ausgegangen, dass diejenige Alternative ausgewählt wird, die nach eigenem Kenntnisstand den eigenen Nutzen bei gegebenen Kosten maximiert bzw. ein Ziel mit minimalen Kosten verwirklicht [24, 27, 29, 30].

Der Ansatz des *Extra-Welfarismus* wendet sich gegen die Berücksichtigung des (gesamten) individuellen Nutzens als alleinige Basis für kollektive Entscheidungen. Es kommt demnach vielmehr darauf an, Voraussetzungen für ein gutes Leben zu schaffen, wobei Gesundheit eine zentrale Komponente darstellt. Das Ziel des Systems stellt folglich die größtmögliche Produktion von Gesundheit dar [22, 28, 30, 31]. Dabei werden neben Gütern und Dienstleistungen, die zur individuellen Wohlfahrt beitragen können, auch intangible Komponenten berücksichtigt, wie z. B. Indikatoren von Wohlbefinden [31]. Im Falle gesundheitsökonomischer Evaluationen liegt das Ziel darin, das zu maximieren, was für Entscheidungsträger entscheidungs- und budgetrelevant ist (z. B. Maximierung der aggregierten Gesundheitseffekte bei einem gegebenen Budget). Hier setzen Kosten-Wirksamkeits- und Kosten-Nutzwert-Analysen als praktische Anwendung der Wohlfahrtstheorie an, um die Alternative mit dem besten Kosten-Nutzen-Verhältnis zu realisieren [18, 22]. Bei der Evaluationsform der Kosten-Nutzwert-Analysen erfolgt die Messung des Gesundheitszustandes überwiegend bei dem betroffenen Individuum selbst. Allerdings erfolgt die Bewertung des Gesundheitszustandes und somit die Ableitung der Nutzwerte⁵ (d. h. der Wert, den ein Individuum einem Gesundheitszustand zuschreibt) oftmals unter Hinzuziehung von Populationen, wie z. B. die Allgemeinbevölkerung. Im Gegensatz zu dem welfaristischen Ansatz hat eine Abkehr der alleinigen Berücksichtigung des Individuums hin zu gesellschaftlichen Bewertungen stattgefunden [25, 29, 31]. Letztere bilden den Ansatzpunkt für die Ermittlung gesellschaftlicher Präferenzen, die im Fokus dieser Dissertation stehen.

⁵ Die Begriffe Nutzwerte und *Utilities* werden in diesem Kontext oftmals synonym verwendet. Eine umfassende definitorische Abhandlung ist u. a. Drummond et al. [24] zu entnehmen.

2.2 Lebensqualität als Outcomeparameter in gesundheitsökonomischen Evaluationen

Im Zuge der steigenden Bedeutung der Lebensqualität im Gesundheitswesen, haben Lebensqualitätseffekte Eingang in den Bereich der gesundheitsökonomischen Evaluationen gefunden. *Modul 2* gibt einen detaillierten Überblick über die Einbezugsmöglichkeiten von Lebensqualitätsdaten bei den verschiedenen Studienformen. So kann (1) verlorene Lebensqualität als Kostenparameter in Kosten- und Krankheitskosten-Analysen eingehen. Ebenso kann Lebensqualität als (2) Vergleichsparameter zur Überprüfung der Übereinstimmung des Nutzens verschiedener Alternativen bei Kosten-Kosten-Analysen hinzugezogen werden. Zudem kann (3) gewonnene Lebensqualität als Ergebnisparameter in Kosten-Wirksamkeits- und Kosten-Nutzwert-Analysen einbezogen werden (z. B. Kosten pro *quality adjusted life year* [QALY]) [32].

Da im Rahmen dieser Dissertation ein Instrument weiterentwickelt wird, um dessen Einsatz in Kosten-Nutzwert-Analysen zu ermöglichen, wird nachfolgend auf diese Evaluationsform vertiefend eingegangen. Der Einsatz von Kosten-Nutzwert-Analysen mit der Messung der aufgeführten QALYs zur Nutzenquantifizierung wird in diversen gesundheitsökonomischen Leitlinien empfohlen [24, 33, 34]. Das QALY-Konzept ist dabei das am häufigsten angewendete Verfahren zur Ermittlung von Nutzwerten [24, 32]. Es erfolgt dabei eine Aggregation der Lebensdauer und der Lebensqualität zu einem eindimensionalen Outcome-Maß. Die quantitative Zeitkomponente umfasst vielfach die Restlebenserwartung vom Zeitpunkt der Beobachtung bis zum Tod des Individuums oder aber den befristeten Zeithorizont unter Erkrankung. Die qualitative Lebensqualitätskomponente ist durch die Werte 1 (vollständige Gesundheit) und 0 (tot) normiert [24, 35–37]. Die Erhebung der Lebensqualität erfolgt dabei durch (kardinale) Lebensqualitätsinstrumente häufig in Form von Selbstangaben des Patienten⁶, die dann anhand spezifischer Berechnungsvorschriften in Nutzwerte umgerechnet werden. Jene Value Sets basieren auf der Bewertung der Lebensqualität gemäß den Präferenzen von Individuen (nutzentheoretische bzw. präferenzbasierte Lebensqualitätsmessung) [2, 24, 25]. Teilweise wird in diesem Kontext auch auf Expertenschätzungen oder bestehende Werte aus der Literatur statt auf eigene Lebensqualitätsbefragungen zurückgegriffen [25, 35, 37].

Das QALY-Konzept beruht auf der aufgeführten Nutzentheorie nach von Neumann und Morgenstern [23]. Ein QALY stellt ein normiertes Nutzenmaß dar und ermöglicht

⁶ Es werden teilweise auch *Proxy*-Versionen eingesetzt, wenn Patienten nicht in der Lage sind, ihre Lebensqualität eigenständig zu beurteilen (d. h. Fremdeinschätzung durch Dritte, wie z. B. Angehörige oder Ärzte) [9, 25].

indikationsübergreifende Vergleiche im Gesundheitswesen. Aus ökonomischer Sicht sollte die Therapiemaßnahme ausgewählt werden, bei der die Kosten pro (normiertem) Nutzwert minimal sind (Minimalprinzip) oder die Anzahl der (normierten) Nutzwerte pro festgelegte Geldeinheit maximal ist (Maximalprinzip; Maximierung der QALYs als extra-welfaristische Zielorientierung) [17, 35]. Es sei angemerkt, dass das QALY-Konzept nicht dafür geeignet ist, über die Behandlung einzelner Personen zu entscheiden, sondern es werden ausschließlich Allokationsentscheidungen für zukünftige Patienten fokussiert, z. B. ob ein Verfahren generell in einem Gesundheitssystem zur Verfügung gestellt werden soll [35] (bevölkerungsbezogene Perspektive der Gesundheitswissenschaften).

Das QALY-Konzept basiert auf den aufgeführten Annahmen der Wohlfahrtstheorie und bedarf weiterer impliziter Annahmen zur Nutzenfunktion von Individuen⁷. Es wird davon ausgegangen, dass ein Individuum risikoneutral in Bezug auf die Lebensdauer ist und dass sich der Nutzen eines Individuums über die Zeit hinweg addiert. Ferner wird angenommen, dass QALYs über Individuen hinweg aggregiert werden können, d. h. ein QALY ist ein QALY unabhängig davon, wer dieses gewinnt oder verliert. Ein gewonnenes QALY hat dabei in jedem Lebensalter denselben Wert, unabhängig von der verbleibenden Restlebensdauer. Die Bewertung eines Gesundheitszustandes sei unabhängig von seiner Dauer. Zudem wird das Ausgangsniveau, von dem aus die Lebensqualitätsverbesserung erfolgt, nicht berücksichtigt [24, 35, 36]. Jene Annahmen finden oftmals eine kritische Betrachtung sowohl aus methodischen Gründen (z. B. Nicht-Berücksichtigung des Gesetzes des abnehmenden Grenznutzens) als auch unter ethischen Gesichtspunkten (z. B. Benachteiligung von Älteren, bei denen weniger QALY-Gewinne zu erwarten sind) [17, 35, 38]. Darüber hinaus werden u. a. folgende Kritikpunkte⁸ für das QALY-Konzept genannt: Die für eine bessere Entscheidungssituation erforderliche Komplexitätsreduktion geht mit einem Informationsverlust gegenüber einer detaillierteren Lebensqualitätserfassung einher. Des Weiteren wird oftmals bemängelt, dass die epidemiologische Datengrundlage über die Lebensqualitätsveränderungen von Patienten, auf welchen die QALY-Ermittlung beruht, für den deutschen Kontext nicht ausreichend bzw. zu ungenau ist. Hinzu kommt, dass bislang keine einheitliche Methodik zur Messung und Bewertung der Lebensqualität existiert. Dies limitiert die Vergleichbarkeit der resultierenden Ergebnisse solcher Studien. Folglich wird eine

⁷ Eine ausführliche Darstellung der Annahmen zur Nutzenfunktion ist u. a. Weinstein et al. [36] zu entnehmen.

⁸ Eine ausführliche kritische Auseinandersetzung mit dem QALY-Konzept sowie möglichen Alternativen ist u. a. Schöffski und Greiner [35] zu entnehmen.

Weiterentwicklung bzw. Standardisierung der Erhebungsmethodik gefordert [17, 35]. Die Weiterentwicklung einer solchen Erhebungsmethodik für den deutschen Kontext bildet den Ansatzpunkt dieser Dissertation.

Die aufgeführten Kritikpunkte implizieren, dass die Anwendung des QALY-Konzeptes nicht problemlos möglich ist und dass Forschungsbedarf zu der Weiterentwicklung des QALY-Konzeptes besteht [39, 40]. Dennoch bewerten Schöffski und Greiner [35] das QALY-Konzept wie folgt: „*Trotzdem stellt dieses Konzept zumindest theoretisch eine gute Möglichkeit dar, die Ressourcenallokation im Gesundheitswesen auf eine objektivere Basis zu stellen. Es ist nicht unwahrscheinlich, dass bei einer konsequenten Allokation nach Kosten pro Nutzwert mehr QALYs und damit mehr Gesundheit ‚eingekauft‘ werden könnte*“ (S. 97). Wenngleich die Praktikabilität und die Anwendungsmöglichkeiten im deutschen Kontext kontrovers diskutiert werden, stellen Kosten-Nutzwert-Analysen bzw. das QALY-Konzept international einen de facto-Standard in der gesundheitsökonomischen Evaluation dar [2, 41]. Der QALY-Ansatz wird u. a. von dem *National Institute for Health and Care Excellence* (NICE), der *Canadian Agency for Drugs and Technologies in Health* (CADTH) und dem *Medical Services Advisory Committee* (MSAC) zur Unterstützung von Allokationsentscheidungen im Gesundheitswesen empfohlen und wird international häufig in randomisiert kontrollierten Studien zur Evaluation von Behandlungsverfahren verwendet [42].

2.3 Vom Fragebogen zum Nutzwert mit dem Methoden-Protokoll der EuroQol-Gruppe

Den bisherigen Ausführungen konnte entnommen werden, dass die Einbeziehung von Lebensqualitätsdaten in Kosten-Nutzwert-Analysen die Ableitung von Gewichtungsfaktoren erfordert, um die einzelnen Lebensqualitätsdimensionen zu einer singulären Maßzahl zu aggregieren. Es existieren dabei folgende drei Möglichkeiten, um diese Gewichtungsfaktoren abzuleiten: (1) durch Plausibilitätsüberlegungen (*judgements*), (2) durch die Entnahme bereits bestehender Werte aus der Literatur oder (3) durch das eigene Messen jener Werte. Für die zuletzt genannte Option besteht die Möglichkeit entweder zur direkten oder indirekten Erhebung der (relativen) Präferenzen bezüglich der entsprechenden Gesundheitszustände, die mit einem Lebensqualitätsinstrument beschrieben werden [27, 43]. Im Rahmen der *direkten Nutzwertbestimmung* werden die aufgeführten Stated-Preference-Methoden zur direkten Bewertung von Gesundheitszuständen eingesetzt, wie z. B. das Time-Trade-Off (TTO)-Verfahren, das im weiteren

Verlauf der Synopse näher erläutert wird. Dabei werden Individuen gebeten, Präferenzurteile zu ausgewählten Gesundheitszuständen zu geben.⁹ Dem Vorgehen liegt die Annahme zugrunde, dass je eher ein Gesundheitszustand gegenüber einem anderen präferiert wird, desto größer ist der Nutzen, der diesem Gesundheitszustand zugesprochen wird (siehe Kapitel 2.1). Anschließend lassen sich aus diesen Präferenzurteilen die gesuchten relativen Gewichtungsfaktoren ermitteln [4, 27, 44]. *Modul 1* gibt einen Überblick über Präferenzmethoden sowie deren theoretische Fundierung, die in diesem Kontext eingesetzt werden.

Die *indirekte Nutzwertbestimmung* erfolgt hingegen über standardisierte Fragebögen (nutzentheoretische Indexinstrumente). Folgende drei Instrumente zählen zu den am häufigsten eingesetzten Lebensqualitätsinstrumenten zur indirekten Bestimmung von Nutzwerten: der EQ-5D, der Health Utility Index (HUI) und der Short Form-6D (SF-6D) [24, 25, 44]. Bei der indirekten Nutzwertbestimmung wird mit einem Referenzkollektiv eine Valuiierungsstudie durchgeführt, bei welcher eine nutzwertbezogene Normierung der Gesundheitszustände auf Basis der Bewertungen der befragten Personen erfolgt. Es finden dabei die Methoden der direkten Nutzwertbestimmung Anwendung [34, 44, 45]. Mithilfe von statistischen Auswertungen lassen sich die mit den gesundheitlichen Einschränkungen assoziierten Abschlagswerte bzw. Nutzwertverluste quantifizieren. Anhand des resultierenden Tarifes bzw. Value Sets ist es möglich, allen mit dem Lebensqualitätsinstrument erfassbaren möglichen Gesundheitszuständen einen Nutzwert (indirekt) zuzuordnen [25, 44].

Die einleitend aufgeführte Entwicklung des EQ-5D-5L erfordert die Ermittlung von Gewichtungsabschlägen, um das Messinstrument zur indirekten Nutzwertbestimmung und somit in Kosten-Nutzwert-Analysen einsetzen zu können. Der EQ-5D-5L beinhaltet fünf Lebensqualitätsdimensionen (Beweglichkeit/Mobilität, Für sich selbst sorgen, Alltägliche Tätigkeiten, Schmerzen/Körperliche Beschwerden, Angst/Niedergeschlagenheit), die durch jeweils fünf Antwortlevel beschrieben werden (keine [Level 1], leichte [Level 2], mäßige [Level 3], starke [Level 4] und extreme Probleme/nicht in der Lage [Level 5]). Durch die Kombination der Dimensionen mit den Antwortleveln lassen sich 3.125 Gesundheitszustände konstruieren. Im Gegensatz zu dem EQ-5D-3L [46] existiert bis dato kein spezifisches Value Set für die deutsche Version des EQ-5D-5L, das gesellschaftliche Präferenzwerte auf Basis der von der EuroQol-

⁹ Bei der beschriebenen Vorgehensweise wird davon ausgegangen, dass nicht-betroffene Probanden hypothetische Gesundheitszustände bewerten. Ebenso kann ein Nutzwert eines „realen“ Gesundheitszustandes durch die Befragung des betroffenen Patienten selbst bestimmt werden [27, 44].

Gruppe empfohlenen methodischen Vorgehensweise ermittelt hat.¹⁰ Derzeit ist ausschließlich ein *Crosswalk-Value Set* als Übergangslösung verfügbar, das auf einem Mapping-Algorithmus zwischen den beiden Fragebogenversionen basiert [48]. Da solche Mapping-Verfahren Limitationen aufweisen [14, 16, 48], kann der EQ-5D-5L nur eingeschränkt als Alternative für den EQ-5D-3L in Deutschland eingesetzt werden. Es besteht daher ein Forschungsdesiderat an einer (neuen) nationalen Valuierungsstudie, um einen auf den Präferenzen der deutschen Bevölkerung basierenden spezifischen Scoring-Algorithmus zur Aggregation der Antworten des erweiterten EQ-5D-5L in einen präferenzbasierten Index abzuleiten. Somit kann jedem der 3.125 möglichen Gesundheitszustände ein singulärer Nutzwert zugeordnet werden [15].

Die Entwicklung des EQ-5D-5L wurde als Möglichkeit gesehen, ein international einheitliches Methoden-Protokoll, inklusive der Erprobung neuer Präferenzmethoden für die Bewertung der Gesundheitszustände und eine dazugehörige Erhebungs-Software (*EuroQol-Valuation Technique*, EQ-VT), zu entwickeln. Das Methoden-Protokoll ermöglicht einen Vergleich der resultierenden nationalen EQ-5D-5L Value Sets im Gegensatz zu den EQ-5D-3L Value Sets, bei welchen inkonsistente Studiendesigns und Methoden für die Valuierung der Gesundheitszustände verwendet wurden [14, 16]. Für die Ermittlung der EQ-5D-3L Value Sets wurde insbesondere das TTO-Verfahren eingesetzt [16, 49].¹¹ Bei dem klassischen TTO werden Individuen gebeten, Lebensjahre in eingeschränkter Gesundheit gegen eine kürzere Lebensspanne in voller Gesundheit abzuwägen [25, 43]. Zur Entwicklung des standardisierten Methoden-Protokolls wurden verschiedene internationale Pilotstudien zur Verbesserung der Valuierungsmethodik durchgeführt [16]. Bei dem bisherigen Einsatz des klassischen TTO-Verfahrens wurde insbesondere der Umgang mit negativen Werten kritisiert (d. h. Transformation der [teilweise großen] negativen Werte zu einer Spannbreite mit -1 als kleinstem Wert) [16, 51, 52]. Vor diesem Hintergrund wurden alternative TTO-Methoden getestet [53, 54] und die Kombinationsmöglichkeiten mit DCE überprüft [55]. Bei der DCE-Methode erhält der Teilnehmer die Aufgabe, eine diskrete Entscheidung zwischen zwei oder mehreren diskreten Alternativen, wie z. B.

¹⁰ Leidl und Reitmeir [47] haben kürzlich ein exploratives Value Set für den deutschen EQ-5D-5L entwickelt, das einen anderen methodischen Ansatz verfolgt als der hier vorgestellte. So wurden gesellschaftliche Präferenzwerte für *erlebte* Gesundheitszustände mittels visueller Analogskala (VAS) erhoben. Im Rahmen des Methoden-Protokolls der EuroQol-Gruppe werden jedoch hypothetische Gesundheitszustände unter Anwendung von cTTO und DCE bewertet. Bei dem VAS-Verfahren werden die mangelnde Kardinalität und somit die Verletzung der in Kapitel 2.1 aufgeführten zweiten ökonomischen Grundannahme kontrovers diskutiert (*Modul 2*). Kapitel 4.1 ist ferner zu entnehmen, warum der Ansatz der hypothetischen Gesundheitszustände im Kontext von Allokationsentscheidungen vorzuziehen ist.

¹¹ Eine detaillierte Abhandlung der Forschungsaktivitäten zur Valuierung des EQ-5D ist Brooks [50] zu entnehmen.

Gesundheitszuständen, entsprechend seiner persönlichen Präferenz zu treffen [25, 56, 57]. Die Erkenntnisse der Pilotstudien mündeten in die Entwicklung des Methoden-Protokolls 1.0 für die EQ-5D-5L Valuierungsstudien von Oppe et al. [16]. Eine ausführliche Darstellung des Methoden-Protokolls ist den *Modulen 4 und 5* zu entnehmen.

Bei dem Methoden-Protokoll wurde eine innovative Kombination aus cTTO und DCE integriert. Das cTTO besteht aus (1) dem aufgeführten klassischen TTO für Gesundheitszustände, die besser als der Tod, und (2) dem *lead-time* TTO (LT-TTO) für Gesundheitszustände, die schlechter als der Tod bewertet werden. Bei den DCE-Aufgaben soll sich der Teilnehmer ohne Angabe einer Zeitkomponente zwischen zwei EQ-5D-5L Gesundheitszuständen entscheiden. Bei dem klassischen TTO hat der Befragte die Auswahl zwischen zwei Leben: (Leben A) 10 Jahre in einem EQ-5D-5L Gesundheitszustand mit anschließendem Tod oder (Leben B) die Dauer x bei vollständiger Gesundheit mit anschließendem Tod zu verbringen (mit $x \leq t$). Die Zeitangabe x in Leben B wird so lange variiert bis der Teilnehmer indifferent zwischen beiden Alternativen ist. Die LT-TTO-Aufgaben folgen dem gleichen Prinzip, allerdings beginnt der zu bewertende EQ-5D-5L Gesundheitszustand erst nach zehn Jahren bei voller Gesundheit und es ist keine Transformation der negativen Werte (mehr) erforderlich [16, 58]. Das Methoden-Protokoll 1.0 wurde 2012/2013 erstmalig zur Ermittlung von nationalen EQ-5D-5L Value Sets eingesetzt. Länderübergreifend wurden u. a. folgende Datenauffälligkeiten insbesondere bei dem cTTO festgestellt: ein hoher Anteil inkonsistenter Bewertungen, hohe Abschlagswerte für milde Gesundheitszustände, Ansammlung von cTTO-Bewertungen an gewissen Datenpunkten („Spikes“), Protokoll-Abweichungen durch die Interviewer sowie starke Interviewer-Effekte [59, 60]. Eine Erörterung der Datenauffälligkeiten der ersten Länderstudien erfolgt in *Modul 4*. Auf Grundlage der Erfahrungen der ersten Länderstudien wurden Verbesserungsvorschläge für die Umsetzung des cTTO in der EQ-VT herausgearbeitet. Dies führte zu einer Weiterentwicklung des Methoden-Protokolls zu Version 1.1 (Einführung von Übungs-cTTO-Aufgaben, Pop-Up-Fenstern zur Verifizierung der cTTO-Antworten und einer Qualitätskontrolle) [59]. Auf Basis eines anschließenden internationalen Forschungsprogramms wurde das verbesserte Methoden-Protokoll 2.0 beschlossen, das in *Modul 5* dargestellt wird. Bei den verschiedenen Protokoll-Versionen blieben die Präferenzmethoden bestehen, allerdings wurde bei den weiterentwickelten Versionen der Fokus auf eine verbesserte Implementierung der Methoden sowie auf eine Einführung einer Qualitätskontrolle gelegt [60].

3. Methodisches Vorgehen der empirischen Studien

Neben der theoretischen Einbettung des Dissertationsthemas in den Selbstbericht im Gesundheitswesen in *Modul 1* und den Bereich der Gesundheitsökonomie in *Modul 2*, liegen dieser Dissertation drei empirische Einzelstudien zur Bearbeitung der aufgeführten Fragestellungen zugrunde. Tabelle 1 gibt einen Überblick über das methodische Vorgehen dieser Studien, bei denen jeweils Präferenzwerte der deutschen Allgemeinbevölkerung zu dem EQ-5D erforscht wurden. Eine ausführliche Darstellung der Methodik ist den *Modulen 3 bis 5* zu entnehmen.

Parallel zur Entwicklung des Methoden-Protokolls wurde in der *explorativen Vorstudie 1 (Modul 3)* überprüft, ob spezifische Value Sets für die kinderfreundliche Version des EQ-5D, den EQ-5D-Youth (EQ-5D-Y) [61], notwendig sind oder ob Synergieeffekte durch die Nutzung der Value Sets der Erwachsenen-Version erzielt werden können. Vor diesem Hintergrund wurde die weltweit erste internationale Pilotstudie zeitgleich in Deutschland, England und Spanien zur Valuierung des EQ-5D-Y durchgeführt, um zunächst den Effekt des Perspektivwechsels – von sich selbst zu anderen – bei der Bewertung von Gesundheitszuständen zu überprüfen. Dabei ist von besonderem Interesse, ob sich die Bewertung der Gesundheitszustände in einem stärkeren Ausmaß ändert, wenn die Anderen nicht als Erwachsene sondern als Kinder beschrieben werden. Es wurde eine schriftliche Befragung durchgeführt, bei welcher ausgewählte EQ-5D-Y Gesundheitszustände mittels einer visuellen Analogskala (VAS) aus drei verschiedenen Perspektiven bewertet wurden: (a) für sich selbst, (b) für einen anderen Erwachsenen oder (c) für ein 10-jähriges Kind. Die englische Version des eingesetzten Fragebogens ist *Modul 3* zu entnehmen.

Da die Entwicklung des Methoden-Protokolls im Ausland erfolgte, wurde im Rahmen der *explorativen Vorstudie 2 (Modul 4)* überprüft, ob sich das Protokoll für die deutsche Bevölkerung eignet. Dabei ist neben der generellen Eignung der Präferenzmethoden, die kritische Reflexion deren Implementierung in der Valuierungsstudie von Interesse. Als Teilstudie des aufgeführten internationalen Forschungsprogramms wurde das (modifizierte) Methoden-Protokoll 1.1 inklusive einer innovativen Qualitätskontrolle in der Kontrollgruppe sowie weiterer Modifikationen in der Testgruppe im deutschen Kontext getestet: (a) *Ranking*-Aufgabe als Aufwärmübung, (b) Trennung der cTTO-Aufgaben für Gesundheitszustände, die besser bzw. schlechter als der Tod bewertet wurden und (c) Verifizierung der cTTO-Antworten in einem *Feedback*-Modul. Es wurde eine Stichprobe der Allgemeinbevölkerung von fünf geschulten Interviewern in den Regionen Bielefeld und Hannover befragt. Die Erkenntnisse dieser Vorstudie mündeten in die Entwicklung

des Methoden-Protokolls 2.0 und bildeten die empirische Basis für die methodische Ausgestaltung der anschließenden EQ-5D-5L Valuierungsstudie in Deutschland.

Tabelle 1: Methodik der empirischen Studien der Dissertation (Quelle: eigene Darstellung)

| | Empirische Studie 1 (Modul 3) | Empirische Studie 2 (Modul 4) | Empirische Studie 3 (Modul 5) |
|------------------------------|--|---|---|
| Art der Studie | explorative Vorstudie | explorative Vorstudie | Hauptstudie |
| Jahr der Erhebung | 2011 | 2014 | 2014/2015 |
| Land/Länder der Erhebung | Deutschland, England, Spanien | Deutschland | Deutschland |
| Studienort/e in Deutschland | 1. Region Hannover und Bielefeld, 2. deutschlandweit | Region Hannover und Bielefeld | deutschlandweit |
| Stichprobengröße | N=1.085 (N=407 in Deutschland) | N=200 | N=1.158 |
| Art der Stichprobe | Allgemeinbevölkerung | Allgemeinbevölkerung | Allgemeinbevölkerung (repräsentativ) |
| Rekrutierung der Stichprobe | Familie und Freunde, zufällige Einladungen anhand von Telefonbucheinträgen | Zeitungsannonce, Aushänge und Email-Einladungen | quotenbasiert (Marktforschungsinstitut) |
| Einschlusskriterien | Mindestalter: 18 Jahre | Mindestalter: 18 Jahre | Mindestalter: 18 Jahre |
| Ausschlusskriterien | nicht ausreichende Deutschkenntnisse | nicht ausreichende Deutschkenntnisse | nicht ausreichende Deutschkenntnisse |
| Primäres Studienziel | Überprüfung der Notwendigkeit eines gesonderten Value Sets für den EQ-5D-Y | Testung des verbesserten Methoden-Protokolls und weiterer Modifikationen im deutschen Kontext | Ermittlung des Value Sets für den EQ-5D-5L in Deutschland (deutsches Value Set) |
| Methoden-Protokoll | Nein | Methoden-Protokoll 1.1 der EuroQol-Gruppe plus 3 weitere Modifikationen | Methoden-Protokoll 2.0 der EuroQol-Gruppe |
| Messinstrument | EQ-5D-Y | EQ-5D-5L | EQ-5D-5L |
| Art der Erhebung | schriftliche Befragung | CAPI | CAPI |
| Präferenzmethode(n) | VAS | cTTO + DCE | cTTO + DCE |
| Art des Gesundheitszustandes | hypothetischer Gesundheitszustand | hypothetischer Gesundheitszustand | hypothetischer Gesundheitszustand |
| Perspektive/Referenzperson | 1. die befragte Person selbst, 2. ein anderer Erwachsener, 3. ein 10-jähriges Kind | die befragte Person selbst | die befragte Person selbst |
| Anzahl der Präferenzaufgaben | 16 VAS-Aufgaben | 10 cTTO- und 7 DCE-Aufgaben | 10 cTTO- und 7 DCE-Aufgaben, 6 explorative DCE-Aufgaben mit Zeitkomponente |
| Dauer der Erhebung | nicht bekannt | ca. 59 Minuten | ca. 48 Minuten |
| Qualitative Nachbefragung | nein (anonyme Befragung) | ja (Vollerhebung; Antwortrate: 48%) | ja (Stichprobe von 10%) |
| Qualitätskontrolle | nein | ja | ja |

CAPI: Computer Assisted Personal Interviewing, cTTO: composite Time-Trade-Off, DCE: Discrete Choice Experiment, VAS: visuelle Analogskala, EQ-5D-Y: EQ-5D-Youth, EQ-5D-5L: fünf-Level-Version des EQ-5D

Bei der dritten Studie bzw. der Hauptstudie dieser Dissertation wurde die *deutsche Valuierungsstudie für den EQ-5D-5L* durchgeführt (Modul 5). Dabei wurde weltweit erstmals das verbesserte internationale EQ-5D-5L Methoden-Protokoll 2.0 angewendet. Die Implementierung der cTTO- und DCE-Aufgaben in der deutschen EQ-VT ist dem Anhang 3 dieser Synopse zu entnehmen. Es wurde eine repräsentative Stichprobe der deutschen Bevölkerung von 19 geschulten Interviewern deutschlandweit befragt. Im Zuge der in der Vorstudie erprobten Qualitätskontrolle wurden täglich die Performance und Compliance der Interviewer entsprechend der Interviewer-Anleitungen analysiert. Mittels Regressionsverfahren wurden die Abschlagswerte für die 3.125 möglichen Gesundheitszustände auf Grundlage der cTTO- und DCE-Daten sowie der Kombination beider Datenformen in einem Hybrid-Modell ermittelt. Anschließend wurde das Value Set für die deutsche Version des EQ-5D-5L selektiert.

4. Ergebnissynthese: Ermittlung gesellschaftlicher Präferenzwerte für die deutsche Version des EQ-5D-5L

4.1 Methodische Entscheidungen und Eignung des Methoden-Protokolls für den deutschen Kontext

Die Ausgestaltung einer Valuierungsstudie für die Valuierung eines Indexinstrumentes zur Messung der Lebensqualität erfordert wichtige methodische Entscheidungen [62]. Hinzu kommt, dass das Methoden-Protokoll ausschließlich außerhalb von Deutschland entwickelt und getestet wurde. Vor diesem Hintergrund wurde im Rahmen dieser Dissertationsschrift der folgenden Fragestellung nachgegangen: Welche methodischen Entscheidungen sind bei einer Valuierungsstudie zu treffen und eignet sich das internationale Methoden-Protokoll für die Schätzung des EQ-5D-5L Index für die deutsche Bevölkerung?

Tabelle 2 gibt einen Überblick über die methodischen Entscheidungen, die bei einer Face-to-Face-Studie zur Valuierung eines Lebensqualitätsinstrumentes zu treffen sind und welche Aspekte das Methoden-Protokoll der EuroQol-Gruppe davon umfasst. Die *Module 1 bis 4* haben sich dabei theoretisch bzw. empirisch mit folgenden methodischen Kernaspekten auseinandergesetzt:

1. Methoden zur Präferenzmessung (*Module 1 und 2*),
2. Referenzkollektiv und Art des zu bewertenden Gesundheitszustandes (*Modul 2*),
3. Perspektive bzw. Referenzperson bei der Valuierungsaufgabe (*Modul 3*) und
4. Eignung eines Methoden-Protokolls für den nationalen Kontext (*Modul 4*).

Eine wesentliche Entscheidung liegt in der Wahl der *Präferenzmethode*, da die resultierenden Lebensqualitätsgewichte von der jeweiligen Methode abhängen und sich diese teilweise stark unterscheiden [27, 63]. So legt *Modul 1* dar, welche Präferenzmethoden generell zum Selbstbericht im Gesundheitswesen existieren und wie diese voneinander abzugrenzen sind. Dabei werden insbesondere die zugrunde liegende ökonomische Theorie, die Messung sowie die Analysemöglichkeiten der einzelnen Methoden erörtert. In *Modul 2* erfolgt eine vertiefende Darstellung der am häufigsten eingesetzten Präferenzmethoden zur Valuierung von Lebensqualitätsinstrumenten (Standard Gamble [SG], TTO, Rating Scale [z. B. VAS]) sowie innovativer Ableitungsformen (DCE). Da es in diesem Kontext weder einen „Goldstandard“ noch eine perfekte Präferenzmethode gibt, ist die Auswahl der Methode unter Berücksichtigung der jeweiligen Vor- und Nachteile zu treffen [16, 63].

Tabelle 2: Checkliste für die Planung und Durchführung einer Face-to-Face-Studie zur Valuierung eines Lebensqualitätsinstrumentes (Quelle: eigene Darstellung)

| Aspekt | Zu treffende methodische Entscheidungen |
|---------------------------------------|--|
| Art der Befragung* | <ul style="list-style-type: none"> ○ Befragtenkreis (z. B. Einzel- versus Gruppeninterviews) ○ Befragungsmedium (z. B. computergestützt versus papierbasiert) |
| Methoden-Protokoll* | <ul style="list-style-type: none"> ○ ja* (Testung im nationalen Kontext bzw. Forschungsfeld) ○ nein (ggf. Entwicklung eines eigenen Methoden-Protokolls) |
| Präferenzmethode(n)* | <ul style="list-style-type: none"> ○ Stated-Preference-Methode(n) (z. B. SG, TTO [z. B. klassisch, LT-TTO], DCE [z. B. mit oder ohne Zeitkomponente]) |
| Referenzkollektiv | <ul style="list-style-type: none"> ○ Allgemeinbevölkerung versus Subpopulation (z. B. Patienten, Angehörige) ○ Ein- und Ausschlusskriterien (z. B. Alter, Sprachkenntnisse) |
| Sampling und Rekrutierung | <ul style="list-style-type: none"> ○ Voll- versus Teilerhebung (Repräsentativität, Stichprobenumfang*) ○ bewusste Auswahl (z. B. Quotenauswahl) versus Zufallsauswahl (z. B. Random Sampling) der Teilnehmer ○ Kontaktierung und Anwerbung der Teilnehmer (z. B. öffentliche Teilnahme-Aufrufe) |
| Interviewer | <ul style="list-style-type: none"> ○ Anzahl und Auswahl der Interviewer (z. B. anhand von Vorerfahrungen, Soft Skills) ○ Schulung der Interviewer ○ Kommunikation/Feedbackgespräche mit den Interviewern |
| Valuierungsaufgabe* | <ul style="list-style-type: none"> ○ Art des Gesundheitszustandes (real/erlebt versus hypothetisch) ○ Perspektive/Referenzperson (z. B. die befragte Person selbst, ein Kind) ○ Auswahl der Gesundheitszustände (z. B. experimentelles Design) ○ Dauer der Gesundheitszustände (z. B. zehn Jahre) ○ Angaben zu dem Zustand nach dem Gesundheitszustand (z. B. Tod) ○ Anzahl und Reihenfolge der Aufgaben (z. B. Randomisierung) ○ ggf. Iteration/Suchprozedur (z. B. Mindestanzahl an iterativen Schritten) ○ ggf. Darstellung der Aufgaben (z. B. rechts-links Randomisierung bei DCE) ○ Aufwärm- und Übungsaufgaben ○ Verifizierung der Aufgaben durch den Teilnehmer (z. B. Feedback-Modul) |
| Beschreibung der Gesundheitszustände* | <ul style="list-style-type: none"> ○ visuelle Hilfen (z. B. Schiebetafel, Software) ○ Einbezug von Patientenberichten |
| Qualitätskontrolle | <ul style="list-style-type: none"> ○ Testaufgaben im Interview (z. B. dominante Aufgaben) ○ Analyse der Compliance und Performance der Interviewer (z. B. Festlegung von Qualitätskriterien, Entwicklung eines Auswertungstools)* ○ Nachbefragung der Teilnehmer bzw. der Interviewer |
| Modellierung | <ul style="list-style-type: none"> ○ Kriterien für den Ein-/Ausschluss von Interviewern, Teilnehmern oder Beobachtungen ○ Modellierung der Präferenzwerte (z.B. Regressionsanalysen) ○ Kriterien für Modellauswahl bzw. Value Set |

DCE: Discrete Choice Experiment, LT-TTO: lead-time Time-Trade-Off, TTO: Time-Trade-Off, SG: Standard Gamble; *durch das Methoden-Protokoll 1.1 bzw. 2.0 der EuroQol-Gruppe vorgegeben

Ebenso determinieren die Wahl des *Referenzkollektivs* und die *Art des zu bewertenden Gesundheitszustandes* die Lebensqualitätsabschläge [64]. Daher wird in *Modul 2* die Frage des geeigneten Personenkreises bei Valuierungsstudien erörtert. Demnach ist bei Allokationsentscheidungen auf kollektiver Ebene die Allgemeinbevölkerung und die Bewertung von hypothetischen Gesundheitszuständen heranzuziehen, u. a. weil im Gegensatz zu der Befragung von Patienten kein Coping- bzw. Adaptionsverhalten zu verzeichnen ist und hinter einem Schleier der Ungewissheit (*veil of ignorance*) bewertet wird. Folglich bleiben strategische Bewertungen

aus, da die Personen erst künftig betroffen sein könnten (*Ex-ante-Bewertung*). Leidl und Reitmeir [47, 65] haben empirisch belegt, dass die Bewertung von erlebten und hypothetischen Gesundheitszuständen zu unterschiedlichen Ergebnissen führt. Vor dem Hintergrund der aufgeführten Argumente wird deutlich, dass die Wahl des Referenzkollektivs und die Art des zu bewertenden Gesundheitszustandes die Abschlagwerte des Value Sets zur Nutzwertbestimmung beeinflussen und somit entscheidend für die Berechnung der Kosten-Effektivität sind [25, 64].

Bei der Bewertung der Gesundheitszustände ist über die jeweilige *Perspektive*, die der Befragte einnehmen soll, zu entscheiden. So analysiert *Modul 3* verschiedene Formen der Perspektive bei der Valuierung von EQ-5D-Y Gesundheitszuständen. Dabei zeigte sich, dass sich die Bewertungen für Gesundheitszustände bei Erwachsenen und Kindern unterscheiden. Die Werte waren länderübergreifend für Kinder geringer als für Erwachsene, d. h. gesundheitliche Einschränkungen implizieren einen größeren Lebensqualitätsabschlag bei Kindern. Die gewählte Perspektive bzw. die Referenzperson nimmt folglich Einfluss auf den Wert, den ein Individuum einem Gesundheitszustand zuschreibt. Bei der Planung einer Valuierungsstudie ist dies zu berücksichtigen und für den Studienteilnehmer explizit zu definieren.

Bei der Anwendung eines bestehenden Methoden-Protokolls ist dieses für den jeweiligen nationalen Kontext im Vorfeld zu testen. Aufgrund der Entwicklung des Methoden-Protokolls der EuroQol-Gruppe außerhalb von Deutschland und der Weiterentwicklung zur Version 1.1, umfasst *Modul 4* eine explorative Vorstudie zur Testung der *Eignung des Methoden-Protokolls 1.1 sowie weiterer Modifikationen* in Bezug auf die Implementierung des cTTO im deutschen Kontext. Das Methoden-Protokoll inklusive der angewendeten Präferenzmethoden – in Kombination mit intensiven Interviewer-Schulungen und einem engen Datenmonitoring – kennzeichnet sich durch eine hohe Durchführbarkeit und Akzeptanz von Teilnehmern und Interviewern in Deutschland. Im Vergleich zu der ersten Welle der EQ-5D-5L Valuierungsstudien, wurde mit dem modifizierten Methoden-Protokoll 1.1 eine hohe Datenqualität in beiden Studienarmen erzielt: ein geringerer Anteil an inkonsistenten Bewertungen, weniger Spikes in der Verteilung der cTTO-Werte, geringere Abschlagswerte für milde Gesundheitszustände sowie eine höhere Protokoll-Compliance der Interviewer. Im Hinblick auf die zusätzlich getesteten Modifikationen in der Testgruppe zeigt sich, dass (1) das Ranking die Komplexität des Interviews steigert, (2) die Trennung der cTTO-Aufgaben für Gesundheitszustände, die besser bzw. schlechter als der Tod bewertet wurden, von den Teilnehmern und Interviewern positiv eingestuft wird (z. B. „keine Unterbrechung des

Denkflusses“) und (3) die Verifizierung der cTTO-Antworten in einem Feedback-Modul das Level an Inkonsistenzen signifikant reduziert. Das internationale Forschungsprogramm bestätigt länderübergreifend das Potential des Feedback-Moduls und infolgedessen wurde diese Modifikation in das standardisierte Methoden-Protokoll 2.0 integriert (siehe *Module 5 und 6*). Da im deutschen Kontext die Trennung der cTTO-Antworten jedoch in den Nachbefragungen positiv bewertet wurde, wurde diese Modifikation für die deutsche Valuierungsstudie ebenfalls übernommen. Die Ergebnisse von *Modul 4* bestätigen die Eignung des Methoden-Protokolls in Deutschland, haben einen Beitrag zu der Entwicklung des Methoden-Protokolls 2.0 der EuroQol-Gruppe geliefert und fungieren als empirische Basis für die methodische Ausgestaltung der deutschen Valuierungsstudie.

Bei einer Face-to-Face-Studie zur Valuierung eines Lebensqualitätsinstrumentes sind weitere methodische Entscheidungen z. B. im Hinblick auf das Sampling der Studienteilnehmer unter Berücksichtigung der nationalen Gegebenheiten bzw. Vorgaben zu treffen. Die Selektion geeigneter Interviewer, deren Schulung sowie ein umfassendes Monitoring im Sinne einer Qualitätskontrolle nehmen eine zentrale Rolle im Rahmen von Valuierungsstudien ein und werden daher in Kapitel 4.4 vertiefend thematisiert. Eine Valuierungsstudie mündet in der Modellierung der Präferenzdaten, um die Abschlagswerte für das Lebensqualitätsinstrument zu schätzen. Neben der Auswahl geeigneter Regressionsverfahren sind ebenso die Entwicklung von Ein- bzw. Ausschlusskriterien für (a) Interviewer, (b) Teilnehmer und/oder (c) Beobachtungen sowie Kriterien für die Selektion des Value Sets zu definieren.

4.2 Einfluss der Präferenzmethoden auf die Lebensqualitätsabschläge

Den bisherigen Ausführungen konnte entnommen werden, dass sich die resultierenden Lebensqualitätsgewichte von verschiedenen Präferenzmethoden teilweise unterscheiden. Bei dem Methoden-Protokoll für die Valuierung des EQ-5D-5L wird eine innovative Kombination aus cTTO und DCE eingesetzt. Allerdings steht die Analyse des Einflusses der Bewertungsverfahren, insbesondere von DCE, noch aus [11]. Vor diesem Hintergrund wurde im Rahmen der Dissertation folgender Fragestellung nachgegangen: Welchen Einfluss haben die Präferenzmethoden cTTO und DCE auf die zu ermittelnden Lebensqualitätsabschläge bei dem EQ-5D-5L?

Im Rahmen von *Modul 5* wurden Abschlagswerte auf Grundlage der cTTO-Daten in einem Tobit-Modell (Modell 1), der DCE-Daten in einem Conditional Logit-Modell (Modell 2) sowie der

Kombination beider Datenformen in einem Hybrid-Modell (Modell 3a) getestet. Für die Modellierung der cTTO-Daten wurde ein zensiertes Modell gewählt, da das Methoden-Protokoll den minimalen cTTO-Wert auf -1 begrenzt. Es kann daher nicht ausgeschlossen werden, dass die Teilnehmer weitere Lebensjahre abgewogen hätten und die Werte folglich in der Spannweite $(-\infty; 1]$ anzusiedeln sind. Vor diesem Hintergrund wurde die Links-Zensierung der cTTO-Antworten bei der Modellierung in einem Tobit-Modell berücksichtigt. Da die DCE-Werte (ohne Zeitkomponente) auf einer latenten Skala ermittelt werden, können diese nicht als alleinige Datenbasis für die Generierung eines Value Sets fungieren. Um die Lebensqualitätswerte der beiden Präferenzmethoden vergleichen zu können, wurden die Koeffizienten des DCE-Modells auf eine Skala mit den Ankerpunkten 1 (volle Gesundheit) und 0 (tot) transformiert. Weitere Informationen zu der Reskalierung und der statistischen Auswertung sind *Modul 5* zu entnehmen.

Die ermittelten Abschlagswerte der cTTO- und der DCE-basierten Modelle weisen eine hohe Übereinstimmung auf. Die absolute Differenz zwischen den Abschlagswerten liegt im Durchschnitt bei 0,02. Bei der Gegenüberstellung der Abschlagswerte wird offenkundig, dass die Rangfolge der Dimensionen über die Präferenzmethoden hinweg konsistent ist (abnehmende Wichtigkeit auf Basis des Level 5): (1) Schmerzen/Körperliche Beschwerden, (2) Angst/Niedergeschlagenheit, (3) Für sich selbst sorgen, (4) Beweglichkeit/Mobilität, (5) Alltägliche Tätigkeiten. Die resultierenden cTTO- und DCE-basierten Indexwerte für den EQ-5D-5L weisen eine starke Korrelation auf ($r=0,998$; $p<0,001$). Die lineare Beziehung zwischen den Indexwerten impliziert die Kombinierbarkeit bzw. Kompatibilität beider Datenformen in einem Hybrid-Modell.

Neben dem Vergleich der cTTO- und DCE-basierten Abschlagswerte, wurde in *Modul 5* der Effekt der Hinzufügung der DCE-Werte zu den cTTO-Werten in einem Hybrid überprüft. Letzteres kombiniert die Modelle 1 und 2 in einem Hybrid-Modell (Modell 3a) durch eine Maximum-Likelihood-Verknüpfung [66, 67]. Die Differenz der absoluten Abschlagswerte zwischen dem cTTO-Modell und dem Hybrid-Modell 3a liegt durchschnittlich bei 0,01. Die aufgeführte Rangfolge der Lebensqualitätsdimensionen sowie die Korrelation der geschätzten Indexwerte bleiben konstant ($r=0,999$; $p<0,001$). Sowohl die Rangfolge der Dimensionen als auch die starke Korrelation mit den cTTO-basierten Indexwerten bleibt bei der zusätzlichen Berücksichtigung der vorliegenden Heteroskedastizität der cTTO-Daten in einem weiteren Hybrid-Modell (Modell 3b) bestehen. Allerdings vergrößert sich dadurch die Spannweite der EQ-5D-5L Indexwerte. Dies ist auf die korrigierten höheren Abschlagswerte für die größeren gesundheitlichen Einschränkungen

in Level 4 und 5 zurückzuführen. Da das Hybrid-Modell 3b, welches für Heteroskedastizität korrigiert, ebenfalls den Standardfehler der Abschlagswerte verringert (d. h. höhere Modellpräzision) und alle erhobenen Präferenzbewertungen der befragten Stichprobe umfasst, basiert das deutsche EQ-5D-5L Value Set auf diesem Modell, das im folgenden Kapitel dargestellt wird.

4.3 Value Set für die deutsche Version des EQ-5D-5L

Im Fokus dieser Dissertationsschrift liegt die Ermittlung der gesellschaftlichen Präferenzwerte für die deutsche Version des EQ-5D-5L und somit die Beantwortung folgender Leitfragestellung: Wie gewichtet die deutsche Bevölkerung die einzelnen Lebensqualitätsabschläge bei dem EQ-5D-5L?

Im Rahmen der empirischen Hauptstudie konnten die bislang fehlenden gesellschaftlichen Präferenzwerte für die gesundheitlichen Einschränkungen, die der EQ-5D-5L beschreibt, unter der erstmaligen Anwendung des Methoden-Protokolls 2.0 für den deutschen Kontext ermittelt werden (*Modul 5*). Die befragte Stichprobe ist repräsentativ für die deutsche Bevölkerung ab 18 Jahren in Bezug auf Alter, Geschlecht, Bildung und Erwerbstätigkeit (maximale proportionale Abweichung von 5,3% bei niedrig gebildeten Personen im Vergleich zu amtlichen Statistiken). Durch das Feedback-Modul¹² wurde das niedrige Level an inkonsistenten Antworten von 2,2% auf 1,5% signifikant verringert ($p < 0,001$). Bei der Analyse der cTTO-Antworten zeigt sich, dass die Teilnehmer bei milden Gesundheitszuständen stärker übereinstimmen als bei schwereren: Je schwerer die zu bewertenden Gesundheitszustände waren, desto größer wurde die Standardabweichung der cTTO-Antworten (Heteroskedastizität).

Bei dem finalen Value Set wurden alle befragten Personen ($N=1.158$) sowie alle Bewertungen auf Grundlage der cTTO- und DCE-Aufgaben in einem Hybrid eingeschlossen, bei welchem die Haupteffekte durch 20 Dummy-Variablen abgebildet und die Zensierung sowie die Heteroskedastizität der cTTO-Werte berücksichtigt werden (siehe Kapitel 4.2). Jedem Antwortlevel der fünf Dimensionen lässt sich ein signifikanter Abschlagswert zuordnen ($p < 0,001$). Die Abschlagswerte sind durchweg konsistent, d. h. je höher die gesundheitliche Einschränkung, desto größer ist der jeweilige Abschlagswert von der Lebensqualität. Mithilfe des Value Sets kann jedem Gesundheitszustand, der durch den EQ-5D-5L beschrieben wird, ein entsprechender

¹² Die Implementierung des Feedback-Moduls in der EQ-VT ist *Modul 6* zu entnehmen.

Nutzwert zugeordnet werden. Die vorhergesagten Werte reichen dabei von -0,661 für den schlechtesten (55555) bis 1 für den besten möglichen Gesundheitszustand (11111).

Bei dem Vergleich der Abschlagswerte der einzelnen Dimensionen wird offenkundig, dass sich die Präferenzen im Hinblick auf die Dimensionen unterscheiden und somit von der befragten Stichprobe unterschiedlich gewichtet werden. Die Abschlagswerte der funktionalen Dimensionen (Beweglichkeit/Mobilität, Für sich selbst sorgen und Alltägliche Tätigkeiten) sind auf einem vergleichbaren Niveau. Die Lebensqualitätsabschlagswerte für die symptomatischen Dimensionen (Schmerzen/Körperliche Beschwerden und Angst/Niedergeschlagenheit) sind hingegen, insbesondere für die Level 4 und 5, höher. Die ermittelten gesellschaftlichen Präferenzwerte geben den Symptomen im Bereich Schmerzen/Körperliche Beschwerden das höchste Gewicht. Ebenso bei der Gegenüberstellung der relativen Wichtigkeit der Dimensionen bestätigt sich die in Kapitel 4.2 aufgeführte Rangfolge der Dimensionen. Aus Sicht der Stichprobe wird ca. ein Drittel der Lebensqualität durch gesundheitliche Einschränkungen im Bereich Schmerzen/Körperliche Beschwerden (36%) beeinflusst. Den Symptomen im Bereich Angst/Niedergeschlagenheit kommen mit 21,66% die zweithöchste relative Wichtigkeit zu. Die funktionalen Dimensionen liegen in Bezug auf die relative Wichtigkeit nah beieinander und weisen einen relativen Anteil von 12,85% (Alltägliche Tätigkeiten) bis 16,34% (Für sich selbst sorgen) auf.

4.4 Einsatz des Methoden-Protokolls in künftigen Valuierungsstudien

Auf Grundlage der empirischen Erfahrungen mit dem Methoden-Protokoll, soll im Rahmen dieser Dissertationsschrift des Weiteren folgende Fragestellung beantwortet werden: Wie sollte das Methoden-Protokoll der EuroQol-Gruppe unter Berücksichtigung der verfügbaren Evidenz bestmöglich eingesetzt werden?

Sowohl in der explorativen Vorstudie zur Testung des Methoden-Protokolls (*Modul 4*) als auch in der späteren Studie zur Valuierung des deutschen EQ-5D-5L (*Modul 5*) kristallisierte sich heraus, dass insbesondere die in Tabelle 3 dargestellten fünf Aspekte bei dem künftigen Einsatz des Methoden-Protokolls zu beachten sind (*Lessons learned*). Die fünf Aspekte tangieren dabei einen erforderlichen Lernprozess auf folgenden drei Ebenen: (1) Ebene des Forscherteams bzw. der Studienplanung, (2) Ebene des Interviewers sowie (3) Ebene des Teilnehmers.

Den Ausführungen zu Kapitel 4.1 konnte entnommen werden, dass die Testung eines bestehenden Methoden-Protokolls im nationalen Kontext erfolgen sollte, um die Eignung des Protokolls, dessen Präferenzmethoden und die Art der Implementierung im Hinblick auf die Akzeptanz des Referenzkollektivs sowie der Interviewer zu überprüfen. Neben der Sicherstellung der Eignung des Methoden-Protokolls, sollte eine umfassende Pilotstudie Bestandteil künftiger Valuierungsstudien sein, da dies einen *Lernprozess bei dem Forscherteam* bei der Planung und Durchführung solch einer Studie ermöglicht. Zudem können die Ergebnisse als empirische Basis der Hauptstudie und die generierten Präferenzdaten als externe Validierung des Value Sets fungieren.

Tabelle 3: Lessons learned bei der Anwendung des Methoden-Protokolls der EuroQol-Gruppe
(Quelle: eigene Darstellung)

| Ebene | Lessons learned |
|---------------------------------|---|
| Forscherteam/ Studienplanung | 1) Eine umfassende Testung des Methoden-Protokolls inklusive der Methodenkombination aus cTTO und DCE sollte im nationalen Kontext erfolgen. |
| Interviewer | 2) Der Interviewer nimmt eine Schlüsselrolle in der Datenerhebung ein, da die Protokoll-Compliance und das Engagement des Interviewers die Präzision der cTTO-Bewertungen bestimmt. |
| | 3) Eine kontinuierliche Qualitätskontrolle im Sinne einer formativen Evaluation ist unerlässlich. |
| Teilnehmer | 4) Übungsaufgaben sind für die späteren cTTO-Bewertungen erforderlich. |
| | 5) Die Verifizierung der cTTO-Bewertungen durch den Teilnehmer kann das Konsistenzniveau der Präferenzwerte erhöhen. |

cTTO: composite Time-Trade-Off, DCE: Discrete Choice Experiment

Das *Modul 6* gibt einen Überblick zu den seit 2012 gewonnenen Erkenntnissen zur Valuierung des EQ-5D-5L mithilfe des Methoden-Protokolls, wobei sowohl die Erkenntnisse der explorativen Vorstudie 2 (*Modul 4*) als auch der deutschen Valuierungsstudie (*Modul 5*) darin einfließen. So wird dargestellt, wie die Methoden für die Valuierung des EQ-5D-5L weiterentwickelt und das Methoden-Protokoll entsprechend angepasst wurde. Bei den bisherigen EQ-5D-5L Valuierungsstudien konnte neue Evidenz generiert werden, dass kein Methoden-Protokoll oder keine (teilweise automatisierte) Software per se ausreicht, um eine gute Datenqualität sicherzustellen. Dem Interviewer kommt eine zentrale Rolle, insbesondere bei den cTTO-Aufgaben, zu. Nur wenn eine hohe Protokoll-Compliance und eine gute Interviewer-Performance sichergestellt sind, kann das Protokoll wie intendiert umgesetzt und können valide Interviews durchgeführt werden. Neben einer reflektierten Auswahl der Interviewer im Hinblick auf methodische und inhaltliche Vorerfahrungen sowie den für ein Valuierungsinterview notwendigen *Soft Skills* (z. B. Geduld) (siehe Kapitel 4.1), ist der *Lernprozess des Interviewers* zu gestalten.

Intensive Schulungen und eine kontinuierliche Betreuung der Interviewer während der Datenerhebung sind unumgänglich. Die Erkenntnis, dass dem Interviewer eine Schlüsselrolle zukommt, motivierte die Entwicklung eines Prozesses zur Qualitätskontrolle [60]. Es wurden dabei sowohl die Präferenzwerte als auch die Meta-Daten in der Software genutzt (u. a. Pfad der iterativen Schritte zu dem Indifferenzpunkt sowie Zeitstempel). Es wurde ersichtlich, dass sowohl das Engagement der Interviewer als auch das der Studienteilnehmer in Kombination mit der adaptiven Eigenschaft der cTTO-Daten¹³ die Präzision der ermittelten Präferenzwerte terminieren. Das Auftreten von Datenmustern, wie Spikes oder eine selektive Nutzung der cTTO-Skala, variiert zwischen Interviewern (Interviewer-Effekte). Folglich kann festgehalten werden, dass das Interviewer-Verhalten die Wahrscheinlichkeit des Auftretens solcher Aufgabenverkürzungen beeinflusst. Es wurden bei der Synthese der Erkenntnisse aller EQ-5D-5L Valuierungsstudien verschiedene Phänomene herausgearbeitet, die die Eigenschaften von cTTO-Daten charakterisieren. Diese können in drei Gruppen zusammengefasst werden: (1) Phänomene in Bezug auf den Mechanismus der cTTO-Aufgabe, (2) Phänomene in Bezug auf das individuelle Antwortverhalten des Teilnehmers und (3) Eigenschaften des gesamten cTTO-Datensatzes. Eine ausführliche Darstellung der identifizierten Phänomene der cTTO-Daten, deren Konsequenzen sowie mögliche Modellierungsansätze als Lösungsstrategien sind *Modul 6* zu entnehmen.

Der entwickelte Prozess der Qualitätskontrolle wurde u. a. in der in *Modul 4* dargestellten Vorstudie empirisch getestet. Damit kann die Compliance und Performance des Interviewers im Sinne einer formativen Evaluation überprüft werden. Zudem können die aufgeführten Phänomene der cTTO-Daten identifiziert und diesen teilweise entgegengewirkt werden. So können mögliche Interviewer-Effekte reduziert werden. Die Evidenz der bisherigen Valuierungsstudien zeigt, dass ein mangelndes Engagement bei dem Interviewer bzw. dem Teilnehmer bei der Aufgabenbearbeitung zu nachteiligen Effekten auf die Datenqualität führen kann. Im Vergleich zu den EQ-5D-5L Valuierungsstudien ohne Qualitätskontrolle (Methoden-Protokoll 1.0) zeigt sich, dass die Studien mit Qualitätskontrolle (seit Methoden-Protokoll 1.1) eine gesteigerte Protokoll-Compliance sowie eine höhere Datenqualität im Sinne von weniger Spikes, Inkonsistenzen und Interviewer-Effekten aufweisen (*Module 4 und 5*). Neben der Entwicklung des Prozesses zur Qualitätskontrolle, wurden innovative Modellierungstechniken generiert, um die Phänomene der

¹³ Dem *Modul 6* ist eine graphische Übersicht der iterativen Sequenz der cTTO-Aufgaben in der EQ-VT zu entnehmen.

cTTO-Daten zu berücksichtigen und deren Einfluss auf die Schätzung des EQ-5D-5L Value Sets zu reduzieren (z. B. Tobit-Modell oder heteroskedastische Modelle) (*Modul 6*).

Des Weiteren ist der *Lernprozess des Teilnehmers* zu unterstützen, da dieser teilweise erstmalig mit den Präferenzmethoden und der Bewertung von Gesundheit konfrontiert wird. Gegebenenfalls bildet sich der Teilnehmer erst während des Interviews seine Meinung zu gesundheitlichen Einschränkungen [68]. Folglich benötigt der Teilnehmer neben der Hilfestellung durch den Interviewer ein gewisses Maß an Übung, um insbesondere das Konzept der cTTO-Aufgaben, die Funktionsweise der Software und die Beschreibung der Gesundheitszustände zu erfassen. Diese Aspekte bilden die Grundvoraussetzung für die Fähigkeit, präzise Präferenzurteile zu geben. Die Einführung von Übungszuständen seit dem Methoden-Protokoll 1.1 gibt dem Teilnehmer die Möglichkeit, die cTTO-Aufgaben zu üben und einen Einblick in verschiedene Schweregrade der zu bewertenden Gesundheitszustände zu erhalten. Auch wenn ein Teilnehmer das Konzept nach den Übungsaufgaben verinnerlicht hat, können dennoch Lerneffekte bei der Bearbeitung der Aufgaben auftreten. Daher ermöglicht die Hinzufügung des Feedback-Moduls seit dem Methoden-Protokoll 2.0 die Verifizierung der cTTO-Bewertungen als interne Validierung durch den Teilnehmer selbst. Die vorliegende Evidenz zeigt, dass das Level an inkonsistenten Antworten im Vergleich zu den Studien ohne Übungsaufgaben reduziert werden konnte (seit Version 1.1; *Module 4 und 5*). Die zusätzliche Möglichkeit zur Verifizierung der cTTO-Antworten in einem Feedback-Modul (seit Version 2.0; *Modul 5*) erzielt eine weitere Reduktion der Inkonsistenzen. Im Gegensatz zu den cTTO-Aufgaben zeigte sich im Hinblick auf die DCE-Aufgaben eine hohe Praktikabilität in Bezug auf das Verständnis und die Anwendung bei den Teilnehmern seit der ersten Version des Methoden-Protokolls, sodass keine Modifikationen zur Gestaltung des Lernprozesses bei dem Teilnehmer in der EQ-VT notwendig waren.

Die gewonnenen Erkenntnisse der bislang durchgeführten EQ-5D-5L Valuierungsstudien resultierten in der Weiterentwicklung des Methoden-Protokolls inklusive einer Qualitätskontrolle und innovativen Modellierungsansätzen. Mithilfe der Anwendung des Methoden-Protokolls 2.0 konnte ein konsistentes Value Set für die deutsche Version des EQ-5D-5L geschätzt werden (*Modul 5*). Demnach ist das verbesserte Methoden-Protokoll 2.0 generell als geeignet einzustufen und kann auf Grundlage der vorliegenden Evidenz für die Ermittlung weiterer nationaler Tarife eingesetzt werden. Dabei sind die aufgeführten Lernprozesse des Forscherteams, der Interviewer sowie der Teilnehmer zu berücksichtigen und aktiv zu gestalten.

5. Diskussion

5.1 Ergebnisorientierte Diskussion

Für die Aggregation der Antworten des deskriptiven Systems des deutschen EQ-5D-5L zu einem singulären Index konnte bisher nur auf ein Crosswalk-Value Set zurückgegriffen werden [48]. Im Rahmen dieser Dissertation konnten die gesellschaftlichen Präferenzwerte für den EQ-5D-5L für den deutschen Kontext ermittelt werden. Die gesundheitlichen Einschränkungen der einzelnen Dimensionen wurden dabei von der befragten Stichprobe unterschiedlich gewichtet. Das *höchste relative Gewicht wurde auf die symptomatischen Dimensionen* Schmerzen/Körperliche Beschwerden und Angst/Niedergeschlagenheit gelegt, wohingegen Einschränkungen im Bereich der funktionalen Dimensionen eine geringere Wichtigkeit in Bezug auf die Lebensqualität zugeschrieben wurde. Zum einen wird dem jeweiligen sozialen und individuellen Wertesystem ein wesentlicher Einfluss auf die Bewertung von Gesundheitszuständen zugesprochen [4]. Zum anderen könnte die Gewichtung auf das umfassende soziale Absicherungssystem der Kranken- und Pflegeversicherung in Deutschland zurückgeführt werden. Im Falle von funktionellen Einschränkungen, wie z. B. bei körperbezogenen Pflegemaßnahmen, kann auf diverse ambulante Pflegedienste zurückgegriffen werden [69], die von der Pflegeversicherung bezuschusst bzw. übernommen werden (§ 4 SGB XI) [70]. Ebenso werden die Kosten für medizinisch erforderliche Hilfsmittel bei einer Gehbehinderung von der Krankenversicherung erstattet (§ 33 SGB V) [71].

Die Anwendung des standardisierten Methoden-Protokolls ermöglicht die Vergleichbarkeit zu *weiteren länderspezifischen EQ-5D-5L Value Sets* sowie die Analyse kultureller Einflüsse auf die Bewertung von Gesundheit [16]. Die bisher publizierten europäischen Value Sets aus England [72], Spanien [73] und den Niederlanden [74] kommen zu einer ähnlichen Rangfolge der symptomatischen und funktionalen Dimensionen wie das deutsche Value Set. Es sind allerdings Unterschiede bei der Gewichtung der einzelnen Dimensionen zu verzeichnen. So kommen z. B. gemäß der Präferenz der niederländischen Bevölkerung Symptomen im Bereich Angst/Niedergeschlagenheit das höchste Gewicht zu [74]. Bei den außereuropäischen EQ-5D-5L Value Sets sind hingegen größere Unterschiede festzustellen. So wird z. B. in Indonesien [75], Uruguay [76] und Südkorea [77] der Dimension Mobilität/Beweglichkeit das höchste Gewicht zugesprochen. Im Gegensatz zu dem deutschen Value Set nehmen im Falle von Indonesien Schmerzen/Körperliche Beschwerden nicht den höchsten sondern den geringsten Einfluss auf die Lebensqualität ein. Purba et al. [75] führen die Unterschiede auf die bessere Infrastruktur und

weniger Handarbeit, für die eine Funktionsfähigkeit erforderlich ist, in Europa zurück. Neben dem sozialen und individuellen Wertesystem zeigt die vorliegende Evidenz, dass die soziale Absicherung, eine behindertengerechte Umgebung sowie die Arbeitsbedingungen des jeweiligen nationalen Kontextes die Bewertung von Gesundheit bestimmen können. Dies gilt es in weiteren Studien zu erforschen. In Anbetracht der Länderunterschiede in Bezug auf die Präferenzwerte wird die Relevanz von länderspezifischen Value Sets deutlich und eine ungeprüfte Übernahme der Value Sets aus anderen Kultur- und Sprachräumen gilt es zu vermeiden [4, 12, 78].

Bei der Gegenüberstellung des deutschen EQ-5D-5L Value Sets mit dem *deutschen TTO-basierten Value Set für den EQ-5D-3L* [46] wird offenkundig, dass insbesondere Symptome im Bereich Angst/Niedergeschlagenheit an Bedeutung gewonnen haben. Dies könnte auf die Zunahme der Diagnosehäufigkeit psychischer Erkrankungen und einer damit gesteigerten Sensibilisierung für deren Krankheitslast in Deutschland zurückzuführen sein [3, 79, 80]. Der höchste Abschlag des EQ-5D-3L Value Sets ist hingegen für Beweglichkeit/Mobilität zu verzeichnen, wohingegen diese Dimension den vierten Rangplatz in dem EQ-5D-5L Value Set einnimmt. Dies ist u. a. auf die Änderung der Formulierung des höchsten Antwortlevels zurückzuführen: „ans Bett gebunden“ (EQ-5D-3L) versus „nicht in der Lage herumzugehen“ (EQ-5D-5L). Zudem sei angemerkt, dass das deutsche EQ-5D-3L Value Set vor mehr als 15 Jahren und mit einem anderen Studienprotokoll ermittelt wurde. Es kann daher nicht abschließend beantwortet werden, ob sich die Präferenzen in Bezug auf gesundheitliche Einschränkungen verändert haben oder ob die Abweichungen der relativen Gewichtung der Dimensionen auf die methodischen Unterschiede zurückzuführen sind.

Die deutsche EQ-5D-5L Valuierungsstudie erzielte im Vergleich zu den ersten Länderstudien, die das Methoden-Protokoll 1.0 verwendeten, eine hohe Datenqualität [60, 72–74]: ein geringerer Anteil an inkonsistenten Bewertungen, niedrige Abschlagswerte für milde Gesundheitszustände, kaum Spikes, weniger Protokoll-Abweichungen sowie geringere Interviewer-Effekte. So konnte in Deutschland jeder gesundheitlichen Einschränkung des deskriptiven Systems ein Abschlagswert zugeordnet und somit zwischen den marginalen Abstufungen der Antwortlevel differenziert werden. Dies stellt ein wesentliches Unterscheidungsmerkmal zu dem TTO-basierten deutschen EQ-5D-3L Value Set dar, da bei diesem für vier gesundheitliche Einschränkungen keine signifikanten Abschlagswerte ermittelt werden konnten [46]. Das EQ-5D-3L Value Set ist daher oftmals Gegenstand kritischer Diskussionen in der Literatur [79]. Dies verdeutlicht das Potential des EQ-5D-5L Value Sets im Hinblick auf eine bessere Diskriminierbarkeit der resultierenden

Nutzwerte. Hinzu kommt, dass die EQ-5D-5L Valuierungsstudie auf einem evidenzbasierten Methoden-Protokoll inklusive einer Qualitätskontrolle basiert, welches in der Vergangenheit nicht existierte. Zudem wurde im Gegensatz zur Vorgängerstudie eine deutschlandweit repräsentative Stichprobe befragt. Es sei jedoch einschränkend erwähnt, dass die Rekrutierung in sechs Regionen erfolgte und dass ein geringer Mittelschichtbias zu verzeichnen war. Dennoch ist die befragte Stichprobe insgesamt als repräsentativ für die Allgemeinbevölkerung in Deutschland einzustufen. Bei dem Vergleich der resultierenden Indexwerte auf Basis der beiden deutschen Value Sets wird ersichtlich, dass das EQ-5D-5L Value Set eine größere Spannweite umfasst als das EQ-5D-3L Value Set. Die Verteilung der EQ-5D-5L Indexwerte folgt einer linksschiefen Normalverteilung, wohingegen die EQ-5D-3L Indexwerte durch eine bi-modale Verteilung mit kaum Differenzierung für milde Gesundheitszustände gekennzeichnet ist. Dem EQ-5D-5L Value Set ist folglich ein größerer Evaluationsbereich mit einer breiteren Verteilung der resultierenden Nutzwerte zuzuordnen. Demnach besitzt das EQ-5D-5L Value Set das Potential, besser zwischen Patienten, insbesondere bei milden gesundheitlichen Einschränkungen, diskriminieren zu können. Dies könnte die Anwendbarkeit des EQ-5D im Allgemeinen und in gewissen Indikationsgebieten in Deutschland steigern. Der praktisch relevante Evaluationsbereich ist jedoch in künftigen Forschungsstudien zu überprüfen. Es bleibt festzuhalten, dass das entwickelte EQ-5D-5L Value Set eine robuste und aktuelle Basis darstellt, um präferenzbasierte Indexwerte für den deutschen EQ-5D-5L zu aggregieren. Folglich sollte es das bevorzugte Value Set für Deutschland sein.

Die Anwendung des ermittelten Value Sets sollte dabei (vorerst) ausschließlich für die indirekte Nutzwertbestimmung mithilfe des EQ-5D-5L erfolgen. So zeigte sich, dass sich die gesellschaftlichen Präferenzwerte für gesundheitliche Einschränkungen bei Erwachsenen und Kindern unterscheiden. Gesundheitlichen Problemen im Kindesalter werden höhere Abschläge von der Lebensqualität und somit ein anderer Wert zugeschrieben. Die Ursache dieser Bewertungsunterschiede sowie die Generalisierbarkeit der Ergebnisse bei der Anwendung anderer Methoden als die VAS sollten in weiteren Forschungsstudien untersucht werden. Das EQ-5D-5L Value Set kann folglich nicht ohne weitere Überprüfung auf die kürzlich entwickelten Gesundheitszustände der *fünf-Level-Version des EQ-5D-Y* (EQ-5D-Y-5L) [81] übertragen werden. Hinzu kommt, dass Unterschiede bei der Formulierung des deskriptiven Systems zwischen der Kinder- und der Erwachsenenversion vorhanden sind. Dies stellt eine weitere Limitation in Bezug auf die Übertragbarkeit der Abschlagswerte des EQ-5D-5L Value Sets auf den EQ-5D-Y-5L dar. Die Valuierung des EQ-5D-Y-5L sollte Gegenstand künftiger Forschungsbestrebungen sein.

5.2 Methodenorientierte Diskussion

Bei dieser Dissertation wurde ein sukzessiver Aufbau gewählt, sodass mithilfe von explorativen Vorstudien eine empirische Basis für die Ermittlung der gesellschaftlichen Präferenzwerte für den EQ-5D-5L gelegt wurde. Durch die gewählte Struktur wurde ein evidenzbasierter Lernprozess im Hinblick auf die Planung und Durchführung der späteren Valuierungsstudie ermöglicht und die Eignung des Methoden-Protokolls im deutschen Kontext sichergestellt. So konnte im Vergleich zu den ersten Länderstudien, bei denen der Einsatz des Methoden-Protokolls 1.0 ohne umfassende Testung der Eignung im nationalen Kontext erfolgte, eine *hohe Datenqualität* erzielt werden [60, 72–74]. Zudem ist die hohe Datenqualität darauf zurückzuführen, dass die deutsche Valuierungsstudie das verbesserte Methoden-Protokoll 2.0 eingesetzt hat. Im Gegensatz zu den ersten Länderstudien kennzeichnet sich die deutsche Studie durch geringe Protokoll-Abweichungen und kaum Interviewer-Effekte. Die dargestellte Evidenz der bisherigen Valuierungsstudien zeigt, dass ein mangelndes Engagement des Interviewers bzw. des Teilnehmers bei der Aufgabenbearbeitung zu nachteiligen Effekten auf die Datenqualität führen kann. Da es einen Zusammenhang zwischen der Compliance bzw. Performance der Interviewer und dem Auftreten einiger der in *Modul 6 identifizierten Phänomene der cTTO-Daten* gibt, waren im Falle der deutschen Valuierungsstudie keine umfassenden Korrekturmaßnahmen durch die in *Modul 6* vorgestellten innovativen Modellierungstechniken erforderlich. Dem Auftreten der Phänomene – insbesondere in Bezug auf das individuelle Antwortverhalten des Teilnehmers – wurde durch eine kontinuierliche Qualitätskontrolle entgegengewirkt. So wurden z. B. ausschließlich in 3% der Interviews das LT-TTO durch den Interviewer nicht erklärt (*Modul 5*). Aufgrund des als gering einzustufenden Levels an Protokoll-Abweichungen waren weder eine Korrektur noch ein Ausschluss von einzelnen Teilnehmern bei der Modellierung erforderlich. Bei der spanischen EQ-5D-5L Valuierungsstudie hingegen fielen die Protokoll-Verletzungen in Bezug auf das LT-TTO höher aus (76,1%). Um der Protokollverletzung Rechnung zu tragen, wurde sich in dieser Studie für eine Zensierung der cTTO-Werte bei 0 entschieden [73]. Im Vergleich zu den ersten Länderstudien zeigte sich, dass in der deutschen Valuierungsstudie kaum Spikes bei der Verteilung der cTTO-Werte vorhanden waren und dass die Teilnehmer zwischen den Gesundheitszuständen differenziert haben. Es kann folglich davon ausgegangen werden, dass keine Verkürzung der Aufgabenbearbeitung zu korrigieren ist, im Gegensatz zu Ländern der ersten Welle, bei denen überwiegend cTTO-Werte bei den Datenpunkten -1 | -0,5 | 0 | 0,5 | 1 ermittelt wurden. Diese Spikes können u. a. auf ein mögliches *Satisficing* zurückgeführt werden, bei welchem der Teilnehmer

seinen Indifferenzpunkt nach wenigen iterativen Schritten festlegt und sich nicht weiter bemüht, einen präzisen Präferenzwert zu geben (*Phänomen in Bezug auf das individuelle Antwortverhalten des Teilnehmers*). In Spanien wurde dieses Phänomen durch eine Intervallregression korrigiert [73]. Bei der Modellierung des deutschen EQ-5D-5L Value Sets wurde durch die Wahl eines Tobit-Modells berücksichtigt, dass eine Links-Zensierung per cTTO-Design bei -1 vorgenommen wurde (*Phänomen in Bezug auf den Mechanismus der cTTO-Aufgabe*). Ebenso erfolgte dies z. B. in den Niederlanden [74] und in Indonesien [75]. Bei dem deutschen cTTO-Gesamtdatensatz zeigte sich, dass die Standardabweichung der cTTO-Bewertungen mit zunehmendem Schweregrad der zu bewertenden Gesundheitszustände steigt (*Eigenschaft des gesamten cTTO-Datensatzes*). Dies ist insbesondere auf unterschiedliche *Bottom*-Werte der Teilnehmer zurückzuführen: Bei zwei Drittel der Befragten lag dieser im negativen Bereich, wohingegen ein Drittel der Personen keinen Gesundheitszustand schlechter als der Tod bewertete. Die resultierende Heteroskedastizität der cTTO-Daten wurde in der Modellierung durch die Schätzung der Varianz für jeden Parameter berücksichtigt, um eine Verzerrung der resultierenden Abschlagswerte zu vermeiden [67]. Dieser Ansatz wurde ebenfalls bei der Modellierung des cTTO-Datensatzes in Indonesien [75] angewendet.

In dem deutschen EQ-5D-5L Value Set wurde sich dafür entschieden, neben den cTTO-Daten auch die DCE-Daten in einen *Hybrid* einfließen zu lassen. *Modul 5* ist eine umfassende Erörterung zu entnehmen, warum ein Hybrid-Modell einem rein TTO-basierten Modell theoretisch vorzuziehen ist und welche Limitationen die Präferenzmethoden cTTO bzw. DCE aufweisen. Ein Hauptargument für die Nutzung eines Hybrid-Modells liegt in der Annahme, dass der „wahre“ Wert nicht von einer singulären Präferenztechnik ermittelt werden kann, sondern dass beide Methoden als komplementär einzustufen sind [66]. Bei dem Vergleich der cTTO-, DCE- und Hybrid-Modelle in der Valuierungsstudie für Deutschland zeigt sich, dass die Abschlagswerte, die Rangfolge sowie die relative Wichtigkeit der Dimensionen über die verschiedenen Modelle hinweg konstant sind. Bei dem Hybrid-Modell ist jedoch die höchste Modellgenauigkeit zu verzeichnen. Vor diesem Hintergrund berücksichtigt das deutsche EQ-5D-5L Value Set die komplementären cTTO- und DCE-Daten in einem Hybrid-Modell, wobei die aufgeführten innovativen Modellierungstechniken angewendet wurden, die die Links-Zensierung der cTTO-Daten und die Heteroskedastizität der Standardfehler berücksichtigen, um eine Verzerrung der geschätzten Parameter zu vermeiden. Die Evidenz der vorliegenden Valuierungsstudie spricht folglich für die *Kombinierbarkeit der cTTO- und DCE-Daten* in einem Hybrid-Modell. Auch wenn nicht

zwangsläufig dieselben Ergebnisse zu erwarten sind, zeigten sich eine hohe Übereinstimmung der geschätzten Abschlagswerte sowie eine starke positive Korrelation der resultierenden Indexwerte für den EQ-5D-5L. Dies impliziert die Kombinierbarkeit beider Datenformen in einem Hybrid-Modell für den deutschen Kontext. Im Hinblick auf die Generalisierung dieser Erkenntnisse sei jedoch einschränkend angemerkt, dass die Kompatibilität der Datenformen im jeweiligen Kontext zu überprüfen ist, bevor eine finale Modellauswahl erfolgt.

Bei dem Vergleich der verfügbaren EQ-5D-5L Value Sets zeigt sich, dass ebenso in Spanien [73], England [72] und Indonesien [75] ein Hybrid-Ansatz für das nationale Value Set gewählt wurde. Andere Länder hingegen, wie z. B. die Niederlande [74] und Uruguay [76], berücksichtigen ausschließlich die cTTO-Daten für die Schätzung des Value Sets. Die einbezogene Datenbasis beruht auf normativen Überlegungen für den spezifischen Kontext aber auch auf datenbasierten Gründen zugunsten der Schätzung eines konsistenten Value Sets. Im Falle der ersten Länderstudien könnte ebenfalls das Fehlen geeigneter Modellierungstechniken die Entscheidung für ein rein cTTO-basiertes Value Set begünstigt haben [74, 76]. Durch die Entwicklung neuer Modellierungsansätze, die beide Datenformen kombinieren und die Zensierung, Trunkierung und Heteroskedastizität der cTTO-Daten berücksichtigen können, wurde der damaligen Forschungslücke begegnet. Jene Modellierungsverfahren sind derzeit als *Best-Practice*-Methoden einzustufen [67, 82]. Auch wenn die Modellierungsansätze einen wesentlichen Fortschritt bei der Modellierung von cTTO-Daten sowie der Einbeziehung zweier Präferenzmethoden in einem Hybrid darstellen, besteht Weiterentwicklungsbedarf in Bezug auf das Hybrid-Modell. In dem jetzigen Ansatz wird die Korrelation der cTTO- und DCE-Werte bei einer Person (*within-person correlation*) nicht berücksichtigt [66]. Dies sollte Gegenstand weiterer Forschungsbestrebungen sein. Ebenso wird diskutiert, inwiefern die Annahme einer konstanten Zeitpräferenz bei cTTO-Bewertungen zu halten ist [53, 83, 84]. In künftigen Studien sollten das Vorliegen von Zeitpräferenzen weiter erforscht werden und je nach Evidenzlage geeignete Methoden zur Berücksichtigung von Zeitpräferenzen (Diskontierung) bei der Schätzung von cTTO-basierten Value Sets entwickelt werden.

Die vorliegende Dissertationsschrift liefert erstmalig Evidenz, dass mit dem Methoden-Protokoll 2.0 kombiniert mit einer innovativen Qualitätskontrolle ein konsistentes nationales EQ-5D-5L Value Set geschätzt werden kann. Jedem Antwortlevel der fünf Dimensionen konnte ein signifikanter Abschlagswert zugeordnet werden. Folglich konnte der Verbesserung der Sensitivität

und Diskriminierbarkeit als intendierte Wirkung des erweiterten EQ-5D-5L ebenso mit dem dazugehörigen Value Set in Deutschland Rechnung getragen werden. Auch wenn das Methoden-Protokoll 2.0 für die Schätzung nationaler Value Sets als geeignet einzustufen ist, sind weitere Forschungsbestrebungen zur Valuierung des EQ-5D-5L seitens der EuroQol-Gruppe zu erwähnen. So werden derzeit neue Designs im Hinblick auf forschungsökonomische Potentiale für kleinere Länderstudien mit weniger Ressourcen geprüft. Dabei wird die Anzahl der DCE-Aufgaben gegenüber der kostenintensiveren cTTO-Aufgaben erhöht (*EQ-VT light-Ansatz*). Zudem werden neue Präferenzmethoden mit DCE als alleiniger Datenbasis getestet, wobei eine Zeitkomponente als weiteres Attribut enthalten ist (*DCE duration-Ansatz*). Weitere Ausführungen zu den künftigen Forschungsaktivitäten der EuroQol-Gruppe sind *Modul 6* zu entnehmen.

Abschließend sei auf einige *Limitationen dieser Dissertationsschrift* hingewiesen. Bei der explorativen Vorstudie 2 wurden neben der Testung des Methoden-Protokolls im deutschen Kontext drei Modifikationen erforscht (*Modul 4*). Durch die simultane Testung der Modifikationen in einer Software-Version konnte der Effekt der einzelnen Modifikationen nur eingeschränkt quantifiziert werden. Aus diesem Grund wurden zum einen die Ergebnisse der Nachbefragung und zum anderen die Erkenntnisse weiterer Studien des Forschungsprogramms hinzugezogen, um über die methodische Ausgestaltung der deutschen EQ-5D-5L Valuierungsstudie und die Weiterentwicklung zu dem Methoden-Protokoll 2.0 zu entscheiden. In künftigen explorativen Studien sollte die Testung einer singulären Modifikation pro Gruppe erfolgen.

Im Hinblick auf die Vergleichbarkeit der Präferenzwerte der drei Studien sei angemerkt, dass die empirische Vorstudie 1 (*Modul 3*) auf einer schriftlichen Befragung unter Anwendung von VAS basiert, wohingegen die empirische Vorstudie 2 (*Modul 4*) und die Hauptstudie (*Modul 5*) ein Computer Assisted Personal Interviewing (CAPI)-Ansatz sowie cTTO und DCE als Präferenzmethoden anwenden. Folglich können die Präferenzwerte der Vorstudie 1 aufgrund der unterschiedlichen Präferenzmethoden sowie der Art der Erhebung nicht uneingeschränkt mit den anderen beiden Studien verglichen werden. Zudem sei angemerkt, dass die Teilnehmer der beiden Vorstudien nicht repräsentativ für die deutsche Allgemeinbevölkerung sind, da dies für den explorativen Forschungszweck nicht erforderlich war. Aus diesem Grund wurde der Datensatz der Vorstudie 2 nicht als externe Validierung für die Modelle der Hauptstudie zur Valuierung des deutschen EQ-5D-5L mithilfe einer repräsentativen Stichprobe der Allgemeinbevölkerung hinzugezogen. Des Weiteren sei angemerkt, dass auf den Vergleich zur Valuierung anderer

Lebensqualitätsinstrumente verzichtet wird. Bis dato sind keine generischen Indexinstrumente außer dem EQ-5D-5L bekannt, deren Abschlagswerte durch eine Kombination aus cTTO und DCE ohne Zeitkomponente ermittelt wurden. Da die Vergleichbarkeit und Reliabilität der mit unterschiedlichen Präferenzmethoden erzielten Ergebnisse als problematisch einzustufen ist [4], erfolgt an dieser Stelle kein Vergleich zu bestehenden Value Sets anderer Messinstrumente.

Der *Effekt der Selbstselektion* kann in den Studien (*Module 3 bis 5*) nicht ausgeschlossen werden. So könnten insbesondere Menschen mit einem höheren persönlichen thematischen Interesse für die Bewertung von Gesundheit teilgenommen haben [68, 85]. Ferner sei aufgeführt, dass bei der Teilnehmerrekrutierung die Erfahrung mit einer Erkrankung nicht berücksichtigt und das Vorhandensein von Erkrankungen nicht erfragt wurde. Auch wenn davon auszugehen ist, dass bei einer Stichprobe der Allgemeinbevölkerung diverse Krankheitsbilder inkludiert sind, kann nicht abschließend beantwortet werden, in welchem Ausmaß eigenes Coping- bzw. Adaptionsverhalten die Bewertung der Gesundheitszustände beeinflusst hat. In künftigen Präferenzbefragungen sollte das Vorliegen von (diagnostizierten) Erkrankungen daher erfasst werden.

Durch die Anwesenheit eines Interviewers in den CAPI-Befragungen (*Module 4 und 5*) ist der *Effekt der sozialen Erwünschtheit* nicht auszuschließen. Die Ausprägung dieses Bias wird jedoch als gering eingestuft, da dies voraussetzt, dass die Teilnehmer die Sanktionierung ihrer Antworten für möglich halten und dass die in der Befragungssituation gültigen Bewertungskriterien bekannt sind [68, 86, 87]. Im Rahmen der Interviews wurde zudem darauf hingewiesen, dass es keine richtigen oder falschen Antworten gibt und dass ausschließlich die persönliche Meinung des Teilnehmers von Interesse ist. Von den Teilnehmern der drei Studien wurde teilweise bemängelt, dass die zu bewertenden hypothetischen Gesundheitszustände zu abstrakt beschrieben und somit als schwer vorstellbar eingestuft wurden. Da bei der schriftlichen Befragung (*Modul 3*) keine Verständnisprobleme geklärt werden konnten und die Interviewer bei den CAPI-Befragungen (*Module 4 und 5*) neutral im Befragungsprozess bleiben sollten, kann nicht ausgeschlossen werden, dass die abverlangte imaginative Leistung zu Missverständnissen geführt haben könnte. Eine Möglichkeit diesem Problem in künftigen Studien zu begegnen, liegt in der Hinzuziehung von Patientenberichten, dem Konzept der „*informed general population values*“ [64]. Auch wenn die *Interviewereinflüsse* in den CAPI-Studien durch die standardisierten Interviewer-Anleitungen, die Interviewerschulung und die enge Qualitätskontrolle generell als gering einzustufen sind, können diese jedoch nicht gänzlich ausgeschlossen werden (z. B. über die Betonung des Interviewers) [88].

6. Implikationen für Forschung, Politik und Praxis

Die *Zielsetzung der vorliegenden Dissertation* bestand darin, ein präferenzbasiertes Value Set für die deutsche Version des EQ-5D-5L für die deutsche Gesellschaft zu ermitteln. In Anbetracht des aufgeführten Bedarfs an geeigneten präferenzbasierten Tarifen für Lebensqualitätsinstrumente in Deutschland leistet diese Dissertationsschrift mit der Entwicklung des deutschen EQ-5D-5L Value Sets einen wichtigen Beitrag, um diese Forschungslücke zu schließen. Dabei lassen sich auf Grundlage der empirischen Investigationen sowie für das entwickelte Value Set Implikationen für Forschung, Politik und Praxis ableiten.

Implikationen für künftige Forschungsprojekte

Die Dissertationsschrift hat einen bedeutsamen *Beitrag zu der Weiterentwicklung eines internationalen Methoden-Protokolls* für die Valuierung eines generischen Lebensqualitätsinstrumentes, den EQ-5D-5L, geliefert. Als Teilstudie des internationalen Forschungsprogramms wurde Evidenz zur Verbesserung der Implementierung des cTTO generiert. So wurde das Feedback-Modul, das Teilnehmern die Verifizierung ihrer cTTO-Antworten ermöglicht, in das Methoden-Protokoll integriert und zu der Version 2.0 weiterentwickelt. Das verbesserte Methoden-Protokoll wurde weltweit erstmalig in der Hauptstudie dieser Dissertation eingesetzt. Die deutsche Valuierungsstudie zeigt, dass das Methoden-Protokoll 2.0 eine hohe Akzeptanz und Durchführbarkeit bei Teilnehmern und Interviewern erzielte und somit für den deutschen Kontext als geeignet einzustufen ist (*Fragestellung 1*). Es konnte eine hohe Datenqualität im Vergleich zu Länderstudien der ersten Generation unter Anwendung der ersten Protokoll-Version erreicht werden. Das deutsche Value Set enthält durchweg konsistente Abschlagswerte für alle gesundheitlichen Einschränkungen, die der EQ-5D-5L beschreibt. Vor diesem Hintergrund kann das *Methoden-Protokoll 2.0 als geeignet* für die Schätzung nationaler EQ-5D-5L Value Sets bewertet werden. In künftigen nationalen Valuierungsstudien für den EQ-5D-5L sollten daher nicht mehr die früheren Protokoll-Versionen, sondern das verbesserte Methoden-Protokoll 2.0 zur Valuierung des EQ-5D-5L verwendet werden.

Ebenso zeigte die Dissertationsschrift, dass die verwendeten Präferenzmethoden *cTTO und DCE eine hohe Übereinstimmung der ermittelten Lebensqualitätsabschläge* erzielten und somit die Kombinierbarkeit beider Datenformen in einem Hybrid ermöglichen (*Fragestellung 2*). Das deutsche Value Set für den EQ-5D-5L berücksichtigt daher beide – als komplementär einzustufenden – Präferenzdaten in einem Hybrid-Modell. In zukünftigen nationalen

Valuierungsstudien ist die Kompatibilität der DCE- und cTTO-Daten für den jeweiligen Kontext zu überprüfen. Auch wenn die verwendeten Modellierungstechniken als derzeitige *Best-Practice* gelten, sollte in künftigen Forschungsprojekten eine Weiterentwicklung des Hybrid-Modells fokussiert werden.

Anhand der Erkenntnisse dieser Dissertationsschrift konnten wichtige *Lessons learned* abgeleitet werden, die aufzeigen, wie das Methoden-Protokoll unter Berücksichtigung der verfügbaren Evidenz in künftigen Länderstudien bestmöglich einzusetzen ist (*Fragestellung 4*). Es zeigte sich, dass kein Methoden-Protokoll inklusive der Interviewer-Anleitungen allein ausreicht, um eine hohe Datenqualität sicherzustellen. Bei der Planung und Durchführung einer Valuierungsstudie sind Lernprozesse auf der Ebene des Forscherteams, des Interviewers sowie des Teilnehmers aktiv zu gestalten. Obwohl TTO bereits seit über 30 Jahren als Präferenzmethode eingesetzt wird, konnten innovative Erkenntnisse generiert werden, die eine Hilfestellung im Sinne von *Best-Practice* für künftige TTO-Studien darstellen. So wurde ersichtlich, dass nicht nur das Engagement des Teilnehmers, sondern auch das des Interviewers einen wesentlichen Einfluss auf die Präzision der TTO-Werte hat. Im Falle von CAPI-Studien können neben den Präferenzwerten auch die erhobenen Meta-Daten dafür genutzt werden, um im Sinne einer formativen Qualitätskontrolle die Protokoll-Compliance zu evaluieren und Interviewer-Effekte zu minimieren. Es wurden verschiedene Phänomene in Bezug auf die TTO-Daten identifiziert, die sich aus dem gewählten Mechanismus der Aufgaben, dem individuellen Antwortverhalten des Teilnehmers und den Eigenschaften des Gesamtdatensatzes ableiten lassen. Zudem wurden Modellierungsverfahren als mögliche Lösungsansätze herausgearbeitet. Diese neuen Erkenntnisse sind nicht nur bei der Valuierung des EQ-5D-5L sondern auch in künftigen TTO-Studien zur direkten Nutzwertbestimmung sowie zur Valuierung anderer Messinstrumente zu berücksichtigen. Zudem sei festgehalten, dass der Einfluss der methodischen Entscheidungen für die Implementierung und Durchführung der Erhebung auf die spätere Datenqualität beträchtlich ist. Daher sollte eine transparente Berücksichtigung der Interviewer-Effekte und der Daten-Verteilung erfolgen, um die Einschätzung der Validität der ermittelten Präferenzwerte vornehmen zu können. Folglich besteht Bedarf an der Entwicklung bzw. Aktualisierung von *Guidelines für die Umsetzung sowie das transparente Publizieren von Valuierungsstudien*. Im Rahmen dieser Synopse erfolgte eine erste Zusammenstellung in Form einer Checkliste der zu treffenden methodischen Entscheidungen, die bei einer Face-to-Face-Studie zur Valuierung eines Lebensqualitätsindexinstrumentes zu treffen sind (*Fragestellung 1*). Diese erhebt keinen Anspruch auf Vollständigkeit, aber könnte dennoch als

Grundlage bei der Entwicklung von Guidelines für Valuierungsstudien fungieren. Für die Berichterstattung von Valuierungsstudien wurde kürzlich eine Checkliste von Xie et al. [62] publiziert. Allerdings ist diese um die Aspekte der Qualitätskontrolle und der Datenqualität im Hinblick auf die Verteilung der beobachteten Daten, Interviewer-Effekte sowie die Protokoll-Compliance zu ergänzen.

Implikationen für Politik und Praxis

Das deutsche Value Set für den EQ-5D-5L bietet eine spezifische Berechnungsgrundlage, um einen singulären EQ-5D-5L Index auf Basis der gesellschaftlichen Präferenzwerte für Deutschland zu aggregieren. Dies ermöglicht den *Einsatz des deutschen EQ-5D-5L sowie des entwickelten Value Sets in verschiedenen Anwendungsbereichen*, wie z. B. gesundheitsökonomischen Studien, und kann somit künftig zur Unterstützung von Allokationsentscheidungen im deutschen Gesundheitswesen eingesetzt werden. Auf diese Weise kann sowohl die Patientenperspektive (durch die selbstberichtete Lebensqualität) als auch die gesellschaftliche Perspektive (durch die Valuierung der Gesundheitszustände) einbezogen werden. Die Entwicklung des EQ-5D-5L Index ermöglicht den Einsatz des deutschen EQ-5D-5L zur indirekten Nutzwertbestimmung und damit in Kosten-Nutzwert-Analysen. Somit kann das Messinstrument für die systematische Evaluation von medizinischen Interventionen unter Effizienzgesichtspunkten indikationsübergreifend angewendet werden und so einen Vergleich zwischen verschiedenen Patientengruppen ermöglichen, wobei auch die Lebensqualität in die ökonomische Beurteilung einbezogen wird. Im Gegensatz zu anderen Ländern spielen Kosten-Nutzen-Bewertungen, die auf QALYs als Maß für den Gesamtnutzen sowie auf den EQ-5D als Basis zurückgreifen, derzeit keine Rolle bei der Entscheidungsfindung für Erstattungsfragen im deutschen Gesundheitswesen. Mit der Durchführung der deutschen Valuierungsstudie für den EQ-5D-5L wurde eine valide und aktuelle Datenbasis zur Nutzwertbestimmung geschaffen. Es ist daher wünschenswert, dass diese von dem Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) und Entscheidungsträgern in Deutschland akzeptiert wird, sodass Lebensqualitätsdaten, die mit dem EQ-5D-5L erhoben werden, bei Entscheidungen im Gesundheitswesen verstärkt einbezogen werden. Ferner ist zu erhoffen, dass eine weitere Implikation der vorliegenden Dissertation darin liegt, dass mehr nationale Studien zur Kosten-Effektivität in Deutschland durchgeführt werden. Auch wenn der Fokus dieser Dissertation auf der Bewertung von Lebensqualität im Kontext gesundheitsökonomischer Evaluationen liegt, sei erwähnt, dass ökonomische Analysen nicht den einzigen Anwendungsbereich von Lebensqualitätsverfahren darstellen. Der deutsche EQ-5D-5L

und das entwickelte Value Set können ebenso in klinischen Studien, Studien zur Versorgungsforschung, nationalen Gesundheitssurveys und bei Routinebefragungen im Rahmen des Qualitätsmanagements, wie z. B. in Krankenhäusern, eingesetzt werden [5, 11]. Folglich kann die präferenzbasierte Lebensqualitätsmessung auf Basis des EQ-5D-5L in verschiedenen Versorgungsbereichen des deutschen Gesundheitswesens Anwendung finden. Das Value Set sollte dabei allerdings (vorerst) ausschließlich für dasjenige Instrument, dessen Gesundheitszustände valuiert wurden – für den EQ-5D-5L – sowie in dem Sprach- und Kulturraum – in Deutschland – verwendet werden, wo die zugrundeliegende Valuierungsstudie durchgeführt wurde. Die *Anwendbarkeit der ermittelten Abschlagswerte* für andere deutschsprachige Länder gilt es zunächst in weiteren Forschungsstudien zu analysieren. Zudem sollte eine ungeprüfte Anwendung des entwickelten Value Sets für die Aggregation der Gesundheitszustände des EQ-5D-Y-5L nicht erfolgen, da die Erkenntnisse dieser Dissertationsschrift gezeigt haben, dass die gesundheitlichen Einschränkungen bei Kindern anders gewichtet werden als bei Erwachsenen. Folglich implizieren die Ergebnisse, dass die Generierung von spezifischen Value Sets für Kinderlebensqualitätsfragebögen, wie der EQ-5D-Y-5L, Gegenstand künftiger Forschungsbestrebungen sein sollte.

Eine weitere mögliche Implikation der generierten Erkenntnisse liegt darin, dass eine Veränderung bei der Lebensqualitätsmessung in Deutschland zu erwarten ist: der *Wechsel von dem EQ-5D-3L zu dem EQ-5D-5L*. Dies ist zum einen auf ein besseres deskriptives System zurückzuführen. Erste komparative Head-to-Head-Studien der beiden Instrumente zeigen, dass bei dem EQ-5D-5L eine Verbesserung der Messeigenschaften in Bezug auf (1) eine gesteigerte Sensitivität und (2) eine gesteigerte Diskriminierbarkeit in Verbindung mit einer größeren Kapazität, Unterschiede sowohl auf Level des Individuums als auch zwischen Gruppen abzubilden, zu verzeichnen ist [89–91]. Zum anderen wird der Wechsel zu dem EQ-5D-5L in Deutschland erwartet, da die Valuierungsstudie für den EQ-5D-5L eine hohe Datenqualität erzielte und den Hauptkritikpunkten der EQ-5D-3L Valuierungsstudie begegnet, die vor mehr als 15 Jahren durchgeführt wurde. Auch wenn der empirische Vergleich der deutschen Value Sets noch aussteht, liefern erste komparative Head-to-Head-Studien anderer nationaler Value Sets Evidenz, dass nicht nur bei dem deskriptiven System, sondern auch bei den EQ-5D-5L Value Sets und den resultierenden Nutzwerten eine höhere Messpräzision zu verzeichnen ist [89]. Zudem wurde im Gegensatz zu dem EQ-5D-3L für die Valuierung des deutschen EQ-5D-5L ein international standardisiertes Methoden-Protokoll mit verbesserten Valuation-Methoden und einer innovativen Qualitätskontrolle eingesetzt. Das EQ-5D-5L Value Set basiert auf den gegenwärtigen gesellschaftlichen Präferenzwerten und stellt nicht

nur eine aktuelle sondern zugleich eine robuste Grundlage zur Aggregation von EQ-5D-5L Indexwerten dar. In Anbetracht der aufgeführten Argumente ist das entwickelte EQ-5D-5L Value Set daher als das bevorzugte Value Set für Deutschland einzustufen. Aufgrund der ausstehenden empirischen Analyse hinsichtlich der Einflussfaktoren auf die Bewertung von Gesundheit, kann die Frage der Gültigkeit des entwickelten EQ-5D-5L Value Sets nicht abschließend beantwortet werden. Einem Value Set sollte ein gewisser *Lebenszyklus* mit antizipierten Aktualisierungen zugesprochen werden, welche mögliche Änderungen bei den gesellschaftlichen Gesundheitspräferenzen durch politische, soziale oder epidemiologische Veränderungen und verbesserte Präferenzmethoden reflektieren. Auf die Frage, wie häufig Value Sets aktualisiert werden sollen, gibt es derzeit jedoch keine endgültige Antwort.

Das ermittelte Set an gesellschaftlichen Präferenzwerten für den EQ-5D-5L zeigt auf, welchen *Stellenwert verschiedene Lebensqualitätsdimensionen entsprechend der Präferenz der deutschen Bevölkerung* einnehmen (*Fragestellung 3*), wobei Angst/Niedergeschlagenheit und insbesondere Schmerzen/Körperliche Beschwerden das größte Gewicht zukommt. Diese Präferenzwerte haben folglich Implikationen für die Bewertung von Behandlungsmaßnahmen, die sich auf die Lebensqualität von Patienten auswirken. Die Erkenntnis, dass insbesondere den aufgeführten Symptomen ein großer Einfluss auf die Lebensqualität zugesprochen wird, sollte in künftigen Entscheidungen im Gesundheitswesen Berücksichtigung finden. So sollte überprüft werden, ob mehr Ressourcen in die Schmerz- und Depressionsforschung unter Berücksichtigung einer allokativen Effizienz investiert werden können. Außerdem sollte die Verbesserung der Versorgungsstrukturen für (chronische) Schmerzpatienten sowie psychische Störungsbilder fokussiert werden. Dabei ist zu prüfen, ob mehr Ressourcen in die Primär-, Sekundär- und Tertiärprävention sowie Therapiemaßnahmen der assoziierten Erkrankungsbilder effizient alloziert werden können.

Für die deutsche Lebensqualitätsforschung stellt die Ermittlung des präferenzbasierten Value Sets für den EQ-5D-5L ein Novum dar und leistet einen wichtigen Beitrag, um patientenrelevante Lebensqualitätseffekte und die dazugehörigen gesellschaftlichen Präferenzen bei (Allokations-) Entscheidungen im Gesundheitswesen einzubeziehen und somit eine effektive und effiziente Gesundheitsversorgung – insbesondere für chronisch erkrankte Patienten – sicherstellen zu können.

Referenzen

1. Hurrelmann K, Laaser U, Razum O. Entwicklung und Perspektiven der Gesundheitswissenschaften in Deutschland. In: Hurrelmann K, Razum O, Hrsg. Handbuch Gesundheitswissenschaften. 6. Auflage. Weinheim, Basel: Beltz Juventa; 2016. S. 15–51.
2. Kohlmann T. Patientenberichtete Studienendpunkte - Stand in Forschung und Praxis. Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen. 2010;104:259–63.
3. Maaz A, Winter MH-J, Kuhlmeier A. Der Wandel des Krankheitspanoramas und die Bedeutung chronischer Erkrankungen (Epidemiologie, Kosten). In: Badura B, Schellschmidt H, Vetter C, Hrsg. Fehlzeiten-Report 2006: Chronische Krankheiten. Berlin, Heidelberg: Springer-Verlag; 2007. S. 5–23.
4. Radoschewski M. Gesundheitsbezogene Lebensqualität - Konzepte und Maße: Entwicklungen und Stand im Überblick. Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz. 2000;43:165–89.
5. Koller M, Neugebauer EAM, Augustin M, Büssing A, Farin E, Klinkhammer-Schalke M, et al. Die Erfassung von Lebensqualität in der Versorgungsforschung - konzeptuelle, methodische und strukturelle Voraussetzungen. Gesundheitswesen. 2009;71:864–72.
6. Valderas JM, Alonso J. Patient reported outcome measures: a model-based classification system for research and clinical practice. Quality of Life Research. 2008;17:1125–35.
7. Sozialgesetzbuch (SGB) Fünftes Buch (V) - Gesetzliche Krankenversicherung. § 35b Kosten-Nutzen-Bewertung von Arzneimitteln. Artikel 1 des Gesetzes vom 20. Dezember 1988, BGBl. I S. 2477.
8. Greiner W, Uber A. Gesundheitsökonomische Studien und der Einsatz von Lebensqualitätsindizes am Beispiel der LQ-Indexes EQ-5D (EuroQol). In: Ravens-Sieberer U, Cieza A, Hrsg. Lebensqualität und Gesundheitsökonomie in der Medizin: Konzepte, Methoden, Anwendung. Landsberg: Ecomed; 2000. S. 336–351.
9. Schöffski O. Lebensqualität als Ergebnisparameter in gesundheitsökonomischen Studien. In: Schöffski O, Graf von der Schulenburg J-M, Hrsg. Gesundheitsökonomische Evaluationen. 4. Auflage. Berlin, Heidelberg: Springer-Verlag; 2012. S. 327–340.
10. Greiner W, Damm O. Die Berechnung von Kosten und Nutzen. In: Schöffski O, Graf von der Schulenburg J-M, Hrsg. Gesundheitsökonomische Evaluationen. 4. Auflage. Berlin, Heidelberg: Springer-Verlag; 2012. S. 23–42.
11. Greiner W. Der EQ-5D der EuroQol-Gruppe. In: Schöffski O, Graf von der Schulenburg J-M, Hrsg. Gesundheitsökonomische Evaluationen. 4. Auflage. Berlin, Heidelberg: Springer-Verlag; 2012. S. 411–422.
12. Antes G, Jöckel K-H, Kohlmann T, Raspe H, Wasem J. Kommentierende Synopse der Fachpositionen zur Kosten-Nutzenbewertung für Arzneimittel. 2007. https://www.mm.wiwi.uni-due.de/fileadmin/fileupload/BWL-MEDMAN/Forschung/SynopseAntes-et-al_neu.pdf. [Stand: 11. Juni 2018].

13. Rabin R, de Charro F. EQ-5D: a measure of health status from the EuroQol Group. *Annals of Medicine*. 2001;33:337–43.
14. Devlin NJ, Krabbe PFM. The development of new research methods for the valuation of EQ-5D-5L. *The European Journal of Health Economics*. 2013;14(Supplement 1):1-3.
15. Herdman M, Gudex C, Lloyd A, Janssen MF, Kind P, Parkin D, et al. Development and preliminary testing of the new five-level version of EQ-5D (EQ-5D-5L). *Quality of Life Research*. 2011;20:1727–36.
16. Oppe M, Devlin NJ, van Hout B, Krabbe PFM, de Charro F. A program of methodological research to arrive at the new international EQ-5D-5L valuation protocol. *Value in Health*. 2014;17:445–53.
17. Graf von der Schulenburg J-M, Greiner W. *Gesundheitsökonomik*. 3. Auflage. Tübingen: Mohr Siebeck; 2013.
18. Schöffski O, Schumann A, Kuhlmann A, Schwarzbach C. Das Schwellenwertkonzept: Theorie sowie Umsetzung beim IQWiG und anderen Institutionen. In: Schöffski O, Graf von der Schulenburg J-M, Hrsg. *Gesundheitsökonomische Evaluationen*. 4. Auflage. Berlin, Heidelberg: Springer-Verlag; 2012. S. 111–153.
19. Kleinewefers H. *Einführung in die Wohlfahrtsökonomie: Theorie - Anwendung - Kritik*. Stuttgart: Kohlhammer; 2008.
20. Drummond MF. Output measurement for resource allocation decisions in health care. *Oxford Review of Economic Policy*. 1989;5:59–74.
21. Greiner W. Methoden der gesundheitsökonomischen Evaluation. In: Hurrelmann K, Razum O, Hrsg. *Handbuch Gesundheitswissenschaften*. 6. Auflage. Weinheim, Basel: Beltz Juventa; 2016. S. 375–402.
22. Schlander M. Gesundheitsökonomie: Der Effizienz auf der Spur. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*. 2009;103:117–25.
23. von Neumann J, Morgenstern O. *Theory of games and economic behavior*. New York: Princeton University Press; 1944.
24. Drummond MF, Sculpher MJ, Claxton K, Stoddart GL, Torrance GW. *Methods for the economic evaluation of health care programmes*. 4. Auflage. Oxford: Oxford University Press; 2015.
25. Brazier J, Ratcliffe J, Salomon JA, Tsuchiya A. *Measuring and valuing health benefits for economic evaluation*. 2. Auflage. Oxford: Oxford University Press; 2017.
26. Mühlbacher AC, Bethge S, Ekert S, Tockhorn A, Nübling M. Der Wert von Innovationen im Gesundheitswesen: Spielen die Patientenpräferenzen eine Rolle? *Recht und Politik im Gesundheitswesen*. 2008;14:53–62.
27. Schöffski O. Nutzentheoretische Lebensqualitätsmessung. In: Schöffski O, Graf von der Schulenburg J-M, Hrsg. *Gesundheitsökonomische Evaluationen*. 4. Auflage. Berlin, Heidelberg: Springer-Verlag; 2012. S. 341–391.

28. Breyer F, Zweifel P, Kifmann M. Gesundheitsökonomik. 6. Auflage. Berlin, Heidelberg: Springer Gabler; 2013.
29. Tsuchiya A, Williams A. Welfare economics and economic evaluation. In: Drummond M, McGuire A, Hrsg. Economic evaluation in health care: merging theory with practice. Oxford: Oxford University Press; 2001. S. 22–45.
30. Hurley J. Welfarism, extra-welfarism and evaluative economic analysis in health sector. In: Barer ML, Getzen TE, Stoddart GL, Hrsg. Health, health care and health economics: Perspectives on distribution. Chichester: John Wiley & Sons; 1998. S. 373–395.
31. Brouwer WBF, Culyer AJ, van Exel NJA, Rutten FFH. Welfarism vs. extra-welfarism. *Journal of Health Economics*. 2008;27:325–38.
32. Schöffski O. Grundformen gesundheitsökonomischer Evaluationen. In: Schöffski O, Graf von der Schulenburg J-M, Hrsg. Gesundheitsökonomische Evaluationen. 4. Auflage. Berlin, Heidelberg: Springer-Verlag; 2012. S. 43–70.
33. Zentner A, Busse R. Internationale Standards der Kosten-Nutzen-Bewertung. *Gesundheitsökonomie & Qualitätsmanagement*. 2006;11:368–73.
34. Feeny D, Krahn M, Prosser LA, Salomon JA. Valuing health outcomes. In: Neumann PJ, Sanders GD, Russell LB, Siegel JE, Ganiats TG, Hrsg. Cost effectiveness in health and medicine. 2. Auflage. New York: Oxford University Press; 2017. S. 167–199.
35. Schöffski O, Greiner W. Das QALY-Konzept als prominentester Vertreter der Kosten-Nutzwert-Analyse. In: Schöffski O, Graf von der Schulenburg J-M, Hrsg. Gesundheitsökonomische Evaluationen. 4. Auflage. Berlin, Heidelberg: Springer-Verlag; 2012. S. 71–110.
36. Weinstein MC, Torrance G, McGuire A. QALYs: the basics. *Value in Health*. 2009;12(Supplement 1):5-9.
37. Stratton KR, Durch JS, Lawrence RS. Vaccines for the 21st century: a tool for decisionmaking. Washington, D.C.: National Academy Press; 2000.
38. Deutscher Ethikrat. Nutzen und Kosten im Gesundheitswesen - Zur normativen Funktion ihrer Bewertung. 2011. <http://www.ethikrat.org/dateien/pdf/stellungnahme-nutzen-und-kosten-im-gesundheitswesen.pdf>. [Stand: 11. Juni 2018].
39. Nord E, Daniels N, Kamlet M. QALYs: some challenges. *Value in Health*. 2009;12(Supplement 1):10-5.
40. Johnson FR. Editorial: Moving the QALY forward or just stuck in traffic? *Value in Health*. 2009;12(Supplement 1):38-9.
41. Graf von der Schulenburg J-M, Vauth C, Mittendorf T, Greiner W. Methoden zur Ermittlung von Kosten-Nutzen-Relationen für Arzneimittel in Deutschland. *Gesundheitsökonomie & Qualitätsmanagement*. 2007;12(Supplement 1):2-25.
42. Tinelli M, Ryan M, Bond C, Scott A. Valuing benefits to inform a clinical trial in pharmacy: do differences in utility measures at baseline affect the effectiveness of the intervention? *Pharmacoeconomics*. 2013;31:163–71.

43. Torrance GW. Measurement of health state utilities for economic appraisal: a review. *Journal of Health Economics*. 1986;5:1–30.
44. Böhmer S, Kohlmann T. Verfahren zur Bewertung von Gesundheitszuständen und Lebensqualität. In: Ravens-Sieberer U, Cieza A, Hrsg. *Lebensqualität und Gesundheitsökonomie in der Medizin: Konzepte, Methoden, Anwendung*. Landsberg: Ecomed; 2000. S. 53–72.
45. Graf von der Schulenburg J-M, Greiner W, Jost F, Klusen N, Kubin M, Leidl R, et al. Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation - dritte und aktualisierte Fassung des Hannoveraner Konsens. *Gesundheitsökonomie & Qualitätsmanagement*. 2007;12:285–90.
46. Greiner W, Claes C, Busschbach JJV, Graf von der Schulenburg J-M. Validating the EQ-5D with time trade off for the German population. *The European Journal of Health Economics*. 2005;6:124–30.
47. Leidl R, Reitmeir P. An experience-based value set for the EQ-5D-5L in Germany. *Value in Health*. 2017;20:1150–6.
48. van Hout B, Janssen MF, Feng Y-S, Kohlmann T, Busschbach J, Golicki D, et al. Interim scoring for the EQ-5D-5L: mapping the EQ-5D-5L to EQ-5D-3L value sets. *Value in Health*. 2012;15:708–15.
49. Szende A, Oppe M, de Charro F. Comparative review of time trade-off value sets. In: Szende A, Oppe M, Devlin N, Hrsg. *EQ-5D value sets: inventory, comparative review and user guide*. Dordrecht: Springer; 2007. S. 21–28.
50. Brooks R. *The EuroQol Group after 25 years*. Dordrecht: Springer; 2013.
51. Lamers LM. The transformation of utilities for health states worse than death: consequences for the estimation of EQ-5D value sets. *Medical Care*. 2007;45:238–44.
52. Tilling C, Devlin N, Tsuchiya A, Buckingham K. Protocols for time tradeoff valuations of health states worse than dead: a literature review. *Medical decision making*. 2010;30:610–9.
53. Devlin NJ, Tsuchiya A, Buckingham K, Tilling C. A uniform time trade off method for states better and worse than dead: feasibility study of the 'lead time' approach. *Health Economics*. 2011;20:348–61.
54. Janssen BMF, Oppe M, Versteegh MM, Stolk EA. Introducing the composite time trade-off: a test of feasibility and face validity. *The European Journal of Health Economics*. 2013;14(Supplement 1):5-13.
55. Stolk EA, Oppe M, Scalone L, Krabbe PFM. Discrete choice modeling for the quantification of health states: the case of the EQ-5D. *Value in Health*. 2010;13:1005–13.
56. Amaya-Amaya M, Gerard K, Ryan M. Discrete choice experiments in a nutshell. In: Ryan M, Gerard K, Amaya-Amaya M, Hrsg. *Using discrete choice experiments to value health and health care*. Dordrecht: Springer; 2008. S. 13–46.
57. McFadden D. Conditional logit analysis of qualitative choice behavior. In: Zarembka P, Hrsg. *Frontiers in econometrics*. New York: Academic Press; 1974. S. 105–142.

58. Oppe M, Rand-Hendriksen K, Shah K, Ramos-Goñi JM, Luo N. EuroQol protocols for time trade-off valuation of health outcomes. *PharmacoEconomics*. 2016;34:993–1004.
59. Shah K, Rand-Hendriksen K, Ramos-Goñi JM, Prause AJ, Stolk E. Improving the quality of data collected in EQ-5D-5L valuation studies: a summary of the EQ-VT research methodology programme. In: Proceedings of the 31st Scientific Plenary Meeting of the EuroQol Group; Stockholm, Sweden, 25.-26. September 2014; 2015. S. 1–18. http://eq-5dpublications.euroqol.org/download?id=0_53918&fileId=54332. [Stand: 11. Juni 2018].
60. Ramos-Goñi JM, Oppe M, Slaap B, Busschbach JJV, Stolk E. Quality control process for EQ-5D-5L valuation studies. *Value in Health*. 2017;20:466–73.
61. Ravens-Sieberer U, Wille N, Badia X, Bonsel G, Burström K, Cavrini G, et al. Feasibility, reliability, and validity of the EQ-5D-Y: results from a multinational study. *Quality of Life Research*. 2010;19:887–97.
62. Xie F, Pickard AS, Krabbe PFM, Revicki D, Viney R, Devlin N, et al. A checklist for reporting valuation studies of multi-attribute utility-based instruments (CREATE). *PharmacoEconomics*. 2015;33:867–77.
63. Bleichrodt H, Johannesson M. Standard gamble, time trade-off and rating scale: experimental results on the ranking properties of QALYs. *Journal of Health Economics*. 1997;16:155–75.
64. Brazier J, Akehurst R, Brennan A, Dolan P, Claxton K, McCabe C, et al. Should patients have a greater role in valuing health states? *Applied Health Economics and Health Policy*. 2005;4:201–8.
65. Leidl R, Reitmeir P. A value set for the EQ-5D based on experienced health states: development and testing for the German population. *PharmacoEconomics*. 2011;29:521–34.
66. Ramos-Goñi JM, Pinto-Prades JL, Oppe M, Cabasés JM, Serrano-Aguilar P, Rivero-Arias O. Valuation and modeling of EQ-5D-5L health states using a hybrid approach. *Medical Care*. 2017;55:51–8.
67. Ramos-Goñi JM, Craig B, Oppe M, van Hout B. Combining continuous and dichotomous responses in a hybrid model. EuroQol working paper series. 2016; 16002. https://euroqol.org/wp-content/uploads/working_paper_series/EuroQol_Working_Paper_Series_Manuscript_16002_-_Juan_Ramos-Goni.pdf. [Stand: 11. Juni 2018].
68. Hlawatsch A, Krickl T. Einstellungen zu Befragungen. In: Baur N, Blasius J, Hrsg. *Handbuch Methoden der empirischen Sozialforschung*. Wiesbaden: Springer Fachmedien; 2014. S. 305–311.
69. Statistisches Bundesamt. *Pflegestatistik 2015. Pflege im Rahmen der Pflegeversicherung Ländervergleich – Ambulante Pflegedienste*. Wiesbaden: Statistisches Bundesamt; 2017. https://www.destatis.de/DE/Publikationen/Thematisch/Gesundheit/Pflege/PflegeDeutschland_ergebnisse5224001159004.pdf?__blob=publicationFile. [Stand: 11. Juni 2018].
70. Sozialgesetzbuch (SGB) Elftes Buch (XI) - Soziale Pflegeversicherung. § 4 SGB XI Art und Umfang der Leistungen. Artikel 1 des Gesetzes vom 26. Mai 1994, BGBl. I S. 1014.

71. Sozialgesetzbuch (SGB) Fünftes Buch (V) - Gesetzliche Krankenversicherung. § 33 SGB V Hilfsmittel. Artikel 1 des Gesetzes vom 20. Dezember 1988, BGBl. I S. 2477.
72. Devlin NJ, Shah KK, Feng Y, Mulhern B, van Hout B. Valuing health-related quality of life: an EQ-5D-5L value set for England. *Health Economics*. 2017;27:7–22.
73. Ramos-Goñi JM, Craig BM, Oppe M, Ramallo-Fariña Y, Pinto-Prades JL, Luo N, et al. Handling data quality issues to estimate the Spanish EQ-5D-5L value set using a hybrid interval regression approach. *Value in Health*. 2018;21:596–604.
74. Versteegh MM, Vermeulen KM, Evers SMAA, de Wit GA, Prenger R, Stolk EA. Dutch tariff for the five-level version of EQ-5D. *Value in Health*. 2016;19:343–52.
75. Purba FD, Hunfeld JAM, Iskandarsyah A, Fitriana TS, Sadarjoen SS, Ramos-Goñi JM, et al. The Indonesian EQ-5D-5L value set. *PharmacoEconomics*. 2017;35:1153–65.
76. Augustovski F, Rey-Ares L, Irazola V, Garay OU, Gianneo O, Fernández G, et al. An EQ-5D-5L value set based on Uruguayan population preferences. *Quality of Life Research*. 2016;25:323–33.
77. Kim S-H, Ahn J, Ock M, Shin S, Park J, Luo N, et al. The EQ-5D-5L valuation study in Korea. *Quality of Life Research*. 2016;25:1845–52.
78. Knies S, Evers SMAA, Candel MJJM, Severens JL, Ament AJHA. Utilities of the EQ-5D: transferable or not? *PharmacoEconomics*. 2009;27:767–79.
79. Günther OH, Friemel S, Bernert S, Matschinger H, Angermeyer MC, König H-H. Die Krankheitslast von depressiven Erkrankungen in Deutschland: Ergebnisse aus dem Projekt European Study of the Epidemiology of Mental Disorders (ESEMeD). *Psychiatrische Praxis*. 2007;34:292–301.
80. Gerste B, Roick C. Prävalenz und Inzidenz sowie Versorgung depressiver Erkrankungen in Deutschland – Eine Analyse auf Basis der in Routinedaten dokumentierten Depressionsdiagnosen. In: Klauber J, Günster C, Gerste B, Robra B-P, Schmacke N, Hrsg. Versorgungs-Report 2013/2014. Schwerpunkt: Depression. Stuttgart: Schattauer; 2014. S. 21–54.
81. Kreimeier S, Åström M, Burström K, Egmar A, Gusi N, Herdman M, et al. Extension of the labels within the EQ-5D-Y. *Value in Health*. 2016;19:480.
82. Feng Y, Devlin NJ, Shah KK, Mulhern B, van Hout B. New methods for modelling EQ-5D-5L value sets: an application to English data. *Health Economics*. 2017;27:23–38.
83. Attema AE, Brouwer WBF. Deriving time discounting correction factors for TTO tariffs. *Health Economics*. 2014;23:410–25.
84. Attema AE, Edelaar-Peeters Y, Versteegh MM, Stolk EA. Time trade-off: one methodology, different methods. *The European Journal of Health Economics*. 2013;14(Supplement 1):53–64.
85. Döring N, Bortz J. Stichprobenziehung. In: Döring N, Bortz J, Hrsg. *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 5. Auflage. Berlin, Heidelberg: Springer; 2016. S. 291–319.

86. Esser H. Die Erklärung systematischer Fehler in Interviews: Befragtenverhalten als „rational choice“. In: Wittenberg R, Hrsg. Person - Situation - Institution - Kultur: Günter Büschges zum 65. Geburtstag. Berlin: Duncker & Humblot; 1991. S. 59–78.
87. Stocké V. Entstehungsbedingungen von Antwortverzerrungen durch soziale Erwünschtheit: Ein Vergleich der Prognosen der Rational-Choice Theorie und des Modells der Frame-Selektion. *Zeitschrift für Soziologie*. 2004;33:303–20.
88. Glantz A, Michael T. Interviewereffekte. In: Baur N, Blasius J, Hrsg. Handbuch Methoden der empirischen Sozialforschung. Wiesbaden: Springer Fachmedien; 2014. S. 313–322.
89. Janssen MF, Bonsel GJ, Luo N. Is EQ-5D-5L better than EQ-5D-3L? A head-to-head comparison of descriptive systems and value sets from seven countries. *Pharmacoeconomics* 2018;36:675-97
90. Janssen MF, Pickard AS, Golicki D, Gudex C, Niewada M, Scalone L, et al. Measurement properties of the EQ-5D-5L compared to the EQ-5D-3L across eight patient groups: a multi-country study. *Quality of Life Research*. 2013;22:1717–27.
91. Pickard AS, de Leon MC, Kohlmann T, Cella D, Rosenbloom S. Psychometric comparison of the standard EQ-5D to a 5 level version in cancer patients. *Medical Care*. 2007;45:259–63.

Anhang

Anhang 1: Deutsche Version des EQ-5D-3L

Bitte geben Sie an, welche Aussagen Ihren heutigen Gesundheitszustand am besten beschreiben, indem Sie ein Kreuz in ein Kästchen jeder Gruppe machen.

Beweglichkeit / Mobilität

- Ich habe keine Probleme herumzugehen
- Ich habe einige Probleme herumzugehen
- Ich bin ans Bett gebunden

Für sich selbst sorgen

- Ich habe keine Probleme, für mich selbst zu sorgen
- Ich habe einige Probleme, mich selbst zu waschen oder mich anzuziehen
- Ich bin nicht in der Lage, mich selbst zu waschen oder anzuziehen

Alltägliche Tätigkeiten (z.B. Arbeit, Studium, Hausarbeit, Familien- oder Freizeitaktivitäten)

- Ich habe keine Probleme, meinen alltäglichen Tätigkeiten nachzugehen
- Ich habe einige Probleme, meinen alltäglichen Tätigkeiten nachzugehen
- Ich bin nicht in der Lage, meinen alltäglichen Tätigkeiten nachzugehen

Schmerzen / Körperliche Beschwerden

- Ich habe keine Schmerzen oder Beschwerden
- Ich habe mäßige Schmerzen oder Beschwerden
- Ich habe extreme Schmerzen oder Beschwerden

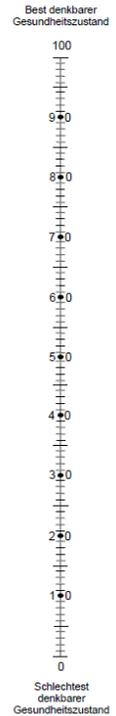
Angst / Niedergeschlagenheit

- Ich bin nicht ängstlich oder deprimiert
- Ich bin mäßig ängstlich oder deprimiert
- Ich bin extrem ängstlich oder deprimiert

Um Sie bei der Einschätzung, wie gut oder wie schlecht Ihr Gesundheitszustand ist, zu unterstützen, haben wir eine Skala gezeichnet, ähnlich einem Thermometer. Der best denkbare Gesundheitszustand ist mit einer "100" gekennzeichnet, der schlechteste mit "0".

Wir möchten Sie nun bitten, auf dieser Skala zu kennzeichnen, wie gut oder schlecht Ihrer Ansicht nach Ihr persönlicher Gesundheitszustand heute ist. Bitte verbinden Sie dazu den untenstehenden Kasten mit dem Punkt auf der Skala, der Ihren heutigen Gesundheitszustand am besten wiedergibt.

Ihr heutiger
Gesundheitszustand



Anhang 2: Deutsche Version des EQ-5D-5L

Bitte kreuzen Sie unter jeder Überschrift DAS Kästchen an, das Ihre Gesundheit HEUTE am besten beschreibt.

BEWEGLICHKEIT / MOBILITÄT

- Ich habe keine Probleme herumzugehen
- Ich habe leichte Probleme herumzugehen
- Ich habe mäßige Probleme herumzugehen
- Ich habe große Probleme herumzugehen
- Ich bin nicht in der Lage herumzugehen

FÜR SICH SELBST SORGEN

- Ich habe keine Probleme, mich selbst zu waschen oder anzuziehen
- Ich habe leichte Probleme, mich selbst zu waschen oder anzuziehen
- Ich habe mäßige Probleme, mich selbst zu waschen oder anzuziehen
- Ich habe große Probleme, mich selbst zu waschen oder anzuziehen
- Ich bin nicht in der Lage, mich selbst zu waschen oder anzuziehen

ALLTÄGLICHE TÄTIGKEITEN (z.B. Arbeit, Studium, Hausarbeit, Familien- oder Freizeitaktivitäten)

- Ich habe keine Probleme, meinen alltäglichen Tätigkeiten nachzugehen
- Ich habe leichte Probleme, meinen alltäglichen Tätigkeiten nachzugehen
- Ich habe mäßige Probleme, meinen alltäglichen Tätigkeiten nachzugehen
- Ich habe große Probleme, meinen alltäglichen Tätigkeiten nachzugehen
- Ich bin nicht in der Lage, meinen alltäglichen Tätigkeiten nachzugehen

SCHMERZEN / KÖRPERLICHE BESCHWERDEN

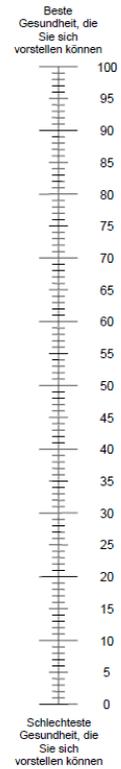
- Ich habe keine Schmerzen oder Beschwerden
- Ich habe leichte Schmerzen oder Beschwerden
- Ich habe mäßige Schmerzen oder Beschwerden
- Ich habe starke Schmerzen oder Beschwerden
- Ich habe extreme Schmerzen oder Beschwerden

ANGST / NIEDERGESCHLAGENHEIT

- Ich bin nicht ängstlich oder deprimiert
- Ich bin ein wenig ängstlich oder deprimiert
- Ich bin mäßig ängstlich oder deprimiert
- Ich bin sehr ängstlich oder deprimiert
- Ich bin extrem ängstlich oder deprimiert

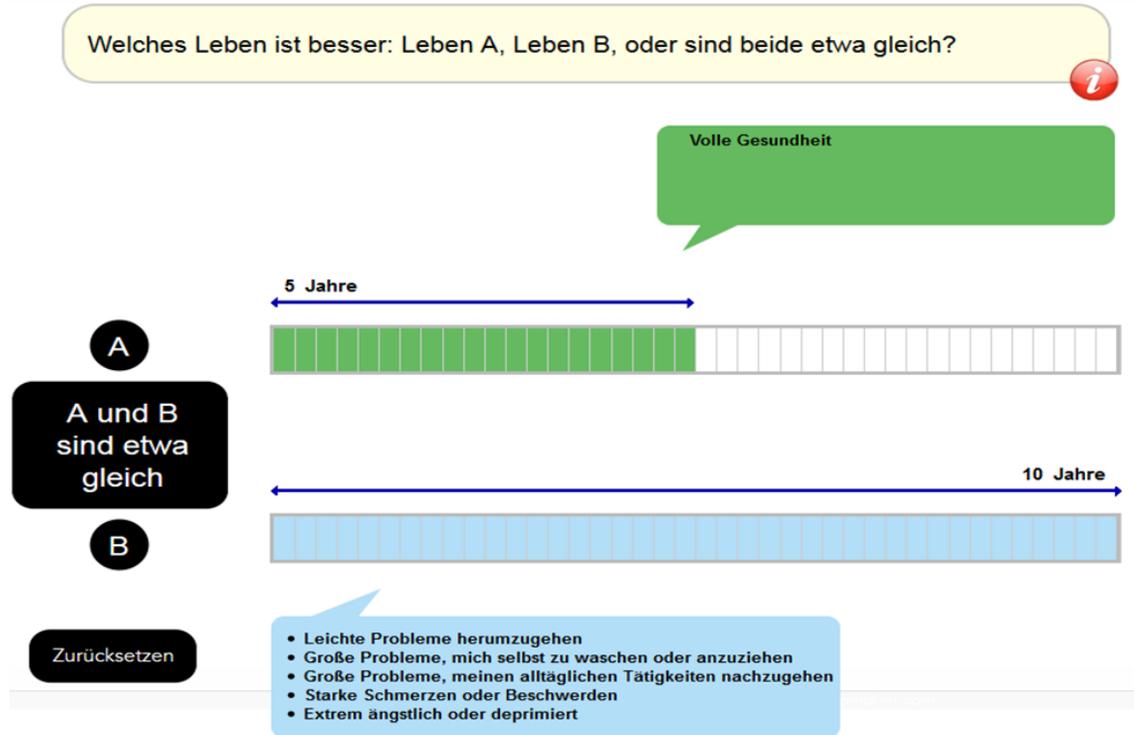
- Wir wollen herausfinden, wie gut oder schlecht Ihre Gesundheit HEUTE ist.
- Diese Skala ist mit Zahlen von 0 bis 100 versehen.
- 100 ist die beste Gesundheit, die Sie sich vorstellen können. 0 (Null) ist die schlechteste Gesundheit, die Sie sich vorstellen können.
- Bitte kreuzen Sie den Punkt auf der Skala an, der Ihre Gesundheit HEUTE am besten beschreibt.
- Jetzt tragen Sie bitte die Zahl, die Sie auf der Skala angekreuzt haben, in das Kästchen unten ein.

IHRE GESUNDHEIT HEUTE =

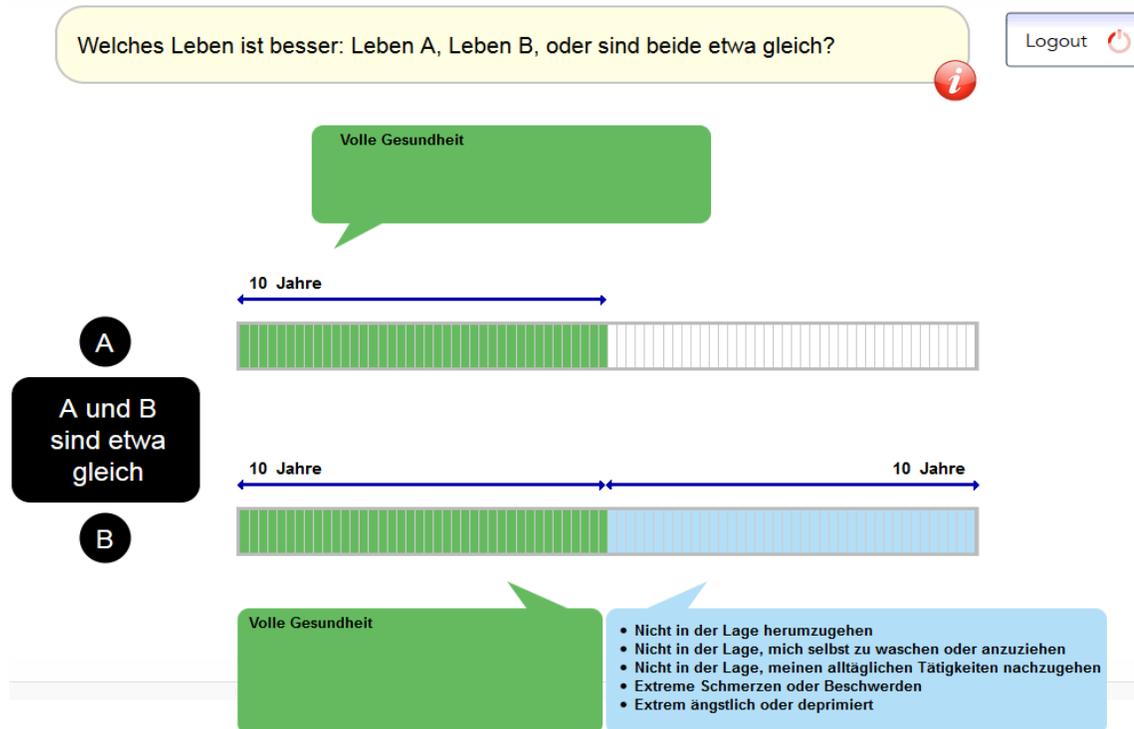


Anhang 3: Implementierung der Präferenzmethoden in der EuroQol-Valuation Technique

Composite Time-Trade-Off für Gesundheitszustände „Besser als der Tod“



Composite Time-Trade-Off für Gesundheitszustände „Schlechter als der Tod“



Discrete Choice Experiment

Welcher Zustand ist besser, A oder B?

Logout 



- Leichte Probleme herumzugehen
- Keine Probleme, mich selbst zu waschen oder anzuziehen
- Leichte Probleme, meinen alltäglichen Tätigkeiten nachzugehen
- Keine Schmerzen oder Beschwerden
- Nicht ängstlich oder deprimiert

A

- Keine Probleme herumzugehen
- Keine Probleme, mich selbst zu waschen oder anzuziehen
- Keine Probleme, meinen alltäglichen Tätigkeiten nachzugehen
- Leichte Schmerzen oder Beschwerden
- Nicht ängstlich oder deprimiert

B

Module der kumulativen Dissertation

Modul 1

Klose, K., Kreimeier, S., Tangermann, U., Aumann, I., Damm, K. (2016):

Patient- and person-reports on healthcare: preferences, outcomes, experiences, and satisfaction. *Health Economics Review*, 6(18) (DOI: 10.1186/s13561-016-0094-6).

Modul 2

Greiner, W., **Klose, K.** (2014):

Lebensqualitätsbewertung und Utilities in der Gesundheitsökonomie. *Zeitschrift für Evidenz, Fortbildung und Qualität im Gesundheitswesen*, 108(2-3): 120-125.

Modul 3

Kind, P., **Klose, K.**, Gusi, N., Olivares, P.R., Greiner, W. (2015):

Can adult weights be used to value child health states? Testing the influence of perspective in valuing EQ-5D-Y. *Quality of Life Research*, 24(10): 2519-2539.

Modul 4

Ludwig, K., Schulenburg, J-M. Graf v. d., Greiner, W. (2017):

Valuation of the EQ-5D-5L with composite time trade-off for the German population - an exploratory study. *Health and Quality of Life Outcomes*, 15(39) (DOI: 10.1186/s12955-017-0617-9).

Modul 5

Ludwig, K., Schulenburg, J-M. Graf v. d., Greiner, W. (2018):

German Value Set for the EQ-5D-5L. *Pharmacoeconomics*, 36(6): 663-674.

Modul 6

Stolk, E., **Ludwig, K.**, Rand, K., van Hout, B., Ramos Goñi, JM.: Overview, update and lessons learned from the international EQ-5D-5L valuation work: version 2 of the EQ-5D-5L valuation protocol. *Value in Health*. Akzeptiert zur Publikation.