

Aus dem  
Deutschen Krebsforschungszentrum  
(Vorstandsvorsitzender: Prof. Dr. med. Michael Baumann)

Welche Faktoren  
– unter besonderer Berücksichtigung der Kostenfolgen („Budget Impact“) –  
beeinflussen die Ergebnisse von offiziellen Health Technology Assessments  
(HTAs) in Deutschland und England?  
Eine multivariate Analyse

Inauguraldissertation  
zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)  
der  
Medizinischen Fakultät Mannheim  
der Ruprecht-Karls-Universität  
zu  
Heidelberg

vorgelegt von  
Ramon Schäfer

aus  
Pforzheim

2022

Dekan: Prof. Dr. med. Sergij Goerd  
Referent: Prof. Dr. med. Michael Schlander

# INHALTSVERZEICHNIS

ABBILDUNGSVERZEICHNIS .....	III
TABELLENVERZEICHNIS .....	IV
ABKÜRZUNGSVERZEICHNIS .....	V
<b>1 EINFÜHRUNG .....</b>	<b>1</b>
1.1 Ausgangssituation und Relevanz des Themas .....	1
1.2 Definition und Abgrenzung zentraler Begrifflichkeiten .....	2
1.2.1 Health Technology Assessment (HTA) versus Evidenzbasierte Medizin .....	3
1.2.2 Kosteneffektivitäts-Analyse versus Budget Impact-Analyse (BIA) .....	4
<b>2 HEALTH TECHNOLOGY ASSESSMENT .....</b>	<b>7</b>
2.1 Historische Entwicklung und gegenwärtige Bedeutung .....	7
2.2 Bewertungsprozess und Ziele .....	9
2.3 Nationale HTA-Institutionen und Evaluationsmethoden im Vergleich .....	11
2.3.1 Gemeinsamer Bundesausschuss (GBA) und Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) in Deutschland .....	12
2.3.2 National Institute for Health and Care Excellence (NICE) in England .....	18
2.3.3 Haute Autorité de Santé (HAS) in Frankreich .....	24
2.4 BIAs als methodischer Bestandteil von HTA? Eine Bestandsaufnahme mit gesondertem Fokus auf Deutschland und England .....	27
<b>3 ZIELSETZUNG UND AUFBAU .....</b>	<b>30</b>
<b>4 DATEN UND METHODEN .....</b>	<b>32</b>
4.1 Quellen und Datenerfassung .....	32
4.1.1 Deutschland: GBA und IQWiG .....	32
4.1.2 England: NICE .....	34
4.1.3 Frankreich: HAS .....	37
4.2 Vergleichende Untersuchung der HTA-Resultate .....	39
4.3 Regressionsmodelle zur Analyse der HTA-Kriterien und -Resultate .....	42
4.3.1 GBA-Beschlussfassungen und IQWiG-Dossierbewertungen .....	42
4.3.2 NICE Single Technology Appraisal (STA) Guidance .....	46

5	ERGEBNISSE .....	51
5.1	Ableich der identifizierten HTAs von GBA, HAS und NICE .....	51
5.1.1	Interventionspaare im Vergleich: Eine Übersicht .....	56
5.1.2	Indikationsgebiete im Vergleich: Onkologische versus nicht-onkologische Therapien .....	57
5.2	Analyse der HTA-Kriterien und GBA-Beschlüsse / IQWiG-Bewertungen.....	60
5.2.1	Deskriptive Untersuchungsergebnisse für GBA und IQWiG.....	61
5.2.2	Ergebnisse der Regressionsanalysen für GBA und IQWiG .....	75
5.2.3	GBA-Beschlussfassungen versus IQWiG-Dossierbewertungen .....	81
5.3	Analyse der STA-Kriterien und NICE Guidance .....	83
5.3.1	Deskriptive Untersuchungsergebnisse für NICE.....	83
5.3.2	Deskriptive Untersuchungsergebnisse für NICE unter Berücksichtigung des Cancer Drugs Fund (CDF) .....	93
5.3.3	Resultate der Regressionsanalyse für NICE.....	100
5.3.4	Resultate der Regressionsanalyse für NICE unter Berücksichtigung des CDF .....	105
6	DISKUSSION .....	112
6.1	Herausstellung und kritische Reflexion der wesentlichen Erkenntnisse .....	112
6.1.1	Vergleichende Untersuchung von GBA, HAS und NICE.....	112
6.1.2	GBA-Beschlussfassungen und IQWiG-Dossierbewertungen .....	117
6.1.3	NICE STA Guidance .....	122
6.2	Die (potenzielle) Rolle des Budget Impact bei GBA/IQWiG und NICE .....	128
6.2.1	Die ökonomische Komponente in der frühen (Zusatz-)Nutzenbewertung...	128
6.2.2	STAs und die Kosteneffektivität als (primärer) Bewertungsfaktor .....	132
6.3	Nationale HTA-Institutionen und -Verfahren – quo vadis? .....	136
6.4	Limitationen und weiterer Forschungsbedarf.....	143
7	ZUSAMMENFASSUNG .....	147
8	LITERATURVERZEICHNIS.....	150
9	VERÖFFENTLICHUNGEN.....	178
10	ANHANG.....	179
11	LEBENS LAUF.....	198
12	DANKSAGUNG.....	199

## ABBILDUNGSVERZEICHNIS

Abbildung 1: Health Technology Assessment (HTA) in der Praxis – mit Fokus auf Deutschland und England.....	4
Abbildung 2: Vereinfachter Verfahrensablauf der Frühen (Zusatz-)Nutzenbewertung nach AMNOG .....	14
Abbildung 3: Single Technology Appraisal – zusammengefasste Prozessdarstellung.....	20
Abbildung 4: Übersicht zu HTA, Preisbildung und Erstattung im französischen Gesundheitssystem .....	25
Abbildung 5: Auswahl der übereinstimmenden HTAs von GBA, HAS und NICE (Januar 2011 – Juni 2018) .....	51
Abbildung 6: Paarweiser Vergleich der Übereinstimmung der HTA-Outcomes von GBA, HAS und NICE .....	56
Abbildung 7: Positive HTA-Outcomes von GBA, HAS und NICE .....	58
Abbildung 8: Budget Impact-Szenarien – GBA-Beschlüsse auf Teilpopulationsebene .....	66
Abbildung 9: Inkrementelle Budget Impact-Szenarien – GBA-Beschlüsse auf Teilpopulationsebene .....	67
Abbildung 10: Budget Impact-Szenarien – IQWiG-Bewertungen und pU-Dossiers auf Teilpopulationsebene .....	68
Abbildung 11: Inkrementelle Budget Impact-Szenarien – IQWiG-Bewertungen auf Teilpopulationsebene .....	68
Abbildung 12: [Inkrementelle] Budget Impact-Szenarien – GBA-Beschlüsse auf Verfahrensebene.....	74
Abbildung 13: [Inkrementelle] Budget Impact-Szenarien – IQWiG-Bewertungen auf Verfahrensebene.....	74
Abbildung 14: GBA-Beschlüsse versus IQWiG-Bewertungen auf Teilpopulationsebene.....	82
Abbildung 15: STA-Empfehlungen zugeordnet nach ICER-Schwellenwerten (Subgruppenebene).....	87
Abbildung 16: Budget Impact-Szenarien für NICE auf Subgruppenebene .....	88
Abbildung 17: Inkrementelle Budget Impact-Szenarien für NICE auf Subgruppenebene .....	89
Abbildung 18: [Inkrementelle] Budget Impact-Szenarien für NICE auf Indikationsebene.....	92
Abbildung 19: Budget Impact-Szenarien für NICE auf Subgruppenebene – inklusive CDF..	96
Abbildung 20: Inkrementelle Budget Impact-Szenarien für NICE auf Subgruppenebene – inklusive CDF .....	97
Abbildung 21: [Inkrementelle] Budget Impact-Szenarien für NICE auf Indikationsebene – inklusive CDF .....	100

**TABELLENVERZEICHNIS**

Tabelle 1: Budget Impact-Analyse versus Kosteneffektivitäts-Analyse.....	6
Tabelle 2: Definition der Regressionsvariablen (1) für GBA und IQWiG .....	43
Tabelle 3: Definition der Regressionsvariablen (2) – Budget Impact-Szenarien für GBA und IQWiG.....	44
Tabelle 4: Definition der Regressionsvariablen (1) für NICE .....	47
Tabelle 5: Definition der Regressionsvariablen (2) – Budget Impact-Szenarien für NICE...	48
Tabelle 6: Abgleich der identifizierten HTAs von GBA, HAS und NICE (Januar 2011 – Juni 2018) .....	52
Tabelle 7: Übereinstimmende HTAs nach therapeutischem Gebiet .....	59
Tabelle 8: Deskriptive Ergebnisse für GBA auf Teilpopulationsebene .....	62
Tabelle 9: Deskriptive Ergebnisse für IQWiG auf Teilpopulationsebene .....	64
Tabelle 10: Deskriptive Ergebnisse für GBA auf Verfahrensebene .....	70
Tabelle 11: Deskriptive Ergebnisse für IQWiG auf Verfahrensebene.....	72
Tabelle 12: Resultate der Regressionsanalyse für GBA auf Teilpopulationsebene.....	76
Tabelle 13: Resultate der Regressionsanalyse für IQWiG auf Teilpopulationsebene .....	77
Tabelle 14: Resultate der Regressionsanalyse für GBA auf Verfahrensebene .....	79
Tabelle 15: Resultate der Regressionsanalyse für IQWiG auf Verfahrensebene.....	80
Tabelle 16: Deskriptive Ergebnisse für NICE auf Subgruppenebene.....	84
Tabelle 17: Deskriptive Ergebnisse für NICE auf Indikationsebene .....	90
Tabelle 18: Deskriptive Ergebnisse für NICE auf Subgruppenebene – inklusive CDF .....	94
Tabelle 19: Deskriptive Ergebnisse für NICE auf Indikationsebene – inklusive CDF.....	98
Tabelle 20: Resultate der Regressionsanalyse für NICE auf Subgruppenebene.....	102
Tabelle 21: Resultate der Regressionsanalyse für NICE auf Indikationsebene .....	105
Tabelle 22: Resultate der Regressionsanalyse für NICE auf Subgruppenebene – inklusive CDF .....	108
Tabelle 23: Resultate der Regressionsanalyse für NICE auf Indikationsebene – inklusive CDF .....	110

## ABKÜRZUNGSVERZEICHNIS

ACT	Avis de la Commission de la Transparence
AMNOG	Arzneimittelmarktneuordnungsgesetz
AM-NutzenV	Arzneimittel-Nutzenbewertungsverordnung
ANAES	Agence Nationale d'Accreditation et d'Evaluation en Santé
ANDEM	Agence Nationale de l'Evaluation Medicale
ASMR	Amélioration du Service Médical Rendu
BI	Budget Impact [Kostenfolgen]
BIA	Budget Impact Analysis [Budget Impact-/ Kostenfolgen-Analyse]
BMG	Bundesministerium für Gesundheit
CBA	Cost-Benefit Analysis [Kosten-Nutzen-Analyse]
CDF	Cancer Drugs Fund
CEA	Cost-Effectiveness Analysis [Kosteneffektivitäts-Analyse]
CEESP	Commission Evaluation Economique et de Santé Publique
CEPS	Comité Economique des Produits de Santé
CHTE	Centre for Health Technology Evaluation
CI	Confidence Interval [Konfidenzintervall]
CT	Commission de la Transparence
CUA	Cost-Utility Analysis [Kosten-Nutzwert-Analyse]
DAHTA	Deutsche Agentur für Health Technology Assessment
DH	Department of Health and Social Care
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information
EbM	Evidenzbasierte Medizin
EMA	European Medicines Agency
EoL	End-of-Life

---

ERG	Evidence Review Group
EU	Europäische Union
EUR	Euro
FAD	Final Appraisal Determination
FNB	Frühe (Zusatz-)Nutzenbewertung
FTA	Fast Track Appraisal
GBA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
GKV-SV	Spitzenverband Bund der Krankenkassen
GBP	British Pound Sterling [Pfund Sterling]
HAS	Haute Autorité de Santé
HL	Hodgkin-Lymphom
HRQoL	Health-Related Quality of Life [Gesundheitsbezogene Lebensqualität]
HST	Highly Specialised Technology
HTA	Health Technology Assessment
ICER	Incremental Cost-Effectiveness Ratio [Inkrementelles Kosteneffektivitäts-Verhältnis]
ISPOR	International Society for Pharmacoeconomics and Outcomes Research
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen
MTA	Multiple Technology Appraisal
NHL	Non-Hodgkin-Lymphom
NHS	National Health Service
NICE	National Institute for Health and Care Excellence
OD	Orphan Drug [Arzneimittel für Seltene Erkrankungen]
pU	pharmazeutisches Unternehmen
QALY	Quality-Adjusted Life Year [Qualitätsadjustiertes Lebensjahr]
RCT	Randomized Controlled Trial [Randomisierte kontrollierte Studie]
RVV	Retinaler Venenverschluss



---

SA	Standardabweichung
SGB	Sozialgesetzbuch
SMR	Service Médical Rendu
STA	Single Technology Appraisal
TA	Technology Appraisal
UNCAM	Union Nationale des Caisses d'Assurance Maladie
URD	Ultra-Rare Disease [Sehr Seltene Erkrankung]
VAV	Venenastverschluss
WHO	World Health Organization
ZVT	Zweckmäßige Vergleichstherapie

# 1 EINFÜHRUNG

## 1.1 Ausgangssituation und Relevanz des Themas

Nationale Gesundheitssysteme stehen vor der immer größer werdenden Herausforderung, den Zugang zu innovativen Therapien sowie deren nachhaltige Finanzierung sicherzustellen. Infolgedessen ist die Einführung von sogenannten Health Technology Assessments (HTAs) zur umfassenden Bewertung medizinischer Interventionen auf das Ziel einer möglichst effektiven und effizienten Nutzung von neuen Gesundheitstechnologien zurückzuführen. Verantwortlich für die systematische Durchführung von HTAs sind dabei in der Regel nationale HTA-Institutionen, die die versorgungspraktische Entscheidungsfindung im Gesundheitssektor unterstützen (Banta, 2003; Drummond et al., 2008; Sorenson and Chalkidou, 2012). Angesichts der stetig steigenden Anzahl neuer medizinischer Interventionen einerseits und der Knappheit der zur Verfügung stehenden finanziellen Mittel andererseits erscheint es jedoch nicht weiter überraschend, dass deren Implementierung bis heute von kontrovers diskutierten Themen wie einer möglichen Rationierung der verfügbaren Ressourcen begleitet wird.

Im Fokus stehen insbesondere die methodischen Vorgehensweisen der verantwortlichen Einrichtungen, die im Rahmen ihres rechtlichen und institutionellen Kontexts sowie innerhalb der bestehenden Versorgungsstruktur zu teils erheblichen Unterschieden in den HTA-Empfehlungen führen können. Verstärkt werden kann die Heterogenität von Bewertungsergebnissen durch länderspezifische Regulierungen, die die unterschiedlichen gesundheitspolitischen Priorisierungen und Zielsetzungen widerspiegeln (Akehurst et al., 2017; Angelis et al., 2018; Nicod and Kanavos, 2012). In seiner praktischen Umsetzung stützt sich HTA dabei vorrangig auf zwei Säulen: Einerseits auf die Grundsätze der Evidenzbasierten Medizin (EbM) zur Bewertung der klinischen Wirksamkeit und andererseits auf eine Bewertung des Kosten-Nutzen-Verhältnisses, welches im Beispiel Deutschlands implizit und im Beispiel Großbritanniens explizit – unter Rekurs auf die Logik der Kosteneffektivität – erfolgt.

Vor diesem Hintergrund konzentriert sich die vorliegende Arbeit auf zwei international anerkannte HTA-Institutionen, die diese unterschiedlichen methodischen Ansätze exemplifizieren. Auf der einen Seite werden der Gemeinsame Bundesausschuss (GBA) und das Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) in Deutschland betrachtet, deren zweistufiges Bewertungsverfahren sich seit Einführung des Arzneimittelmarktneuordnungsgesetzes (AMNOG) im Jahr 2011 auf eine strikte Umsetzung der Prinzipien der EbM stützt. Dabei wird eine Kosten-Nutzen-Bewertung mit dem zusätzlichen Element der Kostenfolgen-Analyse

(Budget Impact Analysis, BIA) allenfalls optional durchgeführt (G-BA, 2020). Auf der anderen Seite wird das National Institute for Health and Care Excellence (NICE) in England und Wales herangezogen, das bei der Bewertung von Gesundheitstechnologien sowohl die klinische Wirksamkeit als auch die Kosteneffektivität einbezieht. Empfehlungen durch NICE begründen sich dabei primär auf dem Nutzenmaß der sogenannten qualitätsadjustierten Lebensjahre (Quality-Adjusted Life Years, QALYs). Neben den bestehenden Leitlinien für eine gesamtheitliche Betrachtung aller innerhalb des englischen National Health Service (NHS<sup>1</sup>) für eine Technologie anfallenden Kosten führt NICE eine Budget Impact (BI)-Projektion durch, die die finanzielle Auswirkung neuer Technologien als Hilfe für die Implementierung der evidenzbasierten Empfehlungen durch NICE – der sogenannten NICE Guidance<sup>2</sup> – auf regionaler Ebene vorhersagen soll (NICE, 2018).

## 1.2 Definition und Abgrenzung zentraler Begrifflichkeiten

Bevor im Anschluss an den theoretischen Hintergrund die Forschungsfrage der Dissertation abgeleitet und die daraus resultierenden Zielsetzungen definiert werden, gilt es zuerst folgende Begrifflichkeiten voneinander abzugrenzen, welche im weiteren Fortgang der Arbeit von besonderer Relevanz sind: Einerseits HTA und EbM, deren kontextabhängige Bedeutung eindeutig darzulegen ist und andererseits die inhaltliche Differenzierung der Kosteneffektivitäts-Analyse (Cost-Effectiveness Analysis, CEA) und der BIA.

---

<sup>1</sup> Das britische NHS unterteilt sich in vier weitgehend unabhängige Organisationen für England, Wales, Schottland und Nordirland. Das NHS England umfasst dabei die Dienstleistungen des englischen Gesundheitswesens.

<sup>2</sup> NICE ist nach vorliegender Gesetzgebung nur innerhalb des NHS England zuständig. Darüber hinaus gelten Vereinbarungen, dass NICE medizinische Produkte und Services für Wales, in Sonderfällen auch für Schottland und Nordirland bereitstellen kann, wobei die beiden letzteren über unabhängige HTA-Institutionen und -Verfahren verfügen (NICE, 2021). Im weiteren Verlauf der Arbeit wird sich demnach auf England und Wales (symbolisch unter Verwendung der Nationalflagge des Vereinigten Königreichs) bezogen, auch wenn begrifflich ausschließlich England oder das NHS England benannt werden.

### 1.2.1 *Health Technology Assessment (HTA) versus Evidenzbasierte Medizin*

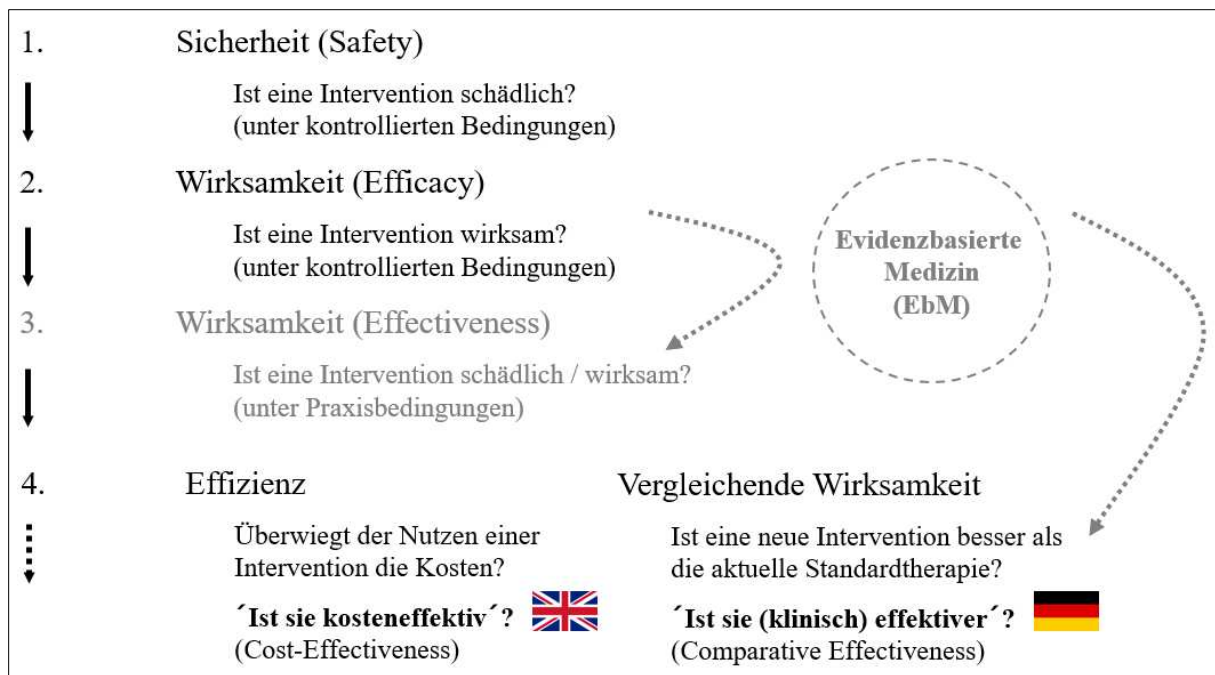
HTA<sup>3</sup> wird allgemein hin als ein multidisziplinärer Prozess zur Unterstützung der Entscheidungsfindung im Gesundheitswesen verstanden, der auf wissenschaftlichen und nicht-wissenschaftlichen Erkenntnissen basiert. HTA stellt Informationen in einer systematischen, transparenten und robusten Weise zur Verfügung, die als Grundlage für Entscheidungen über den Nutzen sowie die Diffusion einer Gesundheitstechnologie (zu verschiedenen Zeitpunkten innerhalb ihres Lebenszyklus) und eine möglichst effiziente Allokation der zur Verfügung stehenden Ressourcen unterstützt (Banta, 2009; Battista and Hodge, 1999; Drummond et al., 2008; O'Rourke et al., 2020).

Zwei wesentliche Elemente von HTA stellen dabei die evidenzbasierte Bewertung der klinischen Effektivität beziehungsweise Wirksamkeit und die Berücksichtigung ökonomischer Implikationen dar, letzteres für gewöhnlich durch die Bestimmung des Verhältnisses von Kosten und Nutzen (Value for Money). In Abbildung 1 sind die inhaltlich aufeinanderfolgenden Schritte inklusive der zugehörigen Schlüsselfragen dargestellt, die erforderlich sind, bevor daran anschließend ein HTA durchgeführt werden kann (siehe unter 4.), um den Nutzen oder vielmehr den 'Value for Money' einer Intervention adressieren zu können (vgl. hierzu auch Kapitel 2.2).

Darüber hinaus gibt es weitere Dimensionen wie sozial-ethische, rechtlich-politische, institutionelle und umweltbezogene Aspekte, die in den HTA-Prozess einfließen (sollten). Das Ziel von HTA besteht darin, eine gerechte, qualitativ hochwertige und zugleich effiziente Gesundheitspolitik zu unterstützen, die sich am Wohl der Patientinnen und Patienten orientiert und damit den bestmöglichen Nutzen respektive Nutzwert anstrebt (Drummond et al., 2008; Henshall and Schuller, 2013; O'Rourke et al., 2020). Dabei handelt es sich um keine einheitliche Methodik, sondern vielmehr um einen ganzheitlichen Ansatz zur systematischen Unterstützung von (politischen) Entscheidungen im Gesundheitswesen (Lühmann et al., 2014).

---

<sup>3</sup> Für HTA hat sich die folgende, konsensbasierte Definition weitgehend etabliert: „*Health technology assessment is a multidisciplinary process that uses explicit methods to determine the value of a health technology at different points in its lifecycle. The purpose is to inform decision-making in order to promote an equitable, efficient, and high-quality health system.*“ (O'Rourke et al., 2020)



**Abbildung 1:** Health Technology Assessment (HTA) in der Praxis – mit Fokus auf Deutschland und England (Eigene Darstellung; Quelle: Luce et al., 2010)

Der Begriff EbM lässt sich nach Sackett et al. (1999) als die Verwendung der bestmöglichen wissenschaftlichen Evidenz bei klinischen Entscheidungen über die Versorgung einzelner Patientinnen und Patienten zusammenfassen. Folglich stützt sich die EbM auf eine Zusammenführung von individuell klinischem Fachwissen und der aktuell besten verfügbaren wissenschaftlichen Erkenntnisse basierend auf systematischer Forschungsarbeit (Eddy, 2005; Sackett et al., 1999).

### 1.2.2 *Kosteneffektivitäts-Analyse versus Budget Impact-Analyse (BIA)*

Sowohl die Kosteneffektivitäts-Analyse als auch die Kostenfolgen-Analyse stellen (gesundheits-)ökonomische Evaluationsformen dar, die sich in ihrer Vorgehensweise jedoch deutlich unterscheiden und zudem verschiedene Outcomes untersuchen. Die Kosteneffektivität beschreibt dabei die Kostenwirksamkeit einer Intervention gegenüber einer Vergleichsintervention. Die Kostenfolgen stellen demgegenüber die Gesamtkosten beziehungsweise budgetwirksamen Kosten einer Intervention innerhalb des gegenwärtigen Versorgungsrahmens dar (Mauskopf, 1998).

CEAs lassen sich als vergleichende Analysen der Kosten zweier (oder mehrerer) Behandlungsalternativen im Hinblick auf einen definierten Outcome definieren. Ihr Ziel ist es, die Behandlungsalternative zu finden, die pro Outcome-Einheit die geringsten Kosten erzeugt und folglich als wirksamer respektive effektiver bezeichnet werden kann (Drummond et al., 2015; Neumann et al., 2016; Weinstein and Stason, 1977). Im gesundheitsökonomischen Kontext steht allen voran die Kosten-Nutzwert-Analyse (Cost-Utility Analysis, CUA) im Vordergrund, die eine spezifische Adaption der CEA darstellt. Bei der CUA wird auf Outcome-Seite die Wirksamkeit in Form eines generischen Nutzenwerts einbezogen.

Am häufigsten wird dabei das Konzept der QALYs<sup>4</sup> verwendet, welches als ein Maß der klinischen Effektivität zur Berechnung des inkrementellen Kosteneffektivitäts-Verhältnisses (Incremental Cost-Effectiveness Ratio, ICER) dient. Es stellt demnach die Relation einer Kostenveränderung zu den QALYs einer neuen Gesundheitstechnologie im Vergleich zu einer (bestmöglichen und zugleich) verfügbaren Therapiealternative dar (Drummond et al., 2015; Neumann et al., 2016). In einigen HTA-Prozessen nationaler Gesundheitssysteme, wie beispielsweise auch innerhalb des englischen NHS, wird die inkrementelle Kosteneffektivitäts-Relation als ein offizieller Entscheidungsfaktor berücksichtigt (NICE, 2018).

Im Vergleich zur CEA ist das Ziel der BIA, die sich durch Aufnahme (oder Entfernen) einer Intervention ergebenden finanziellen Konsequenzen innerhalb einer Versorgungsstruktur abzuschätzen. Im Mittelpunkt steht folglich die Bewertung der Bezahlbarkeit (oder auch Erschwinglichkeit, im Englischen 'Affordability') einer Intervention für ein Gesundheitssystem, welche aus Perspektive der Ausgaben- und Entscheidungsverantwortlichen für einen möglichst adäquaten Zeithorizont getroffen werden sollte. Die BIA berücksichtigt in der Regel ausschließlich direkte Gesundheitskosten, die in Zusammenhang mit der zu evaluierenden Technologie stehen (Mauskopf et al., 2007; Sullivan et al., 2014; Trueman et al., 2001). In nachstehender Tabelle 1 sind die wesentlichen Gesichtspunkte der BIA und der CEA nochmals differenziert gegenübergestellt.

---

<sup>4</sup> Das Konzept der QALYs geht auf eine Arbeit von Klarman et al. (1968) zurück, die in einem Index gewonnene Lebensjahre und Veränderungen der Lebensqualität zusammenführten. Für die Bestimmung der QALYs benötigt man die Wirkung einer Behandlung in Bezug auf die Lebenszeitverlängerung als eine quantitative und den Lebensqualitätseffekt als eine qualitative Dimension. Dabei werden die Nutzwerte zu unterschiedlichen Gesundheitszuständen aus Befragungen gewonnen, die auf einer normierten Skala von 0 (Gesundheitszustand Tod) bis 1 (Zustand perfekter Gesundheit) entsprechend zugeordnet werden. Für die Qualitätsanpassung wird die erwartete Dauer der potenziellen Gesundheitszustände mit den entsprechenden Gewichten multipliziert. Aus der anschließenden Addition der Produkte ergeben sich dann die QALYs, die als generische Ergebnisparameter indikationsübergreifend verwendet werden können (vgl. unter anderem bei Drummond et al., 2015; Neumann et al., 2016).

**Tabelle 1:** Budget Impact-Analyse versus Kosteneffektivitäts-Analyse  
(Eigene Darstellung; in Anlehnung an: Garattini and van de Vooren, 2011; Neyt et al., 2015)

	Budget Impact-Analyse	Kosteneffektivitäts-Analyse
Fokus	Kostenfolgen (Bezahlbarkeit)	Kostenwirksamkeit/ -effektivität
Perspektive	Kostenträger	Kostenträger (oder Patienten, Gesellschaft, System)
Zeithorizont	Kurz- bis mittelfristig (1–5 Jahre)	Kurz- bis langfristig (variabel)
Zielpopulation	Abhängig von Erstattungsanforderung	Abhängig von Erstattungsanforderung
Vergleich	Vorhandene Behandlungsalternativen	Vergleichstherapie(n)
Kosten	Direkte (gesundheitsbezogene) Kosten	Direkte (gesundheitsbezogene) Kosten (plus andere direkte / indirekte Kosten)
Effektivität	Kurzfristige Kosten; Einsparungen (bei den Gesundheitsleistungen)	Natürliche Einheiten; Lebensqualität (Kostenwirksamkeits-Analyse)
Ergebnis(darstellung)	(Inkrementelle) Kostenfolgen	Kosten / Kosteneffektivitäts-Verhältnis (inkrementell)
Diskontierung (Kosten)	Keine Diskontierung	Diskontierung nach Zeithorizont
Sensitivitätsanalyse	Extremwerte; Szenarioanalyse	Verschiedene Formen (einseitig, mehr- seitig, probabilistisch, Schwellenwerte)

## 2 HEALTH TECHNOLOGY ASSESSMENT

### 2.1 Historische Entwicklung und gegenwärtige Bedeutung

Der Ursprung von HTA ist eng mit der Gründung des US Congress Office of Technology Assessment (OTA<sup>5</sup>) verknüpft, welches bereits im Jahr 1976 einen ersten HTA-Bericht veröffentlichte. Auch wenn der Fokus zu Beginn primär auf sozialen beziehungsweise gesellschaftlichen Implikationen der betrachteten Gesundheitstechnologie lag, stellt der Report die erste systematische Konzeptualisierung einer solchen Bewertung dar (Banta, 2003; vgl. hierzu auch Schäfer, 2015).

In den darauffolgenden Jahren kam es durch die Implementierung nationaler HTA-Programme zur sukzessiven Institutionalisierung von HTA, insbesondere auch im europäischen Raum. Das Swedish Council on Technology Assessment in Health Care (SBU), welches 1987 als unabhängige, staatliche Behörde entstand, gilt weltweit als eines der ältesten HTA-Institutionen und ist bis heute für die Evaluation neuer Interventionen in Schweden verantwortlich. Parallel begann mit der 1985 gegründeten International Society of Technology Assessment in Health Care (ISTAHC<sup>6</sup>) und dem 1993 etablierten International Network of Agencies for Health Technology Assessment (INAHTA<sup>7</sup>) die Förderung internationaler Kollaborationen und Partnerschaften der weltweit aufkommenden HTA-Aktivitäten (vgl. Banta, 2003; Banta and Jonsson, 2009).

Mit der Gründung zahlreicher nationaler HTA-Institutionen und einer zugleich global wachsenden HTA-Gemeinschaft stehen seither sowohl die Standardisierung der verschiedenen HTA-Aktivitäten als auch die methodische Weiterentwicklung verstärkt im Vordergrund. Als ein wesentlicher Baustein gilt hierbei die Verbreitung systematischer Übersichtsarbeiten (Systematic Reviews) und qualitativ hochwertiger Forschungsberichte, die maßgeblich durch die Entstehung der Cochrane Collaboration<sup>8</sup> gefördert wurden. Der Bedarf nach vergleichbaren und verlässlichen HTA-Reports führte darüber hinaus zu einer engeren Zusammenarbeit nationaler

---

<sup>5</sup> Das US-amerikanische OTA wurde bereits im Jahr 1995 geschlossen.

<sup>6</sup> ISTAHC war eine interdisziplinäre wissenschaftliche Fachgesellschaft, die im Jahr 2003 unter dem Namen Health Technology Assessment international (HTAi) neu gegründet wurde (HTAi, 2021).

<sup>7</sup> Das INAHTA-Netzwerk verbindet eine Vielzahl nationaler HTA-Institutionen, um Kooperation und Informationsaustausch zur Erstellung und Verbreitung von HTA-Berichten weltweit zu ermöglichen (INAHTA, 2020).

<sup>8</sup> Cochrane wurde 1993 gegründet und ist ein globales, unabhängiges Netzwerk von Wissenschaft, Medizin sowie der Vertretung von Patientinnen und Patienten. Die daraus als internationale, gemeinnützige Organisation entstandene Cochrane Collaboration liefert durch die Erstellung, Aktualisierung und Dissemination systematischer Reviews evidenzbasierte Berichte und Informationsquellen als Entscheidungsunterstützung im Gesundheitswesen (Cochrane, 2021).



HTA-Institutionen. Daraus resultierten Annäherungen unterschiedlicher Methoden (zum Beispiel der EbM zur Bewertung klinischer Effektivität) sowie die Erstellung strukturierter HTA-Berichte angelehnt an die Vorgehensweise systematischer Reviews oder (teilweise auch) Meta-Analysen (Drummond et al., 2008; Lühmann et al., 2014; vgl. hierzu auch Schäfer, 2015).

Seit 1993 fördert zudem auch die Europäische Kommission die Kollaboration nationaler HTA-Träger durch länderübergreifende Projekte<sup>9</sup>. Zentrale Elemente stellen dabei die Verbesserung und Harmonisierung der Bewertungsgrundlagen und HTA-Methoden sowie eine effektivere Verbreitung von HTA-Resultaten dar (Banta and Jonsson, 2009). Als eines der bekanntesten Vorhaben gilt das 2005 gestartete Projekt European Network Health Technology Assessment (EUnetHTA<sup>10</sup>), das die genannten Kernpunkte aufgreift und sich bis heute mitunter verstärkt auf eine effektivere und nachhaltigere Vernetzung nationaler HTA-Entwicklungen bezieht. Mit EUnetHTA 21 (2021-2023) liegt der aktuelle Schwerpunkt der Projektarbeit auf einer fundierten Unterstützung eines zukünftigen 'EU HTA System' unter Einbindung der bestehenden nationalen HTA-Regulierungen (EUnetHTA, 2021). Darüber hinaus kündigte die Europäische Kommission in ihrem 'Arbeitsprogramm 2017' an, eine Initiative zur Stärkung der EU-weiten Kooperation bei der Bewertung von Gesundheitstechnologien einzuleiten, um den Binnenmarkt für medizinische Produkte zu verbessern (EU-Kommission, 2018). Im September 2021 ist die Umsetzung der angekündigten Initiative eingeleitet worden, deren für 2029 geplanter Abschluss und eine damit (mögliche) Implementierung von (zumindest einzelnen) HTA-Prozesselementen folglich abzuwarten bleibt (EU-Kommission, 2021).

Seit seiner Entstehung hat HTA als ein Instrument zur Unterstützung der gesundheitspolitischen Entscheidungsfindung sukzessive an Bedeutung gewonnen. Dies ist sowohl auf politischer als auch auf wissenschaftlicher Ebene unverkennbar, da die internationale Zusammenarbeit zwischen einzelnen Institutionen und Programmen sowie auch die nachhaltige Etablierung weltweiter Netzwerke einen festen Platz in der HTA-Praxis einnehmen.

---

<sup>9</sup> Dazu zählen Projekte wie das EUR-ASSESS-Project (1994-1997) und Kooperationen wie HTA-Europe (1997-1998) oder The European Collaboration for Assessment of Health and Interventions and Technology (ECHTA/ECAHI im Jahr 2000; vgl. hierzu bei Banta and Jonsson, 2009).

<sup>10</sup> Seit 2005 existiert EUnetHTA als ein europäisches Netzwerk, dessen HTA-Aktivitäten durch die Gründung der EUnetHTA Collaboration (2009), der EUnetHTA Joint Action 1 (2010-2012), EUnetHTA Joint Action 2 (2012-2015) und EUnetHTA Joint Action 3 (2016-2021) organisiert wurden (EUnetHTA, 2021).

## 2.2 Bewertungsprozess und Ziele

HTA als Bewertungsprozess umfasst keine einheitliche Methodik, die in einer festgelegten Form angewandt wird und infolgedessen ohne weiteres von einem Gesundheitssystem auf ein anderes übertragbar ist. Dennoch haben sich für den Ablauf von HTA grundsätzliche Vorgehensschritte und allgemeingültige Leitlinien durchgesetzt und im Laufe der letzten beiden Jahrzehnte international etabliert. Die HTA-Methodik bedient sich dabei standardisierter wissenschaftlicher Verfahren, deren Schwerpunkt auf der systematischen Analyse und Synthese der einschlägigen Literatur und von klinischen Studien liegt. Die Einhaltung einer durchgängig strukturierten Vorgehensweise sowie eine übersichtliche Darstellung und Verbreitung der Ergebnisse in (einer gegliederten) Berichtsform gelten darüber hinaus als wichtige Voraussetzungen (Drummond et al., 2008; Zentner et al., 2014; vgl. hierzu auch Schäfer, 2015).

Um den grundlegenden Bewertungsprozess<sup>11</sup> einer Gesundheitstechnologie vereinfacht darstellen zu können, wird HTA in der Regel in drei Phasen gegliedert (Zentner et al., 2014): [1] Die Identifizierung und Priorisierung der zu bewertenden Technologien, [2] die Durchführung von HTA und die [3] Dissemination und Implementierung der Ergebnisse (in die Praxis).

Die Auswahl der zu bewertenden Technologien bildet den ersten Schritt dieses Prozesses. Das zentrale Element stellt die Priorisierung zwischen den zur Verfügung stehenden Technologien dar. Die Priorisierung von HTA-Themen unterscheidet sich ausgehend von den jeweiligen Gegebenheiten in den einzelnen Ländern und den zuständigen HTA-Institutionen. Zudem gibt es unterschiedliche Ansätze<sup>12</sup>, um identifizierte Technologien zu priorisieren, die wiederum als Grundlage für die Entscheidungsfindung der HTA-Themen dienen (Perleth et al., 2014).

Mit dem festgelegten Thema kann in einem zweiten Schritt die systematische Durchführung eines HTAs stattfinden. In seiner praktischen Umsetzung lässt sich dieser Prozess in die Bewertung (Assessment) und die Beurteilung der Bewertung (Appraisal) unterteilen, wobei die Verantwortung bei einer oder mehreren Institutionen liegen kann. Während das Assessment einen wissenschaftlich-analytischen Prozess auf Grundlage der zur Verfügung stehenden Evi-

---

<sup>11</sup> Es handelt sich dabei um den grundsätzlichen Ablauf von systematischen HTAs, weshalb eine Bewertung nicht zwangsläufig jeden einzelnen Schritt des Prozesses (vollumfänglich) durchlaufen muss.

<sup>12</sup> Die Priorisierung von HTA-Themen erfolgt entweder durch die Erstellung von Kandidatenlisten oder mithilfe spezieller Instrumente für die Prioritätensetzung. Bei ersterem handelt es sich um Strategien zur Suche und Filterung von Technologien, wie zum Beispiel durch Expertenbefragungen oder ein Horizon Scanning System (systematisches Frühwarnsystem). Letzteres beschreibt Modelle für eine auf objektiven Kriterien gestützte Prioritätensetzung, wie beispielsweise medizinisch-epidemiologische oder ökonomische Ansätze (Perleth et al., 2014).

denz beschreibt, werden beim Appraisal neben den Informationen der Bewertung weitere Quellen hinzugenommen, zum Beispiel klinische Fachleute oder auch Patientengruppen<sup>13</sup>. Entsprechend können sich die Evidenzanforderungen, analytischen Rahmenbedingungen und Bewertungsmethoden in der Durchführungsphase, insbesondere beim Assessment, grundsätzlich unterscheiden. Jedes HTA-System erfordert bestimmte Varianten und Qualitätskriterien der Evidenz für die klinische und ökonomische Evaluierung (Zentner et al., 2014). Aus diesem Grund unterstreicht eine Mehrheit der HTA-Institutionen die Bedeutung systematischer Reviews (oder Meta-Analysen), die, wenn möglich, auf methodisch hochwertigen, randomisiert kontrollierten Studien (Randomized-Controlled Trials, RCTs<sup>14</sup>) basieren sollten (Kleijnen et al., 2012). Im Regelfall haben die pharmazeutischen Unternehmen (pUs) zusätzlich eine Zusammenfassung zur Produkt- und Kosteneffektivität vorzulegen, die in die Bewertung mit einfließt (vgl. hierzu auch Schäfer, 2015).

Bereits in Kapitel 1.2.1 dargelegte Faktoren wie Sicherheit, ökonomische Auswirkungen, ethisch-soziale Implikationen und die Wirksamkeit oder der Nutzen einer Technologie stellen den zentralen Gegenstand des Bewertungsprozesses dar (Drummond et al., 2008). Dabei lässt sich die Wirksamkeit grundsätzlich nach drei Ebenen aufschlüsseln (Drummond et al., 2015): Erstens, mit Bezug auf den klinischen Nutzen, also nach der Wirkung unter studiengegebenen Idealbedingungen (Efficacy); zweitens nach der (klinischen) Wirksamkeit unter Alltagsbedingungen (Effectiveness); drittens nach der Effektivität unter Kostenbetrachtung (Efficiency; vgl. hierzu auch Kapitel 1.2.1, Abbildung 1). HTA basiert zwar de facto auf 'Efficacy', sein praktischer Schwerpunkt liegt jedoch primär auf 'Effectiveness' mit dem Kernelement der Bewertung der (vergleichenden) klinischen Wirksamkeit sowie darüber hinaus der Evaluation des Kosten-Nutzen-Verhältnisses (vgl. hierzu auch Schäfer, 2015).

Die nationalen HTA-Programme versuchen sich heute möglichst umfassender, integrativer Ansätze – basierend auf 'Best Practices' oder internationalen Guidelines – zu bedienen, die folglich sowohl in den methodischen Vorgehensweisen als auch in den grundlegenden Anforderungen an die wissenschaftliche Evidenz vorliegender Studien für die systematische Bewertung

---

<sup>13</sup> Einzelne wiederkehrende Ausdrücke und Begrifflichkeiten, die in offiziellen Dokumenten wie den GBA-Beschlusstexten oder IQWiG-Dossierbewertungen ausschließlich im generischen Maskulinum verwendet werden, wurden im weiteren Verlauf der Arbeit durchgängig übernommen. Dazu zählen Patienten(sub)gruppe, Patientenzahl, Patientenpopulation und 'pro Patient' sowie auch die Begriffe patientenberichtet und patientenrelevant.

<sup>14</sup> RCTs gelten als bestes Studiendesign (sogenannter 'Goldstandard'), um bei einer klinischen Fragestellung (zum Beispiel zur Wirksamkeit eines Medikaments in einer Zielgruppe gegenüber einem Placebo in einer Vergleichsgruppe) eine eindeutige Aussage zu erhalten und zudem die Kausalität belegen zu können. Sie werden folglich bei der Erhebung der klinischen Wirksamkeit präferiert, da sie eine hohe interne Validität mit zugleich niedriger Ergebnisverzerrung aufweisen (vgl. Concato et al., 2000).

neuer Technologien inhaltliche Überschneidungen aufweisen (Akehurst et al., 2017; Kleijnen et al., 2012; vgl. hierzu auch Kapitel 2.3).

Mit dem Abschluss der Durchführung folgt im dritten und letzten Schritt die Diffusion der Resultate an die Verantwortlichen der politischen Entscheidungsprozesse. Im Mittelpunkt steht eine gezielte und aktive Kommunikation gegenüber den Entscheidungsträgerinnen und -trägern, damit die erarbeiteten Empfehlungen in der Praxis implementiert werden können. Zudem wird in der Regel das Ausmaß des Einflusses von HTAs sowohl für die praktische Umsetzung als auch für Entscheidungsfindung im Gesundheitswesen verfolgt und dahingehend beurteilt (Gerhardus et al., 2014; vgl. hierzu auch Schäfer, 2015).

Der Entscheidungsbedarf aufgrund medizinischer Innovationen spielt eine zentrale Rolle in den nationalen Versorgungsstrukturen, um deren bestmöglichen Nutzen für die Patientinnen und Patienten nachhaltig sicherzustellen (vgl. hierzu auch bei Hofmann et al., 2021). Neben neuen Gesundheitstechnologien sind aber auch etablierte Interventionen Gegenstand systematischer Bewertungen. Folglich können medizinische Produkte und Therapien, die Teil der Versorgungsstruktur sind und damit bereits durch den Kostenträger erstattet werden, den HTA-Prozess (erneut) durchlaufen, zum Beispiel wenn sich durch Modifikationen neue Indikationen ergeben oder nach einem bestimmten Zeitraum eine Re-Evaluation (auf Basis einer erweiterten, verbesserten Studienlage) stattfinden soll (Zentner et al., 2014).

### **2.3 Nationale HTA-Institutionen und Evaluationsmethoden im Vergleich**

In den folgenden Unterkapiteln werden die offiziellen HTA-Agenturen in Deutschland und England mit den zugehörigen Bewertungsverfahren innerhalb der jeweils bestehenden Versorgungsstruktur vorgestellt.<sup>15</sup> Erweiternd wird die Gesundheitsbehörde HAS herangezogen, die für die Umsetzung sowie Durchführung von HTA in Frankreich zuständig ist und von besonderem Interesse für einen Vergleich von GBA/IQWiG und NICE sein dürfte, da der zugrunde liegende Evaluationsprozess wesentliche Bewertungselemente beider HTA-Ansätze aufweist.

Der Bearbeitungsschwerpunkt liegt ausschließlich auf den zuständigen Institutionen und nationalen HTA-Prozessen. Die vorausgehenden Zulassungsverfahren neuer medizinischer Therapien, die sowohl dezentral durch nationale Regulierungsbehörden als auch zentral durch die

---

<sup>15</sup> Die nachfolgenden beiden Unterkapitel 2.3.1 (GBA und IQWiG in Deutschland) und 2.3.2 (NICE in England) sind im Wesentlichen an die Arbeit von Schäfer (2015) angelehnt und wurden entsprechend an einzelnen Textstellen inhaltlich modifiziert beziehungsweise aktualisiert und in Teilen ergänzt.

European Medicines Agency (EMA) durchgeführt werden können, bleiben demnach weitgehend unberücksichtigt. Dabei kommt den nationalen Zulassungsprozessen heute im Vergleich zur gesamteuropäischen Zulassung durch die EMA (vgl. hierzu EU-Verordnung Nr. 726/2004) eine eher nachrangige Bedeutung zu. Die Beantragung der formalen Erstattung einer neu zugelassenen Gesundheitstechnologie durch die Kostenträger ist demgegenüber wiederum ausschließlich auf Länderebene möglich, weshalb sie für Deutschland, England<sup>16</sup> und Frankreich unabhängig voneinander zu erfolgen hat.

### ***2.3.1 Gemeinsamer Bundesausschuss (GBA) und Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) in Deutschland***

Die offizielle Einbindung von HTA in das beitragsfinanzierte Sozialversicherungssystem ('Bismarck-Modell'), welches in seinen Grundzügen auf einer Versicherungspflicht (die unter anderem für die Krankenversicherung gilt) und dem Prinzip der Solidarität basiert, begann mit der Gesundheitsreform der Gesetzlichen Krankenversicherung (GKV) im Jahr 2000 (Busse et al., 2017). Neben der Implementierung einer HTA-Datenbank wurde mit Gründung der Deutschen Agentur für Health Technology Assessment (DAHTA), die innerhalb des Deutschen Instituts für Medizinische Dokumentation und Information (DIMDI<sup>17</sup>) angesiedelt war, ein erstes wissenschaftliches Arbeitsprogramm zu HTA etabliert. In Folge des GKV-Modernisierungsgesetzes wurden im Jahr 2004 die beiden zentralen Institutionen gegründet, die heute für die Umsetzung der frühen (Zusatz-)Nutzenbewertung (FNB) verantwortlich sind (Perleth et al., 2009). Zum einen handelt es sich um den GBA<sup>18</sup>, der im Januar 2004 als das höchste Beschlussgremium der gemeinsamen Selbstverwaltung implementiert wurde und durch Richtlinien sowohl Leistungs- als auch Zulassungsfragen in der GKV bestimmt, um eine nachhaltige Sicherung der Qualität im deutschen Gesundheitswesen zu gewährleisten (G-BA, 2021). Zum anderen

---

<sup>16</sup> Zum 31. Januar 2020 (gefolgt von einem Übergangszeitraum bis 31. Dezember 2020) ist das Vereinigte Königreich formal aus der EU ausgetreten (vgl. EUR-Lex, Brexit: Beziehungen EU – Vereinigtes Königreich, 30.4.2021). Demzufolge hatten die (spätestens) seit 01. Januar 2021 geltenden gesetzlichen Änderungen respektive Neuerungen – beispielsweise auch bei der Zulassung von Arzneimitteln, die damit nicht mehr der EMA unterliegen – keine Auswirkungen auf die vorliegenden Bearbeitungen und deren Ergebnisse respektive Interpretation.

<sup>17</sup> Das DIMDI war eine nachgeordnete Behörde des BMG, das unter anderem HTA-Berichte erstellte und am 26. Mai 2020 im Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) aufging (BfArM, 2020).

<sup>18</sup> Der GBA, für den das BMG die Rechtsaufsicht wahrnimmt, wird von den vier größten Selbstverwaltungsorganisationen im deutschen Gesundheitswesen gebildet: Aus der Deutschen Krankenhausgesellschaft, der Kassenärztlichen und Kassenzahnärztlichen Bundesvereinigung sowie dem GKV-Spitzenverband (G-BA, 2021).

folgte im Juni 2004 die Gründung des fachlich unabhängigen IQWiG<sup>19</sup>, um die Entscheidungen im Bereich der Nutzen-, Qualitäts- und Wirtschaftlichkeitsbewertung von medizinischen Leistungen innerhalb der GKV wissenschaftlich zu unterstützen. Im Jahr 2007 folgte das GKV-Wettbewerbsstärkungsgesetz, das den anfangs definierten IQWiG-Aufgabenkatalog um die Durchführung formaler Kosten-Nutzen-Bewertungen von neuen Arzneimitteln ergänzte (IQWiG, 2021; vgl. hierzu Schäfer, 2015).

Mit dem Inkrafttreten des Gesetzes zur Neuordnung des Arzneimittelmarktes (AMNOG) am 01. Januar 2011 wurde ein zweistufiges Verfahren zur FNB neu zugelassener Arzneimittel mit neuen Wirkstoffen eingeführt. Gemäß § 35a SGB V (Bewertung des Nutzens von Arzneimitteln mit neuen Wirkstoffen, Verordnungsermächtigung) hat der GBA seither die gesetzliche Aufgabe, innerhalb von drei Monaten nach Markteintritt eines neuen Arzneimittels zu bewerten, ob ein möglicher Zusatznutzen gegenüber der zweckmäßigen Vergleichstherapie (ZVT) anerkannt wird. Zu diesem Zweck sind die pUs verpflichtet, dem GBA spätestens zum Zeitpunkt der Markteinführung ein umfassendes Dossier zur klinischen Bewertung vorzulegen, das einen Zusatznutzen des neuen Arzneimittels gegenüber der vorab bestimmten ZVT belegt. Die formale Bewertung des eingereichten Dossiers kann der GBA entweder selbst durchführen oder er beauftragt damit das – in der praktischen Umsetzung dafür vorgesehene – IQWiG. Die Dossierbewertung wird daraufhin auf der offiziellen Website des GBA veröffentlicht, sodass das pU, Verbände, Sachverständige und Fachkreise die Möglichkeit erhalten, Stellung zu nehmen. Nach weiteren drei Monaten fasst der GBA einen Beschluss auf Grundlage der Dossierbewertung und der eingegangenen Stellungnahmen. Die dann ebenfalls unmittelbar auf der Website veröffentlichte GBA-Beschlussfassung entscheidet folglich über die Preisbildung im Rahmen der sozialversicherungsbasierten Versorgungsstruktur (G-BA, 2020).

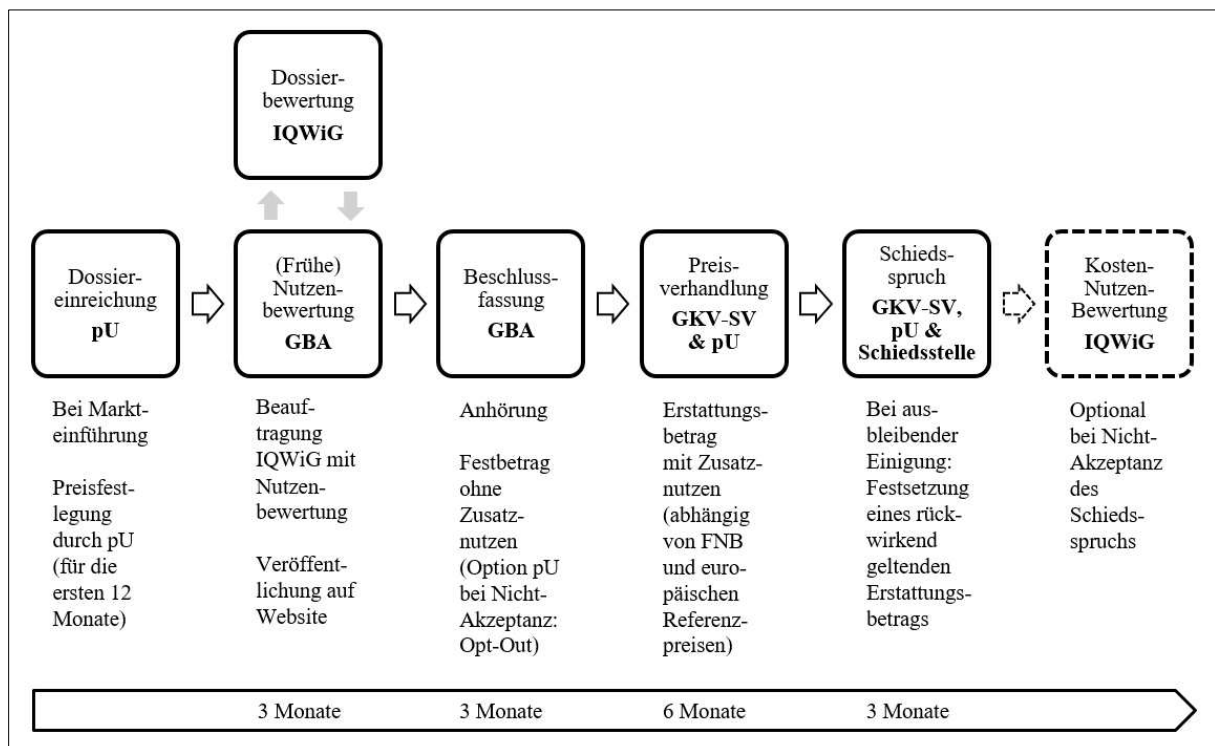
Wenn der GBA den Antrag durch das pU auf Zusatznutzen akzeptiert, verhandelt das pU mit dem GKV-Spitzenverband (GKV-SV) innerhalb von sechs Monaten einen Erstattungspreis für die GKV. Dieser Erstattungsbetrag gilt als Rabatt auf den ursprünglichen, durch das pU für die ersten zwölf Monate nach Markteinführung festgelegten Abgabepreis. Wird in den Verhandlungen keine Einigung erzielt, legt eine Schiedskommission den Erstattungspreis fest. Als Maßstab gilt hierfür seit 25. Oktober 2012 das europäische Preisniveau (vgl. Zweites Gesetz zur

---

<sup>19</sup> Das IQWiG ist ein wissenschaftliches Institut, das Aufträge ausschließlich vom BMG oder dem GBA annehmen darf (oder alternativ aus Eigeninteresse arbeitet) und über die Vor- und Nachteile von medizinischen Leistungen in Form wissenschaftlicher Berichte oder Gutachten sowie über allgemeine Gesundheitsinformationen informiert (IQWiG, 2021).

Änderung arzneimittelrechtlicher und anderer Vorschriften, 2. AMGÄndG). Beide Parteien haben bei Nichtakzeptanz des Schiedsspruchs zudem die Option, eine Kosten-Nutzen-Bewertung beim GBA einzufordern, der daraufhin das IQWiG mit der Durchführung beauftragen kann (G-BA, 2020; vgl. hierzu Schäfer, 2015).

Wenn der GBA dem Antrag durch das pU keinen Zusatznutzen bestätigt, wird das neue Arzneimittel direkt in das Festbetragssystem aufgenommen. Kann ein neues Arzneimittel ohne Zusatznutzen dabei keiner Festbetragsgruppe zugeordnet werden, muss ein Erstattungsbetrag vereinbart werden, welcher die Kosten der ZVT jedoch nicht übersteigen darf (vgl. Gesetz zur Stärkung der Arzneimittelversorgung in der GKV, GKV-AMVSG). Nach Einführung des AM-NOG folgten neben den bereits genannten noch weitere Gesetzesaufträge zur FNB neuer Arzneimittel, die sowohl formale als auch technische Erweiterungen darstellen. Die Gesetze bleiben jedoch unberücksichtigt, da sie für das vorliegende Bearbeitungsthema nicht weiter von Relevanz sind. In nachfolgender Abbildung 2 ist der schematische Ablauf der FNB sowie die daran anschließende Preisfindung gemäß § 35a SGB V nochmals vereinfacht dargestellt.



**Abbildung 2:** Vereinfachter Verfahrensablauf der Frühen (Zusatz-)Nutzenbewertung nach AMNOG (Eigene Darstellung; in Anlehnung an: Schäfer, 2015; Quelle: Bundesgesetzblatt Jahrgang 2010 Teil I Nr. 67, Gesetz zur Neuordnung des Arzneimittelmarktes in der gesetzlichen Krankenversicherung (Arzneimittelmarktneuordnungsgesetz – AMNOG) nach § 35a,b SGB V, S. 2262–2277)

AMNOG, Arzneimittelmarktneuordnungsgesetz; FNB, Frühe (Zusatz-)Nutzenbewertung; GKV-SV, GKV-Spitzenverband; pU, pharmazeutisches Unternehmen.

Wie zuvor ausführlich erläutert obliegt die FNB in Deutschland dem GBA. In einem etablierten, zweistufigen Verfahrensprozess beauftragt der GBA im Regelfall das IQWiG mit der systematischen Durchführung der Bewertung des pU-Dossiers und fasst, basierend auf den (Zusatz-)Nutzenempfehlungen durch das IQWiG, verbindliche Beschlusstexte.

Im Rahmen der FNB verfolgen GBA und IQWiG einen streng evidenzbasierten Ansatz. Die methodischen Anforderungen an die wissenschaftliche Bewertung des (Zusatz-)Nutzens sind dabei in der Verfahrensordnung des GBA geregelt (vgl. G-BA 2020). Diese legt in Verbindung mit der Arzneimittel-Nutzenbewertungsverordnung (AM-NutzenV nach § 35a Absatz 1 SGB V) die formalen und inhaltlichen Vorgaben an die pU-Dossiers fest und definiert die Auswahlbedingungen der ZVT.

Das einzureichende Nutzen-Dossier besteht dabei aus mehreren Modulen<sup>20</sup>, in denen die zugrunde liegenden Studienergebnisse nach den Methoden der EbM zusammenzustellen sind. Beispielsweise müssen in den pU-Dossiers Angaben zu den Kostenfolgen gemacht werden, die bei Aufnahme einer neuen Therapie in die GKV jährlich zustande kommen würden (vgl. hierzu auch in Kapitel 2.4). Die ZVT wird als ein evidenzbasierter, nach internationalen Maßstäben ermittelter Versorgungsstandard verstanden, dessen Voraussetzung sowohl eine bestehende Zulassung im zugrunde liegenden Anwendungsgebiet als auch eine Erstattung innerhalb der GKV ist (G-BA, 2020; vgl. hierzu Schäfer, 2015).

IQWiG bezieht sich für die formale Nutzenbewertung auf ein Methodenpapier, welches unter anderem angelehnt an die Anforderungen der GBA-Verfahrensordnung die wissenschaftlichen Grundsätze respektive Kriterien seiner eigenen Arbeitsweise beschreibt (IQWiG, 2020). Vergleichbar zu internationalen Standards präferieren GBA und IQWiG systematische Übersichtsarbeiten basierend auf direkten Vergleichsstudien, insbesondere RCTs, die eine möglichst hohe Ergebnissicherheit gewährleisten. Grundlegend sollte die Nutzenbewertung – vergleichbar den Anforderungen an die pU-Dossiers – auf der besten zur Verfügung stehenden Evidenz aufbauen. Indirekte Vergleichs- sowie randomisierte Interventions- oder Beobachtungsstudien sollten nur verwendet werden, wenn eine qualitativ hochwertigere Evidenz nicht zur Verfügung steht (G-BA, 2020; IQWiG, 2020; vgl. hierzu Schäfer, 2015).

---

<sup>20</sup> Die Struktur der einzureichenden pU-Dossier-Module für die FNB nach AMNOG setzt sich wie folgt zusammen (IQWiG, 2020): Modul 1, Administrative Informationen und Zusammenfassungen der Bewertungsaussagen; Module 2-4, Angaben zu zugelassenen Anwendungsgebieten inklusive Anforderungen an eine qualitätsgesicherte Anwendung, klinischem Nutzen, medizinischem Zusatznutzen gegenüber der ZVT, Patientenzahl (Populationsgröße) sowie den Kosten und Kostenfolgen für die GKV; Modul 5 (optional), Anlagen.



Nach AM-NutzenV § 2 Abs. 3 SGB V richtet sich die Nutzenbestimmung, bei der indikations-spezifische Nutzenparameter untersucht werden, an drei patientenrelevanten Endpunkten aus: Mortalität (Sterblichkeit), Morbidität (Krankheitshäufigkeit) und der gesundheitsbezogenen Lebensqualität (Health-Related Quality of Life, HRQoL). Als weitere Nutzendimensionen können Ausmaß und Schwere möglicher Nebenwirkungen sowie patientenberichtete Outcomes hinzukommen.

Auf Basis der vom pU dargelegten Angaben quantifiziert IQWiG sowohl den therapeutischen Zusatznutzen als auch dessen Ergebnissicherheit gegenüber der ZVT. Zum einen bewertet IQWiG die Wahrscheinlichkeit für das Vorliegen eines Zusatznutzens aufsteigend nach Anhaltspunkt, Hinweis und Beleg gegenüber der ZVT. Die Bewertung der Aussagesicherheit klinischer Studien ist dabei ein zentraler Bestandteil der methodischen Vorgehensweise der EbM. Zum anderen bewertet IQWiG das Ausmaß des Zusatznutzens, das sowohl die Vorgaben der AM-NutzenV konsistent umsetzen als auch Transparenz für die nachfolgenden AMNOG-Verfahrensschritte schaffen soll. Das Ausmaß lässt sich bei einem festgestellten Zusatznutzen nach erheblich, beträchtlich, gering und nicht quantifizierbar differenzieren. Neben der inhaltlichen Bewertung der Dossiers durch Prüfung der Validität und Vollständigkeit der pU-Angaben wird von externen Sachverständigen klinisch-medizinische Expertise eingeholt und zusätzlich die Perspektive von Patientinnen und Patienten miteinbezogen. Darüber hinaus kann das IQWiG eigene Recherchen und Analysen zur Unterstützung der Nutzenbewertung durchführen (IQWiG, 2020; vgl. hierzu Schäfer, 2015).

Die Operationalisierung der Feststellung des Zusatznutzens durch IQWiG umfasst demnach drei elementare Schritte, die sich wie folgt zusammenfassen lassen (IQWiG, 2020):

[1] Prüfung der qualitativen Aussage

Einstufung der Wahrscheinlichkeit für das Vorliegen eines Zusatznutzens:  
Anhaltspunkt, Hinweis, Beleg;

[2] Prüfung der quantitativen Aussage

Ausmaß des Zusatznutzens nach Effektstärke:  
erheblich, beträchtlich, gering, nicht quantifizierbar, nicht vorhanden, geringer;

[3] Gesamtaussage beziehungsweise abschließende Empfehlung zum Zusatznutzen

Zusammenführung von Wahrscheinlichkeit und Ausmaß des Zusatznutzens.

In einzelnen Fällen werden zudem Addenda zu bereits abgeschlossenen Dossierbewertungen durch IQWiG erstellt, die anschließend an die durchgeführte Nutzenbewertung aufgrund eines zusätzlichen Bearbeitungsbedarfs veröffentlicht werden und folglich erweiternde Informationen enthalten. Dies ist unter anderem dann der Fall, wenn sich die Studienlage während des Stellungnahme-Verfahrens ändert oder der GBA für die Beschlussfassung ergänzende Aspekte anfordert (IQWiG, 2020).

Ausgenommen von der FNB sind grundsätzlich Arzneimittel mit einem geringfügigen Jahresumsatz von bis zu einer Mio. EUR. Ebenfalls unterstehen die Wirkstoffe für seltene Erkrankungen, sogenannte Orphan Drugs (ODs), einer Ausnahmeregelung. Die gesetzliche Bestimmung sieht vor, dass bei Arzneimitteln mit einem OD-Status per Definition ein Zusatznutzen als belegt gilt. Voraussetzung hierfür ist, dass jährliche GKV-Kosten von 50 Mio. EUR nicht überschritten werden. Dementsprechend bewerten GBA oder IQWiG nur das Ausmaß des vorhandenen Zusatznutzens, mit dem primären Schwerpunkt auf die Größe der GKV-Zielpopulation und die jährlichen Therapiekosten pro Patient (G-BA, 2020; vgl. hierzu Schäfer, 2015).

Im Gegensatz zu anderen nationalen HTA-Prozessen (wie zum Beispiel in England) stellt die ökonomische Evaluation keinen obligatorischen Bestandteil der FNB nach AMNOG dar (vgl. Bundesgesetzblatt Jahrgang 2010 Teil I Nr. 67, Arzneimittelmarktneuordnungsgesetz – AMNOG nach § 35a,b SGB V). Sowohl die GBA-Beschlussfassungen als auch die IQWiG-Dossierbewertungen fokussieren sich nach den Leitlinien der EbM ausschließlich auf die klinische Effektivität einer neuen Therapie gegenüber der ZVT. Dabei lehnen sowohl GBA als auch IQWiG die Einbindung der Kosteneffektivität – unter Verwendung eines indikationsübergreifenden Nutzenmaßes wie dem QALY – aus verschiedenen Gründen ab. Dazu zählen neben wissenschaftlich-methodischen Kontroversen insbesondere auch kulturelle Aspekte, sowie darüber hinaus, dass anwendungsübergreifende Outcome-Parameter – auf der Basis von beispielsweise rechtlichen Vorgaben – Werturteile über den Stellenwert unterschiedlicher Indikationen implizieren (IQWiG, 2020).

Allenfalls bei einer optionalen Beauftragung durch den GBA (oder das Bundesministerium für Gesundheit, BMG) führt IQWiG eine Kosten-Nutzen-Bewertung<sup>21</sup> nach § 35b SGB V durch, die jedoch nur zustande kommen kann, wenn ein pU nach Erstattungspreisverhandlungen mit dem GKV-SV den Schiedsspruch nicht akzeptieren sollte (G-BA, 2020). Methodisch greift

---

<sup>21</sup> Die optionale Durchführung einer Kosten-Nutzen-Bewertung durch IQWiG spielt bis dato (noch) keine Rolle im Kontext der FNB in Deutschland (G-BA, 2019).

IQWiG dann auf die sogenannte 'Analyse der Effizienzgrenze' zurück, die das Verhältnis zwischen Kosten und Nutzen einer Therapie gegenüber Interventionen desselben Indikationsgebiets bewertet (vgl. hierzu bei Caro et al., 2010). Ergänzend wird für eine Einschätzung der möglichen finanziellen Auswirkungen die BIA herangezogen, indem IQWiG die auf einzelne Patientinnen und Patienten bezogene Kosten-Nutzen-Bewertung (Cost-Benefit Analysis, CBA) auf die gesamte Zielpopulation innerhalb der GKV abbildet (IQWiG, 2020).

Abschließend ist mit Blick auf die unmittelbaren Auswirkungen der FNB noch die Option der pUs für ein sogenanntes 'Opt-Out' zu nennen. Opt-Out bedeutet, dass ein pU den Vertrieb für sein Arzneimittel innerhalb von vier Wochen nach Beschlussfassung durch den GBA einstellen kann (vgl. § 4 Abs. 7 Rahmenvereinbarung nach § 130b Abs. 9 SGB V). Die Vertriebeinstellung kann dabei einen wirtschaftlichen Hintergrund haben (zum Beispiel aufgrund einer Erstattungspreisfestsetzung bei nicht festgestelltem Zusatznutzen, wodurch die Kosten der ZVT nicht überschritten werden dürfen), aber auch komplexere Gründe, wenn eine spätere Rücknahme vom Markt deutschland- oder gegebenenfalls europaweit erfolgte.

### ***2.3.2 National Institute for Health and Care Excellence (NICE) in England***

Beim englischen NHS handelt es sich demgegenüber um ein zentral verwaltetes, steuerfinanziertes Gesundheitssystem ('Beveridge-Modell'), das auf eine universelle Grundversorgung für die gesamte Bevölkerung abzielt (Prinzip der Unentgeltlichkeit) und zusätzlich durch Eigenleistungen ergänzt werden kann (Grosios et al., 2010). Nach ersten thematischen Auseinandersetzungen in den 1970er Jahren folgte mit der Einführung des Research and Development Programme Anfang der 1990er Jahre die Integration von HTA in das NHS. Aufgrund begrenzter Kapazitäten sowie mangelnder Ressourcen wurden jedoch wesentliche Teile des HTA-Prozesses, wie zum Beispiel die Dissemination der Resultate, an das externe National Coordinating Centre for Health Technology Assessment abgetreten, wodurch immer wieder Kritik bezüglich der Effektivität der Umsetzung von HTAs aufkam (Drummond and Banta, 2009).

Folglich geht die Gründung von NICE<sup>22</sup> im April 1999 – zuerst als offizielle Gesundheitsbehörde mit besonderen Befugnissen, später als nicht-öffentliche Regulierungsinstitution – primär auf das Ziel einer effektiveren Nutzung der zur Verfügung stehenden Ressourcen innerhalb des NHS zurück. Als unabhängige, beratende Einrichtung, die vom Department of Health and

---

<sup>22</sup> Bei gleichbleibendem Akronym wurde NICE ursprünglich unter dem Namen National Institute for Clinical Excellence gegründet und heißt seit April 2013 National Institute for Health and Care Excellence.

Social Care (DH) finanziert wird, stellt NICE allen voran Empfehlungen respektive Richtlinien (Guidance) für die medizinische Versorgung innerhalb des NHS England bereit. Die Grundlage der NICE Guidance stellen systematische Bewertungen durch unabhängige wissenschaftliche Ausschüsse (interner und) externer Fachleute dar (NICE, 2021).

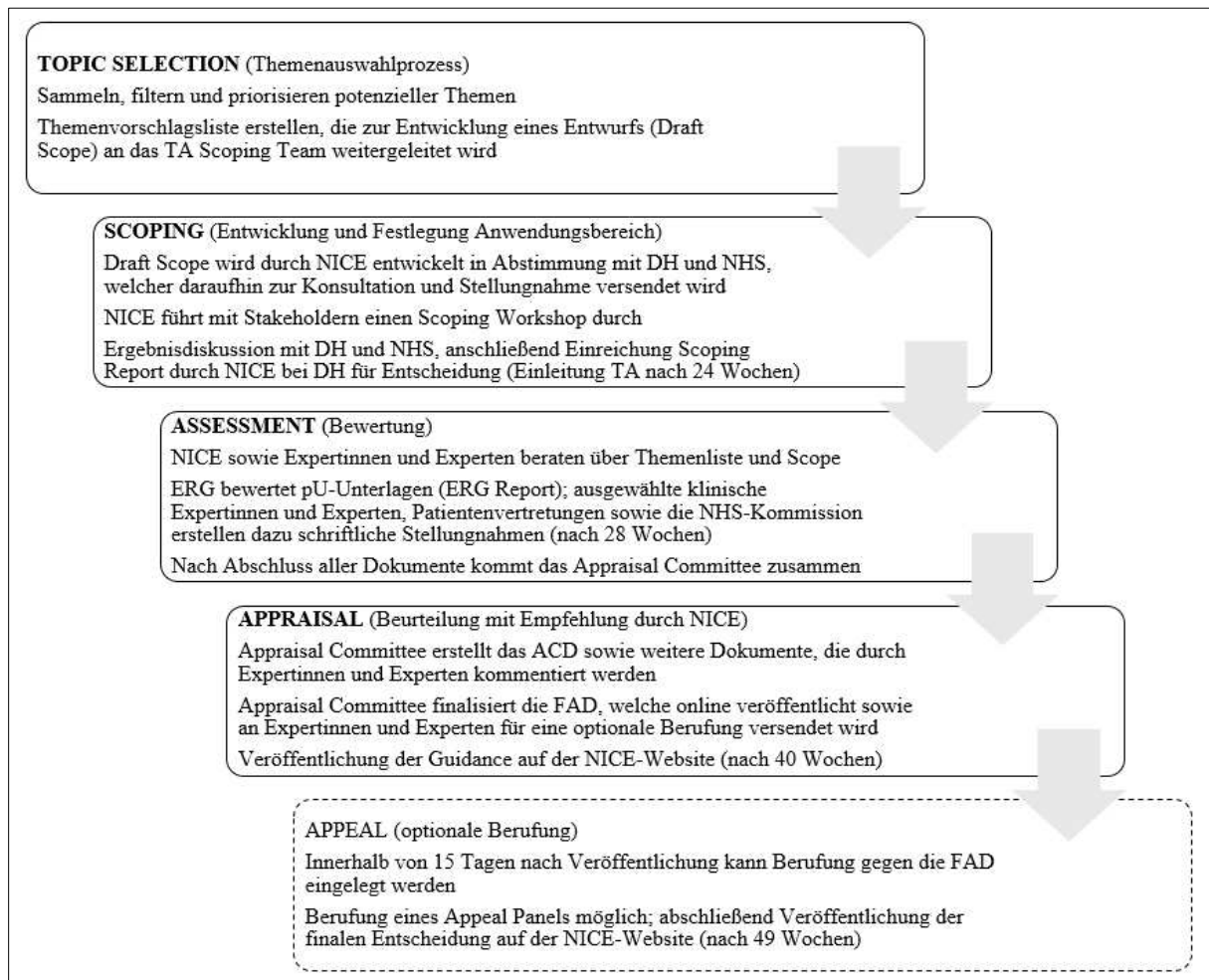
NICE unterteilt sich in seinem Aufbau in insgesamt drei Kompetenzzentren (Centre for Clinical Practice, Centre for Public Health, Centre for Health Technology Evaluation) mit vier verschiedenen Bereichen respektive Leitungsebenen (Communication, Health and Social Care, Evidence Resources, sowie Business, Planning and Resources), die sowohl organisatorisch als auch thematisch unterschiedliche Schwerpunkte haben. Im weiteren Fortgang liegt der Fokus ausschließlich auf dem Centre for Health Technology Evaluation (CHTE), das für die Entwicklung der Guidance von Arzneimitteln und Behandlungsmethoden zuständig ist. In Form sogenannter 'Technology Appraisals' (TAs) gibt das CHTE dabei Empfehlungen für die Verwendung bestehender und neuer Gesundheitstechnologien heraus, die innerhalb des NHS England rechtsverbindlich sind (Schlander, 2007). Folglich müssen die durch NICE TAs positiv evaluierten medizinischen Technologien mittels lokal budgetierter Ressourcen (innerhalb des NHS) zur Verfügung gestellt werden (vgl. hierzu Schäfer, 2015).

Der zugrunde liegende Prozess der TAs, der auf den bestmöglichen zur Verfügung stehenden Daten und Informationen basiert, gliedert sich in die drei Phasen 'Scoping', 'Assessment' und 'Appraisal'. Ein 'Appeal' kann optional als vierte Phase hinzukommen, in der eine Berufung gegen die sogenannte Final Appraisal Determination (FAD) eingelegt werden kann (NICE, 2018).

Der Ablauf lässt sich dabei nach Multiple Technology Appraisals (MTAs), Single Technology Appraisals (STAs) und Fast Track Appraisals (FTAs) unterscheiden. MTAs bilden die ursprüngliche Vorgehensweise, die entweder mehr als eine Technologie oder eine Technologie für mehr als eine Indikation abdecken (NICE, 2013). Aufgrund der jedoch verhältnismäßig langen Dauer von etwa 52 Wochen führte NICE zusätzlich den STA-Prozess ein. Anfangs primär zur Bewertung onkologischer Erkrankungen installiert, werden STAs heute praktisch für alle therapeutischen Gebiete genutzt und stellen den am häufigsten verwendeten Bewertungsprozess durch NICE dar. Die Durchführung eines STAs, die im weiteren Verlauf dieses Unterkapitels ausführlich dargelegt wird (vgl. Abbildung 3), erfasst dabei ausschließlich eine Technologie für eine einzige Indikation und dauert insgesamt 40 Wochen (NICE, 2018).

Bei FTAs für ausschließlich sehr kosteneffektive Gesundheitstechnologien ( $ICER < 10.000$  GBP/QALY oder ein positiver Kostenvergleich gegenüber einer TA Guidance innerhalb der

gleichen Indikation), welche seit April 2017 zusätzlich zu den MTAs und STAs existieren, verpflichtet sich das NHS im Falle einer positiven Bewertungsempfehlung zur Kostenübernahme spätestens 30 Tage nach Veröffentlichung der NICE Guidance. Sogenannte Highly Specialised Technologies (HSTs) sind demgegenüber Teil eines separaten Bewertungsverfahrens, das ausschließlich bei der Bewertung von Interventionen für sehr seltene Erkrankungen Anwendung findet (NICE, 2018).



**Abbildung 3:** Single Technology Appraisal – zusammengefasste Prozessdarstellung (Eigene Darstellung; in Anlehnung an: Schäfer, 2015; Quelle: NICE, 2018)

ACD, Appraisal Consultation Document; DH, Department of Health; ERG, Evidence Review Group; FAD, Final Appraisal Determination; TA, Technology Appraisal.

Vor dem Scoping, das die erste Phase eines TAs darstellt, findet ein Auswahlprozess (Topic Selection) statt, bei dem Themenvorschläge mithilfe eines Scanning zuerst gefiltert und an-

schließlich mittels spezifischer Kriterien<sup>23</sup> priorisiert werden. Auf dieser Basis erfolgen konkrete Entwürfe politischer Themenstellungen, aus welchen NICE den Fokus des Aufgaben- und Geltungsbereichs (Scope) entwirft. In Abstimmung mit dem DH und dem NHS werden die priorisierten Themen zur Konsultation an externe Expertinnen und Experten versendet. Daraufhin werden die kommentierten Themenvorschläge in einen von NICE initiierten Scoping Workshop integriert. Die Ergebnisse dienen als Grundlage für eine abschließende Diskussion mit dem NHS und DH, bevor NICE alle zur Verfügung stehenden Informationen zu einem Scoping Report zusammenfasst und beim DH einreicht. Das DH trifft die finale Priorisierungsentscheidung und beauftragt NICE mit der Themenveröffentlichung und anschließenden Bewertung (NICE, 2018).

Zentrale Komponenten der Scoping-Phase stellen die Definition der klinischen Fragestellung, der Patientenpopulation und der Vergleichstherapie(n) dar. Zudem sind eine Maßeinheit für die Kosten, relevante Ergebnisvariablen und der zeitliche Horizont des anschließenden Bewertungsprozesses festzulegen. Vom Scoping Draft bis zur Übermittlung des Scoping Report an das DH vergehen in der Regel etwa 18 Wochen, wobei die schrittweise Erstellung bis hin zum finalen Scope vom Themenbereich (onkologisch, nicht-onkologisch) abhängt. Weitere sechs Wochen nimmt der Prozess von der Entscheidungsübermittlung bis zur Einleitung eines TAs durch NICE in Anspruch (NICE, 2018; vgl. hierzu Schäfer, 2015).

Während des Assessment (STA-Prozess) erfolgt durch eine externe, fachlich heterogene Evidence Review Group (ERG) die Ausarbeitung eines unabhängigen Reports, der in erster Linie Untersuchungsergebnisse zu Kosteneffektivität und klinischer Wirksamkeit der zu evaluierenden Technologie beinhaltet. Der ERG Report beruht auf den eingereichten Unterlagen der pUs (Manufacturer Submission Document), die systematisch evaluiert werden. Zur abschließenden Bewertung werden schriftliche Stellungnahmen ausgewählter klinischer Fachleute sowie betroffener Interessengruppen hinzugezogen. Nach Abschluss des ERG Report (nach etwa 16 Wochen) und dem Vorliegen aller Stellungnahmen kommt ein auf drei Jahre gewählter Ausschuss (NICE Appraisal Committee) zusammen. Die Dauer beträgt etwa 28 Wochen bis alle relevanten Dokumente konsolidiert versendet werden können (NICE, 2018; vgl. hierzu Schäfer, 2015).

---

<sup>23</sup> Die Kriterien für eine Priorisierung werden nach vier Kategorien eingeteilt: Population, Schwere der Krankheit, Ressourcenverwendung und (geforderter) therapeutischer Nutzen. Mittels Punktesystem wird innerhalb der einzelnen Kategorien priorisiert, sodass unterschiedliche Themen bei der anschließenden Auswahl vergleichbar gemacht werden können. Dabei wird im Auswahlprozess (Filterung und Priorisierung) zwischen onkologischen und nicht-onkologischen Themen differenziert (NICE, 2018).

Mit dem Zusammenkommen des NICE Appraisal Committee startet die Appraisal-Phase (STA-Prozess). Unter Berücksichtigung des ERG Report und allen weiteren bewertungsrelevanten Dokumenten aus der Assessment-Phase sowie der Anhörung ausgewählter Fachleute wird entweder vorab, sofern keine oder eine optimierte Empfehlung (das heißt nicht für die gesamte Zielpopulation, sondern lediglich für eine Subpopulation) vorliegt, ein Appraisal Consultation Document (ACD) entwickelt oder es folgt die direkte Erstellung einer finalen Bewertungsempfehlung (Final Appraisal Determination, FAD). Mit der Veröffentlichung der FAD auf der NICE-Website (nach etwa 36 Wochen) wird die Option auf eine Berufung (Appeal) ermöglicht. Auf Grundlage der FAD erfolgt nach etwa 40 respektive 49 Wochen die abschließende Veröffentlichung der NICE STA Guidance, die innerhalb des NHS rechtlich bindend ist (NICE, 2018; vgl. hierzu Schäfer, 2015).

Seit seiner Gründung vor mehr als 20 Jahren gilt NICE als ein internationaler Vorreiter für die Umsetzung von HTA unter Bezugnahme der Kosteneffektivität, das hohe methodische Standards mit einem transparenten Vorgehen unter Einbindung weiterer Bewertungsfaktoren, wie zum Beispiel sozial-ethischer Aspekte, verknüpft (Charlton, 2020; 2022; de Folter et al., 2018; Littlejohns et al., 2019; Rawlins et al., 2010). Die explizite Berücksichtigung eines Schwellenwerts von 20.000–30.000 GBP/QALY deutet jedoch zugleich daraufhin, dass sich Empfehlungen von NICE primär durch die inkrementelle Kosteneffektivitäts-Relation begründen (Cerri et al., 2014; Dakin et al., 2015; 2006; Devlin and Parkin, 2004; Tappenden et al., 2007; Walton et al., 2019). In aller Regel werden Technologien mit einem ICER  $< 20.000$  GBP/QALY durch NICE vollumfänglich oder (zumindest) eingeschränkt<sup>24</sup> empfohlen. Bei einem ICER von 20.000–30.000 GBP/QALY werden weitere Faktoren wie der Grad der Unsicherheit bei der Berechnung des ICERs, das vorliegende Innovationspotenzial, die gesellschaftlichen Kosten oder Unsicherheiten bezüglich der definierten Patientenpopulation berücksichtigt, um eine abschließende Empfehlung zu treffen. Interventionen mit einem ICER  $> 30.000$  GBP/QALY werden grundsätzlich nach ergänzenden Bewertungskriterien geprüft, worunter insbesondere die sogenannten End-of-Life (EoL)-Kriterien fallen (NICE, 2013; vgl. hierzu Schäfer, 2015).

Die Kriterien für EoL-Interventionen gelten als erfüllt, wenn eine Technologie für eine kleine Patientenpopulation ( $\leq 7.000$ ) indiziert ist und die durchschnittliche Lebenserwartung weniger

---

<sup>24</sup> Eine restriktive oder auch optimierte Empfehlung beschreibt Bewertungen durch NICE, wenn die Anwendung einer Intervention innerhalb des NHS als nicht uneingeschränkt empfohlen wird. Ihre Klassifizierung begründet sich wie folgt: i) Zweitlinien- oder anschließende Therapie, ii) intolerant für eine weitere Behandlung, iii) die Antwort(rate) liegt innerhalb eines spezifischen Zeitraums, oder iv) innerhalb einer zugelassenen Indikation wurden Subgruppen definiert, die unterschiedliche Bewertungsergebnisse aufweisen. Weitere Subkriterien können die Klassifizierung zusätzlich ergänzen (Raftery, 2006).

als 24 Monate beträgt. Zudem müssen ausreichend Belege vorliegen, die der Technologie im Vergleich zu bereits verfügbaren Behandlungsoptionen derselben Indikation eine Lebenszeitverlängerung von mindestens drei Monaten zusprechen (NICE, 2009). Die ICER-Schwellenwerte stellen demnach keinen fixen Grenzbereich dar, der per se zu (einer Einschränkung oder) Ablehnung durch NICE führt. Vielmehr bedarf es des Nachweises weiterer Evidenz, die zusätzlich zur Kosteneffektivität und der klinischen Wirksamkeit in den TAs berücksichtigt wird.

In der Praxis finden die EoL-Kriterien hauptsächlich bei der Bewertung von Krebstechnologien Berücksichtigung, die einen ICER (über dem Schwellenwert) von bis zu 50.000 GBP/QALY aufweisen (vgl. hierzu bei Bovenberg et al., 2021). Die Implementierung des Cancer Drugs Fund (CDF) im Jahr 2011 ist auf das Ziel zurückzuführen, innovative und zugleich sehr teure Krebstherapien zu finanzieren, die NICE (trotz Erfüllung der EoL-Kriterien) als nicht kosteneffektiv einstuft und folglich auch nicht für eine routinemäßige Verwendung innerhalb des NHS empfiehlt. Mit einem Relaunch im Juli 2016 folgte die Konkretisierung für eine nachhaltigere Umsetzung des CDF, welcher einen früheren (und damit schnelleren) Zugang zu vielversprechenden neuen Technologien sicherstellen soll, bis ein pU ausreichend Evidenz vorlegen kann (NHS England, 2016).

NICE bevorzugt die durch die pUs einzureichenden Dokumente in einer integrierten Darstellung, das heißt in Form systematischer Reviews, die auf relevanten RCTs basieren sollten. Alternativ können Studiendaten auch zu Meta-Analysen synthetisiert werden. NICE berücksichtigt jedoch auch nicht-randomisierte Studien oder Beobachtungs- und Interventionsstudien, sofern keine RCTs vorgelegt werden können. Zudem kann NICE auch epidemiologische und experimentelle Studien sowie andere Formen qualitativer Forschungsergebnisse akzeptieren. Mittels Sensitivitätsanalysen werden darüber hinaus zugrunde liegende Unsicherheiten in den Ergebnissen ökonomischer Evaluationen untersucht (NICE, 2013; vgl. hierzu Schäfer, 2015).

Weiterhin sind die pUs dazu angehalten, konkrete Angaben zur Auswirkung der neuen Technologie auf das vorhandene NHS Budget zu machen (vgl. hierzu auch Kapitel 2.4). Für positiv bewertete Technologien veröffentlicht(e) NICE zusammen mit der TA Guidance sogenannte 'Costing Templates', um die ressourcenwirksamen Folgen ('Resource Impact') auf lokaler Ebene besser abschätzen zu können. Seit April 2017 führt NICE zudem einen offiziellen BI-Test durch, um den potenziellen finanziellen Einfluss neuer Technologien innerhalb der ersten drei Jahre bei routinemäßiger Anwendung innerhalb des NHS zu bewerten (NICE, 2017). Das wiederum bedeutet für Interventionen, die in den ersten drei Jahren einen BI von 20 Mio. GBP überschreiten, dass NICE Verhandlungen zwischen dem NHS und dem pU einleitet. Ziel dieser



kommerziellen Verhandlungen soll es sein, den budgetwirksamen Einfluss der neuen Technologie auf das NHS-Budget zu reduzieren. Sollte dabei keine Einigung zwischen den beiden Parteien erzielt werden, kann der NHS (in begründeten Ausnahmefällen) noch eine Sonderform der gesetzlichen Finanzierungsanforderung beantragen (NICE, 2018).

### **2.3.3 *Haute Autorité de Santé (HAS) in Frankreich***

Das französische Gesundheitswesen basiert in seinen Grundzügen ebenfalls auf einem Sozialversicherungsmodell ('Bismarck-Modell'), das ähnlich wie in Deutschland von einer gesetzlichen Krankenversicherung (l'Assurance Maladie) geprägt ist und darüber hinaus durch Zusatzversicherungen individuell erweitert werden kann (Chevreul et al., 2015).

Die ersten thematischen Auseinandersetzungen mit HTA reichen in Frankreich bis in die späten 1970er Jahre zurück. Mit der Gründung der Agence Nationale de l'Evaluation Medicale (ANDEM) im Jahr 1989 folgte die Entwicklung eigener Evaluationen sowie geeigneter Methoden zur Bewertung medizinischer Technologien und Praktiken. Später wurde das Mandat der ANDEM auf die Akkreditierung von Krankenhäusern ausgeweitet, mit der zugleich eine Namensänderung in Agence Nationale d'Accreditation et d'Evaluation en Santé (ANAES) einherging (Orvain et al., 2004; Weill and Banta, 2009).

Einige zentrale Reformen des Gesundheitswesens, die insbesondere auch die bestehenden HTA-Bemühungen beeinflussten, folgten nach 1990. Von gesondertem Interesse ist hierbei das Krankenversicherungsreformgesetz 'Réforme Douste-Balzy' aus dem August 2004, in Zuge dessen das französische Gesundheitsministerium auch die HAS<sup>25</sup> gründete (Weill and Banta, 2009). Seit 2005 bündelt die HAS alle nationalen HTA-Bemühungen, um die politische Entscheidungsfindung in Frankreich mittels Empfehlungen (und Stellungnahmen) hinsichtlich der Kostenerstattung und Preisgestaltung von neuen Therapien zu unterstützen (HAS, 2018).

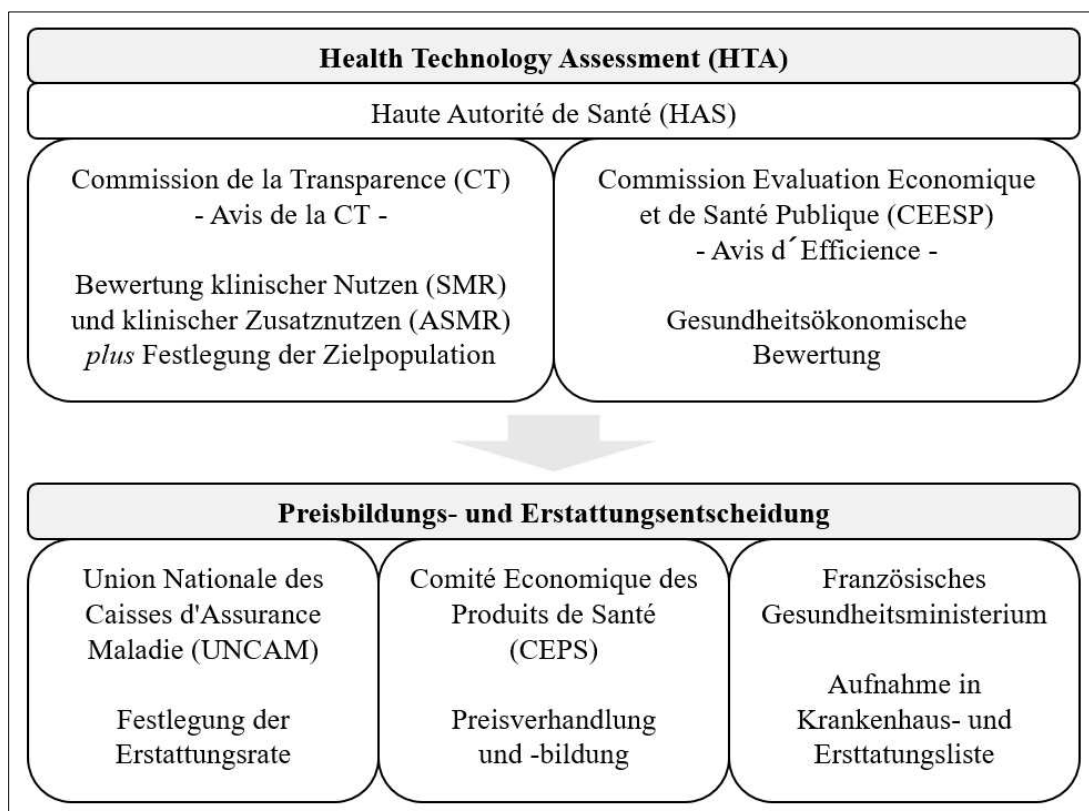
Der offizielle HTA-Prozess, der in seinen Grundzügen auch als Vorbild für die methodische Umsetzung der FNB in Deutschland diente, unterliegt folglich der HAS, die die systematische Bewertung neuer medizinischer Produkte durchführt und daraufhin eine konkrete Empfehlung (Avis de la Commission de la Transparence, ACT) ausstellt. Auf Grundlage der ACT folgt vor

---

<sup>25</sup> Die HAS ist eine öffentliche und rechtlich weitgehend unabhängige, wissenschaftlich arbeitende Gesundheitsbehörde, deren grundlegende Zielsetzung darin besteht, die Regulierung des französischen Gesundheitswesens weiter zu verbessern und zugleich effizienter zu gestalten (HAS, 2018).

Markteinführung eines neuen Medikaments die Erstattungsentscheidung durch das Gesundheitsministerium beziehungsweise die Union Nationale des Caisses d'Assurance Maladie (UNCAM), sowie eine Festsetzung des Preises durch die Comité Economique des Produits de Santé (CEPS) und eine entsprechende Festlegung der Höhe der Zuzahlungen durch die gesetzliche Krankenversicherung (HAS, 2015; vgl. Abbildung 4).

HAS ACTs basieren dabei zum einen auf einer Bewertung des tatsächlichen klinischen Nutzens (Service Médical Rendu, SMR) und zum anderen auf dem klinischen Zusatznutzen (Amélioration du Service Médical Rendu, ASMR) eines neuen Medikaments im Vergleich zu einer (bestmöglichen) verfügbaren (Standard-)Therapie (HAS, 2020). Während SMR eine Empfehlung zur Aufnahme in die Erstattungsliste in Frankreich impliziert, evaluiert ASMR den (Zusatz-)Nutzen eines neuen Medikaments gegenüber den verfügbaren Arzneimitteln. Darüber hinaus spielen die Indikation respektive das Anwendungsgebiet und die zugehörige Patientenpopulation (das heißt vor allem die definierte Zielpopulation) eine gewichtige Rolle bei der Bewertung (Detournay et al., 2021; Dubromel et al., 2018; Rémuzat et al., 2013).



**Abbildung 4:** Übersicht zu HTA, Preisbildung und Erstattung im französischen Gesundheitssystem (Eigene Darstellung; Quelle: HAS, 2015)

ASMR, Amélioration du Service Médical Rendu; SMR, Service Médical Rendu.

Beim SMR, das den Nutzen des neuen Arzneimittels hinsichtlich einer Kostenübernahme durch die gesetzliche Krankenversicherung untersucht, werden Faktoren wie der Schweregrad der Krankheit, das Verhältnis von Wirksamkeit zu Nebenwirkungen (anhand klinischer Endpunkte), die Rolle in der therapeutischen Strategie gegenüber anderen verfügbaren Therapien, der Nutzen für die öffentliche Gesundheit und die Art der Behandlung (präventiv, kurativ oder symptomatisch) berücksichtigt. Die Empfehlung in der ACT für eine Aufnahme in die Erstattungsliste ist dabei entweder hinreichend (positive Nutzenempfehlung inklusive prozentualem Erstattungsanteil: beträchtlich, 65 %; moderat, 30 %; gering, 15 %) oder unzureichend (bei negativer Nutzenempfehlung liegt der Erstattungsanteil bei 0 %). Sofern keine privaten Zusatzversicherungen vorliegen, verbleibt grundsätzlich ein gewisser Prozentsatz der Kosten bei den Versicherten ('Ticket Modérateur'). Medizinische Leistungen für chronische (Langzeit-)Erkrankungen oder für besonders schwerwiegende Krankheiten können zudem auch vollständig erstattet (100 %) werden (HAS, 2020).

Beim ASMR, das daraufhin den Umfang des klinischen Zusatznutzens untersucht, werden primär (klinische) Wirksamkeits- sowie Sicherheitsdaten des neuen Medikaments in Bezug auf die verfügbare Vergleichstherapie berücksichtigt. Der Umfang der ASMR in der ACT unterteilt sich dabei in fünf Ebenen (I–V). Während I–IV einen Zusatznutzen gegenüber der Vergleichstherapie bestätigen (I, erheblich; II, beträchtlich; III, moderat; IV, gering), liegt bei V kein klinischer Zusatznutzen vor. Ein Medikament, das keinen klinischen Zusatznutzen aufweist, kann nur dann in die Liste der erstattungsfähigen Arzneimittel aufgenommen werden, wenn es direkte Einsparungen in den Behandlungskosten vorweist (HAS, 2020).

Zudem ist seit Oktober 2013 eine CEA durch die HAS beziehungsweise genauer gesagt durch die Commission Evaluation Economique et de Santé Publique (CEESP) verpflichtend, wenn bestimmte Voraussetzungen bei der Bewertung der klinischen Effektivität als erfüllt gelten. Zum einen hängt die Durchführung einer CEA (Avis d'Efficiencie) vom Innovationsgrad der medizinischen Leistung ab (pU-Antrag mit ASMR I–III) und zum anderen von der Wahrscheinlichkeit, dass das Medikament eine womöglich signifikante Auswirkung auf die Ausgaben des Gesundheitssystems haben wird (HAS, 2021; vgl. hierzu auch bei Toumi et al., 2017). Im Gegensatz zur Vorgehensweise anderer HTA-Agenturen (wie zum Beispiel von NICE in England) begründet sich die resultierende Empfehlung dabei nicht auf einem ICER-Schwellenwert, sondern stellt vielmehr eine unterstützende Evaluationskomponente unter expliziter Heranziehung der Kosten(effektivität) dar.

Vergleichbar den Anforderungen in England und Deutschland müssen die pUs auch bei der HAS ein Dossier einreichen, das auf der bestmöglichen zur Verfügung stehenden klinischen Evidenz basieren sollte. RCTs in Form systematischer Reviews oder Meta-Analysen stellen dabei den präferierten Standard dar, auch wenn nicht-randomisierte Studien oder Interventionsanalysen akzeptiert werden können (HAS, 2020). Zudem müssen Angaben in Form einer CEA eingereicht werden, die die Kosteneffektivität der neuen Therapie gegenüber einer verfügbaren Vergleichstherapie aufzeigt (HAS, 2021). Seit Januar 2016 sind die pUs darüber hinaus verpflichtet, für ein neues Arzneimittel, das aller Wahrscheinlichkeit nach einen Umsatz von mehr als 50 Mio. EUR innerhalb der ersten zwei Jahre erzielen wird, ergänzend eine BIA einzureichen. Für Medikamente, die mehrere Indikationen mit unterschiedlichem Zusatznutzen aufweisen, betrifft die BIA ausschließlich die zum vorliegenden Zeitpunkt eingereichte(n) Indikation(en). Zusätzlich kann die HAS für den Evaluationsprozess auf die öffentlich zugängliche Fachliteratur zurückgreifen und entsprechend weitere Informationen heranziehen, die über die vorgelegten pU-Daten hinausgehen (Ghabri et al., 2018).

#### **2.4 BIAs als methodischer Bestandteil von HTA? Eine Bestandsaufnahme mit gesundem Fokus auf Deutschland und England**

Wie in den Ausführungen zuvor dargelegt hat sich unter den traditionellen (gesundheits-)ökonomischen Evaluationsmethoden das Konzept der QALYs als ein De-facto-Standard für die Messung der Effektivität bei HTAs durchgesetzt, das mittels inkrementeller Kosteneffektivitäts-Verhältnisse den Vergleich zweier Behandlungsalternativen unterschiedlicher Indikationen ermöglichen soll (Angelis et al., 2018; Beletsi et al., 2018; Stafinski et al., 2011a). Dies deutet wiederum darauf hin, dass bei (ausschließlicher) Analyse der Kosteneffektivität gesellschaftliche beziehungsweise soziale Präferenzen nicht (vollumfänglich) abgebildet werden (können). Hierzu zählt beispielsweise eine möglichst gerechte Allokation von begrenzten finanziellen Ressourcen im Gesundheitswesen (*'efficiency or fairness first'*; Richardson and Schlander, 2018), die insbesondere auch den Bestimmungszweck der Kostenfolgen von zu erstattenden Arzneimitteln und Therapien verdeutlicht.

BIAs werden (demnach) sukzessive als ergänzende Komponente einer umfassenden ökonomischen Bewertung gefordert, um den finanziellen Einfluss einer neuen Technologie auf ein zur Verfügung stehendes Budget oder die Gesamtkosten eines Versorgungssystems abschätzen zu können. In der einschlägigen Literatur wird folglich darauf verwiesen, die BIA zusätzlich zu einer CEA durchzuführen, weshalb sich die Diskussionen in der Regel um deren praktische

Umsetzung sowie Integration in den bestehenden HTA-Prozess drehen (Mauskopf, 1998; Mauskopf et al., 2007; 2017; Sullivan et al., 2014; Trueman et al., 2001).

Die fortschreitende Entwicklung von sowohl länderspezifischen Leitlinien als auch allgemeingültigen Empfehlungen lässt dabei zugleich eine zunehmende Berücksichtigung von BIAs für die HTA-Praxis vermuten (vgl. bei Foroutan et al., 2018). Während spezifische Leitlinien in Ländern wie Australien (PBAC, 2016), Belgien (Neyt et al., 2015), Kanada (Marshall et al., 2008) oder Polen (Orlewska and Mierzejewski, 2004) seit vielen Jahren fester Bestandteil von HTA (beziehungsweise Erstattungsentscheidungen) sind, spielt der BI in anderen Nationen, wie beispielsweise in Deutschland (vgl. Kapitel 2.3.1) oder den USA, eine bislang eher untergeordnete Rolle. In wiederum anderen Ländern wie England (NICE, 2017; vgl. Kapitel 2.3.2) oder Frankreich (Ghabri et al., 2018; vgl. Kapitel 2.3.3) etablierten sich erst in den letzten Jahren nationale Guidelines, die dem BI eine zumindest gewichtigere Bedeutung innerhalb des HTA-Prozesses zukommen lassen. Auf internationaler Ebene sind insbesondere die International Society for Pharmacoeconomics and Outcomes Research (ISPOR) Task Force Reports I (Mauskopf et al., 2007) und II (Sullivan et al., 2014) hervorzuheben, die als langjährig etablierte Good-Practice-Leitfäden zu einer Vereinheitlichung der Anforderungen an die Durchführung von BIAs<sup>26</sup> geführt und deren Umsetzung in die Praxis verbessert haben.

Sowohl GBA/IQWiG (G-BA, 2020; IQWiG, 2020) in Deutschland als auch NICE (NICE, 2018) in England verlangen von den pUs bei Einreichung der Bewertungsunterlagen die Durchführung von BIAs, die für den einzelnen Evaluationsprozess jedoch nur begrenzt herangezogen werden (vgl. hierzu in Kapitel 2.3). Demzufolge lässt auch der empirische Forschungsstand weitgehend offen, inwieweit und in welchem Ausmaß BIAs (trotzdem) in die Empfehlungen und Entscheidungen der Institutionen einfließen (könnten). Länderspezifische Analysen und Vergleichsstudien zu HTAs in Deutschland und England setzten sich bisher primär mit methodischen Vorgehensweisen und den (offiziellen) Bewertungskriterien sowie daraus resultierenden Empfehlungen und Entscheidungsfindungen auseinander, weshalb Untersuchungen zu möglichen Zusammenhängen zwischen dem (potenziellen) BI und den Bewertungsergebnissen durch NICE beziehungsweise GBA/IQWiG nur vereinzelt vorliegen.

Bislang berichteten zwei empirische Studien relevante Ergebnisse im Zusammenhang mit dem BI als möglichen Einflussfaktor der Empfehlungen durch NICE (vgl. hierzu auch Kapitel 6.2.2). Die Arbeit von Dakin et al. (2006) stellte dabei keine signifikante Auswirkung durch den BI

---

<sup>26</sup> Eine ausführliche Darstellung zur technischen Umsetzung beziehungsweise Durchführung von BIAs findet sich darüber hinaus bei Mauskopf et al. (2017).

fest, jedoch zeigen die für eine eingeschränkte respektive optimierte Anwendung empfohlenen Technologien einen signifikant höheren BI auf als die durch NICE uneingeschränkt positiv bewerteten Interventionen. Eine Analyse von Mauskopf et al. (2013) zeigte einen signifikant negativen Zusammenhang zwischen dem BI und den Erstattungsrestriktionen durch NICE, die auch nach Kontrolle weiterer (offizieller) Einflussfaktoren wie der Kosteneffektivität und der klinischen Wirksamkeit vorzufinden gewesen ist.

Zur (potenziellen) Rolle des BI innerhalb des deutschen HTA-Kontexts wurden bisher keine weiterführenden Studienergebnisse veröffentlicht (vgl. hierzu auch Kapitel 6.2.1). Lediglich eine Arbeit von Fischer *and* Stargardt (2014) identifizierte einen signifikanten Zusammenhang von positiven GBA-Beschlussfassungen und (sehr) hohen jährlichen Behandlungskosten pro Patient, dabei konnte jedoch kein eindeutiges Muster für den (potenziellen) BI festgestellt werden. Sie schlussfolgerten vielmehr, dass sowohl die jährlichen Behandlungskosten pro Patient als auch der maximal mögliche BI in den pU-Dossiers gegenüber den (geschätzten) Angaben aus den GBA-Beschlusstexten als signifikant höher einzuschätzen sind (Fischer *and* Stargardt, 2014). Vergleichbar den Resultaten von Fischer *and* Stargardt (2014) konnten auch Dintsios et al. (2019), die den Schwerpunkt ihrer Untersuchung insbesondere auf Verfahren mit nachträglich veröffentlichten Addenda legten, keinen signifikanten Einfluss des BI auf die Nutzenbewertungen durch GBA und IQWiG feststellen.

Vor diesem Hintergrund erscheinen die Kostenfolgen innovativer Gesundheitstechnologien von besonderem Interesse für eine genauere Betrachtung der methodischen Umsetzung sowie den (offiziellen) HTA-Outcomes von sowohl GBA/IQWiG als auch NICE. Zugleich wird jedoch auch erkennbar, dass der BI keinesfalls isoliert sondern ausschließlich im Zusammenhang mit anderen Kriterien innerhalb der nationalen Bewertungs- und Entscheidungsprozesse berücksichtigt werden kann.

### 3 ZIELSETZUNG UND AUFBAU

Im Rahmen der vorliegenden Dissertation werden die abgeschlossenen HTAs von GBA und IQWiG in Deutschland und NICE in England für den Zeitraum von Januar 2011 bis einschließlich Juni 2018 systematisch untersucht. Im Mittelpunkt stehen die veröffentlichten HTA-Resultate, die Aufschluss über die Implikationen der unterschiedlichen methodischen Vorgehensweisen und zugehörigen Einflussfaktoren bei der Bewertung durch GBA/IQWiG und NICE geben sollen. Ein gesonderter Schwerpunkt liegt auf den unterschiedlichen Indikationsgebieten, insbesondere den onkologischen und hämatologischen Interventionen, um möglichst weitreichende Aussagen zu indikationsspezifischen HTA-Ergebnissen und möglicherweise zugrunde liegenden Besonderheiten treffen zu können. Mit einer vergleichenden Untersuchung der definitionsgemäß übereinstimmenden HTAs von GBA und NICE – unter zusätzlicher Hinzuziehung der HAS in Frankreich – sollen darüber hinaus Unterschiede aufgezeigt und hinreichend spezifiziert werden, um die Aussagekraft der voneinander unabhängigen Analyseergebnisse von GBA/IQWiG und NICE in einen gesamtheitlichen Kontext stellen zu können. Davon ausgehend leitet sich die zentrale Forschungsfrage der Arbeit wie folgt ab:

**Welche Bewertungsfaktoren – unter besonderer Berücksichtigung der Kostenfolgen (Budget Impact) – beeinflussen die HTA-Resultate von GBA/IQWiG in Deutschland und NICE in England?**

Zum einen werden die Zusammenhänge der jeweils – offiziellen oder empirisch als relevant eingestuften – Bewertungskriterien und der HTA-Outcomes analysiert, um zu einem grundlegend besseren Verständnis sowie einem konstruktiven Diskurs von HTA in der Praxis, insbesondere in Deutschland und England, beitragen zu können. Zum anderen wird die potenzielle Rolle der Kostenfolgen mittels unterschiedlicher BI-Szenarien herausgestellt, um mögliche Rückschlüsse aus den Bewertungsergebnissen respektive Empfehlungen ziehen und damit den gegenwärtigen Status – auch im Hinblick auf zukünftige Anwendungsmöglichkeiten der BIA – diskutieren zu können.

Die Dissertation gliedert sich in insgesamt sieben Abschnitte. Auf die thematische Einführung in Kapitel 1 folgte in Kapitel 2 eine umfassende Betrachtung des Instruments HTA, mit einem gesonderten Fokus auf die BIA und deren Stellenwert innerhalb des HTA-Prozesses. Darüber hinaus wurden die relevanten HTA-Institutionen mit den jeweiligen nationalen Evaluationsansätzen vorgestellt, die den wissenschaftlichen Hintergrund für die im vorliegenden Kapitel 3 definierten Forschungsziele bilden. Im nachfolgenden Kapitel 4 wird die Methodik beschrieben, untergliedert nach einerseits der vergleichenden Untersuchung übereinstimmender HTAs (von GBA, HAS und NICE) und andererseits den spezifischen Regressionsmodellen für GBA/IQWiG und NICE. Daraufhin werden in Kapitel 5 sowohl die Resultate der vergleichenden Untersuchung als auch die einzelnen Analyseergebnisse für GBA und IQWiG in Deutschland und NICE in England dargestellt. Die daran anschließende Diskussion in Kapitel 6 konzentriert sich auf die Herausstellung und Reflektion der wesentlichen Erkenntnisse, die in einen wissenschaftlichen sowie praktischen Kontext von HTA gestellt werden. Abschließend folgt in Kapitel 7 eine Zusammenfassung der Arbeit.



## 4 DATEN UND METHODEN

Auf Grundlage des theoretischen Hintergrunds, dessen inhaltlicher Schwerpunkt auf den offiziellen HTA-Institutionen und den nationalen Evaluationsverfahren lag, wird nachfolgend die Extraktion aller relevanten Daten<sup>27</sup> für GBA und IQWiG in Deutschland, NICE in England und HAS in Frankreich dargestellt. Daran anschließend folgt eine ausführliche Darlegung der verwendeten Methodik für die vergleichende Untersuchung sowie die statistischen Analysen.

### 4.1 Quellen und Datenerfassung

#### 4.1.1 Deutschland: GBA und IQWiG

Für GBA und IQWiG sind ausschließlich Verfahren der FNB<sup>28</sup> berücksichtigt worden, die im Untersuchungszeitraum von Januar 2011 bis einschließlich Juni 2018 abgeschlossen und zugleich auf der offiziellen Website des GBA (G-BA, 2019) veröffentlicht wurden. Die primäre Voraussetzung hierfür ist gewesen, dass der abschließend gültige GBA-Beschlusstext vorlag. Zudem sollten in aller Regel die erforderlichen Dokumente (IQWiG-Dossierbewertung und pU-Dossier), die – bei einer Beauftragung der Nutzenbewertung durch IQWiG – Teil eines vollständigen FNB-Verfahrens sind, online vollumfänglich zugänglich gewesen sein. Folglich sind alle aufgehobenen, zurückgezogenen oder nicht mehr aktuellen FNBS<sup>29</sup> unberücksichtigt geblieben. Auch bei erneuter Bewertung eines Wirkstoffes für dieselbe Indikation ist immer nur die aktuelle respektive gültige GBA-Beschlussfassung sowie IQWiG-Nutzenbewertung<sup>30</sup> herangezogen worden. Darüber hinaus wurden relevante Informationen aus den zugehörigen pU-Dossiers der einzelnen Verfahren extrahiert. Im Fokus stand dabei das jeweilige Modul 1, welches die zusammenfassenden Bewertungsaussagen aller eingereichten Module beinhaltet.

---

<sup>27</sup> Alle relevanten Daten und Informationen der identifizierten sowie berücksichtigten HTAs von GBA/IQWiG, HAS und NICE wurden im Zeitraum von August 2018 bis einschließlich Januar 2019 extrahiert und daran anschließend auf Vollständigkeit sowie stichprobenartig auf ihre Richtigkeit geprüft.

<sup>28</sup> FNBS, die von Januar 2011 bis einschließlich April 2015 abgeschlossen und öffentlich zugänglich gemacht wurden, basieren auf einer früheren Bearbeitung. Nach nochmaliger Prüfung und entsprechender Aktualisierung im Juli 2018 wurden die Untersuchungsergebnisse veröffentlicht (vgl. Schaefer and Schlander, 2019).

<sup>29</sup> FNBS, die während des vorliegenden Studienzeitraums eingeschlossen und zu einem späteren Zeitpunkt, das heißt nach Juni 2018, aufgehoben oder zurückgezogen wurden, sind weiterhin Bestandteil der vorliegenden Analysen und Untersuchungen.

<sup>30</sup> In der Regel liegen für alle durchgeführten FNBS nach AMNOG Dossierbewertungen durch IQWiG vor. Für Wirkstoffe mit einem OD-Status, die zugleich jährliche GKV-Kosten von 50 Mio. EUR nicht überschreiten, wurden demzufolge lediglich die Zielpopulation innerhalb der GKV sowie die jährlichen Therapiekosten bewertet.

Die folgenden Daten wurden aus den einzelnen GBA-Beschlussfassungen und IQWiG-Dossierbewertungen, sowie ergänzend aus den pU-Dossiers, extrahiert:

- Wirkstoff
- Veröffentlichungsdatum
- Therapeutisches Gebiet
- Zugelassenes Anwendungsgebiet
- OD-Status
- Zusatznutzen nach Wahrscheinlichkeit und Ausmaß:
  - Wahrscheinlichkeit: Anhaltspunkt, Hinweis, Beleg
  - Ausmaß: Mit Zusatznutzen: erheblich, beträchtlich, gering, nicht quantifizierbar
  - Ohne Zusatznutzen: nicht vorhanden, geringer
- Klinische Studienevidenz
  - => Vorhandensein relevanter RCTs
- Vergleichende klinische Wirksamkeit
  - => Vorhandensein signifikanter Unterschiede in den patientenrelevanten Endpunkten: Morbidität, Mortalität, HRQoL (Nebenwirkungen; *wenn akzeptiert*: Surrogatendpunkte)
- Vergleichstherapie
  - => Vorhandensein einer ZVT
- Patientenpopulation:
  - GKV-Zielpopulation mit Zusatznutzen
  - GKV-Teilpopulation ohne Zusatznutzen
  - Gesamte theoretisch behandelbare Population in der GKV mit und / oder ohne Zusatznutzen (GKV-)Zielpopulation aus pU-Dossiers [*zusätzliche Daten*]
  - Plus*: Extraktion der ZVT-Daten
- Jahrestherapiekosten pro Patient
  - Jahrestherapiekosten (insgesamt) aus pU-Dossiers [*zusätzliche Daten*]
  - Plus*: Extraktion der ZVT-Daten.

Bei der Patientenpopulation handelt es sich um die Prävalenzdaten, die in den GBA-Beschluss-texten und IQWiG-Nutzenbewertungen beziehungsweise pU-Dossiers angegeben werden. Dies gilt sowohl für die Daten zur GKV-Zielpopulation (mit Zusatznutzen) als auch für die GKV-Teilpopulationen (ohne Zusatznutzen) sowie die Gesamtanzahl der theoretisch behandelbaren

Population innerhalb der GKV (das heißt die theoretisch maximal behandelbare Patientenzahl mit und / oder ohne Zusatznutzen).

GBA und IQWiG definieren die Subgruppen von GKV-Patientenpopulationen teilweise unterschiedlich, wodurch Diskrepanzen in ihrer jeweiligen Anzahl für eine bestimmte Intervention vorliegen können. Um folglich eine konsistente und möglichst einheitliche Darstellung beim Vergleich und bei der Analyse der GBA-Beschlüsse sowie IQWiG-Bewertungen sicherzustellen, wurden – sofern vorhanden – die durch IQWiG zusätzlich gebildeten Subgruppen (in aller Regel basierend auf den relevanten Studien aus den pU-Dossiers) zusammengefasst, sodass sie den durch den GBA definierten Teilpopulationen eindeutig zugeordnet werden konnten. Für den Fall, dass unterschiedliche HTA-Ergebnisse innerhalb der betroffenen Subgruppen festgestellt werden konnten, wurde auf die positivste Empfehlung durch IQWiG zurückgegriffen.

Die Jahrestherapiekosten pro Patient für die neue Therapie und die ZVT werden grundsätzlich sowohl in den GBA-Beschlussfassungen als auch in den IQWiG-Dossierbewertungen angegeben. Die in den Nutzenbewertungsdossiers angegebenen Kosten beruhen (zumeist) direkt auf den Angaben aus den pU-Dossiers. IQWiG prüft und kommentiert diese Schätzungen, indem sie als plausibel, zu niedrig oder zu hoch eingeschätzt werden. In einzelnen Fällen beruhen die Angaben auf institutionseigenen Berechnungen und Einschätzungen, die entsprechend in der Nutzenbewertung angegeben werden.

#### **4.1.2 England: NICE**

Für NICE sind ausschließlich abgeschlossene STAs<sup>31</sup> berücksichtigt worden, die entsprechend von Januar 2011 bis einschließlich Juni 2018 auf der offiziellen Website (NICE, 2019) veröffentlicht wurden. Seit April 2017 sind zusätzlich die FTAs herangezogen worden, sofern die Bewertungsunterlagen in vergleichbarer Form wie bei einem vollständigen STA zugänglich waren. STA Guidance<sup>32</sup>, die bis zum Ende des Untersuchungszeitraums aufgehoben oder durch ein pU zurückgezogen wurden, sind nicht Bestandteil der Bearbeitungen. Demgegenüber wur-

---

<sup>31</sup> STAs, die von Januar 2011 bis einschließlich April 2015 abgeschlossen und öffentlich zugänglich gemacht wurden, basieren auf einer früheren Bearbeitung. Nach nochmaliger Prüfung und entsprechender Aktualisierung im Juli 2018 wurden die Untersuchungsergebnisse veröffentlicht (vgl. Schaefer and Schlander, 2019).

<sup>32</sup> STAs (oder entsprechend berücksichtigte FTAs), die während des Studienzeitraums eingeschlossen und zu einem späteren Zeitpunkt, das heißt nach Juni 2018, aufgehoben oder zurückgezogen wurden, sind weiterhin Bestandteil der vorliegenden Analysen und Untersuchungen.

den erneut bewertete oder aktualisierte STAs sowie hinzugekommene Anwendungsbewertungen bereits vorgenommener Technologiebewertungen miteinbezogen. Neben den Guidance wurden, sofern durch NICE veröffentlicht, sowohl auf die 'Costing Templates' als auch die 'Resource Impact Statements' zurückgegriffen, die wesentliche Informationen zu Patientenpopulationen und Kosten von positiv bewerteten Technologien beinhalten. Die zugrunde liegenden ERG Reports wurden ausschließlich hinzugezogen, wenn relevante Daten fehlten. Die eingereichten Unterlagen der pUs wurden nicht weiter durchsichtet, da einzelne relevante Inhaltselemente nicht vollumfänglich zugänglich gewesen sind. Insbesondere die Angaben zu Kostendaten werden bei der Veröffentlichung aufgrund vorliegender Vertraulichkeitsvereinbarungen geschwärzt (vgl. hierzu bei Bullement et al., 2019).

Die folgenden Daten wurden aus den NICE STA Guidance (inklusive der zugehörigen Costing Templates und / oder Resource Impact Statements) und, sofern erforderlich, aus den ERG Reports extrahiert:

- Wirkstoff (Technology)
- Veröffentlichungsdatum (Publication date)
- Indikationsgebiet (Topic area / Category)
- Zugelassenes Anwendungsgebiet (Treatment option)
- Empfehlung (Recommendation):
  - empfohlen (recommended)
  - restriktiv *empfohlen* (optimized/ restricted)<sup>33</sup>
  - nicht empfohlen (not recommended)
- Kosteneffektivitäts-Verhältnis (ICER per QALY gained)
- Klinische Wirksamkeit (Clinical effectiveness)
- Klinische Studienevidenz (Clinical study evidence)
  - => Vorhandensein relevanter RCTs
- EoL-Kriterien (EoL criteria)
  - => Erfüllung der Kriterien [*sofern ausdrücklich berücksichtigt*]
- CDF (CDF reconsideration/ reimbursement)
  - => Erstattung negativ bewerteter Krebstherapien durch den CDF [*seit Juli 2016*]

---

<sup>33</sup> Die Einteilung der 'Empfehlungen' durch NICE erfolgte anhand spezifischer Kriterien (vgl. hierzu bei Raftery, 2006). Die Variable 'restriktive Empfehlung' (optimized/ restricted) wird für die Regressionsanalyse als separate Outcome-Variable berücksichtigt, um Einschränkungen respektive Optimierungen, die zu Unterschieden innerhalb der (potenziellen) Patientenpopulation einer Indikation führen, darstellen zu können (vgl. Kapitel 4.3.2).

- Vergleichstherapie (Comparator drug)
- Patientenpopulation (Patient population):
  - Geeignete, erstattungsfähige Behandlungspopulation innerhalb des NHS  
(Reimbursed-eligible population for treatment)
  - Geeignete, nicht erstattungsfähige Zielpopulation innerhalb des NHS  
(Target population eligible for treatment)
  - Geeignete (erstattungs- und nicht-erstattungsfähige) Zielpopulation innerhalb des NHS  
(Target population for treatment)
  - Gesamte theoretisch behandelbare/ geeignete NHS-Population mit vorliegender Indikation  
(Total number of NHS patients with the condition of interest)
  - Plus:* Extraktion der Vergleichstherapiedaten => mögliche Hinzunahme ERG Report
- Jährliche Arzneimittelkosten pro Patient  
(Annual drug acquisition cost per patient)
- Jährliche Behandlungskosten pro Patient [*sofern ausgewiesen*]  
(Annual treatment cost per patient)
- Plus:* Extraktion der Vergleichstherapiedaten => mögliche Hinzunahme ERG Report.

Die Zuordnung der einzelnen Indikationsgebiete durch NICE ist durchweg einheitlich definiert gewesen. Lediglich HTAs zu malignen hämatologischen Neoplasien (Blood and bone marrow cancer) wurden nochmals geprüft, damit die Interventionen für Krebserkrankungen dem Indikationsgebiet der 'Onkologie' eindeutig zugeordnet werden konnten.

Sofern kein ICER-Wert in den veröffentlichten NICE Guidance ausgewiesen wurde, mussten weitere, zugehörige Informationen miteinbezogen werden. Dies ist zum Beispiel der Fall gewesen, wenn NICE ausschließlich angab, dass die ICER einer bewerteten Intervention im kosteneffektiven respektive positiv zu bewertenden Grenzbereich gelegen hatte (daraufhin getroffene Annahme:  $ICER < 20.000 \text{ GBP/QALY}$ ). Bei mangelhafter Datenlage und zugleich fehlenden Informationen zu einer plausiblen ICER-Schätzung durch NICE wurde erweiternd auf den jeweiligen ERG Report zurückgegriffen. Bei einer ebenfalls unzureichenden Datenlage im ERG Report blieb der Bewertungsfaktor 'Kosteneffektivität' unberücksichtigt (entsprechendes Vorgehen bei anderen Faktoren). EoL-Kriterien sind ausschließlich herangezogen worden, wenn deren Berücksichtigung in der NICE Guidance explizit angegeben wurde. Dasselbe gilt entsprechend für die Berücksichtigung des CDF, wodurch eine durch NICE negativ bewertete Krebstherapie für einen begrenzten Zeitraum durch den NHS erstattet werden kann.

Bei den Patientenpopulationsdaten sind auf Inzidenzebene die erstattungsfähige NHS-Zielpopulation (mit Empfehlung) und die NHS-Zielpopulation (ohne Empfehlung) berücksichtigt worden. Sofern eindeutig durch NICE ausgewiesen, sind zusätzlich die Angaben zu den NHS-Patienten herangezogen worden, die für die Technologie innerhalb des NHS indiziert wurden und damit (theoretisch) für eine Behandlung in Frage kommen könnten (NHS patients with the condition of interest). Die Darstellung auf Indikationsebene basiert auf der NHS-Zielpopulation mit und / oder ohne Empfehlung (erstattungs- und / oder nicht-erstattungsfähig). Daten zur theoretisch maximal behandelbaren, prävalenten Patientenzahl sind zusätzlich extrahiert worden, wenn durch NICE entsprechend ausgewiesen. Bei Krebstherapien konnten dabei ausschließlich Inzidenzdaten herangezogen werden – das heißt, die theoretisch maximal behandelbare, inzidente Patientenpopulation, da Prävalenzdaten für gewöhnlich nicht zur Verfügung stehen.

Bei den Kostendaten ist zwischen den jährlichen Arzneimittelkosten pro Patient und den jährlichen Behandlungskosten pro Patient zu unterscheiden. Letztere werden nicht immer angegeben, weshalb sich im weiteren Verlauf der Bearbeitungen ausschließlich auf die jährlichen Arzneimittelkosten pro Patient bezogen wird. Die monatlich oder pro Behandlungszyklus angegebenen Kosten sind (sofern möglich) immer auf Jahresbasis abgebildet worden. Für die monatlichen Angaben wurde dabei von vier Wochen beziehungsweise 28 Tagen ausgegangen, die im Regelfall auf den standardisierten Werten für eine durchschnittliche (männliche) Person<sup>34</sup> beruhen. Steuerliche Abgaben (Value Added Taxes, VAT) bleiben dabei unberücksichtigt, da sie für gewöhnlich nicht in den von NICE angegebenen Kostendaten beinhaltet sind. Konkrete Behandlungskosten (wie in der Guidance oder dem ERG Report spezifiziert) sind nur dann unverändert übernommen worden, wenn eine Krebstherapie eine bestimmte Anzahl an Therapiezyklen für ein spezifisches Krankheitsbild auswies.

#### **4.1.3 Frankreich: HAS**

Als Datengrundlage für die ACTs wurde die offizielle Website der HAS (HAS, 2019) herangezogen. In der Regel liegt für jede Teilindikation eines Wirkstoffes eine eigene Bewertung vor, für die sowohl eine französische Volltextversion als auch eine englische Kurzfassung veröffentlicht wird. Sofern durch die CEESP zusätzlich vorgenommen, wurden die zugehörigen Effizienzbewertungen (Avis d'Efficiency) ebenfalls herangezogen.

---

<sup>34</sup> Die standardisierten Werte für eine (männliche) Person entsprechen 70 Kilogramm und 1,73 Quadratmeter / Standard-Körperoberfläche.

Im Gegensatz zu GBA/IQWiG und NICE sind die Daten der HAS ausschließlich für die vergleichende Untersuchung berücksichtigt worden. Dies ist einerseits auf die empirische Vorarbeit zu den offiziellen HTAs von GBA/IQWiG und NICE zurückzuführen (vgl. Schaefer and Schlander, 2019). Andererseits liegen die (primäre) Zielsetzung und damit der thematische Schwerpunkt der Arbeit auf der Analyse zweier methodisch (weitgehend) heterogener HTA-Systeme, weshalb die offiziellen Bewertungen der HAS eine inhaltliche Ergänzung darstellen. Vor diesem Hintergrund wurde ausschließlich nach den durch die HAS veröffentlichten ACTs<sup>35</sup> gesucht, die beim systematischen Abgleich der GBA FNBs und der NICE STAs identifiziert werden konnten. Folgende Daten wurden dabei aus den identifizierten HAS ACTs, der offiziellen Website des französischen Gesundheitsministeriums (Solidarites-sante, 2019) sowie gegebenenfalls erweiternd aus der Effizienzbewertung des CEESP extrahiert:

- Wirkstoff (Ingrédient)
- Veröffentlichungsdatum (Date)
- Indikationsgebiet (Indication)
- Zugelassenes Anwendungsgebiet inklusive OD-Status (Indications thérapeutiques)
- Empfehlung der Kommission (Recommandations de la Commission)
- SMR inklusive zugehöriger Erstattungsrate (Taux de remboursement):  
Hinreichender klinischer Nutzen: erstattungsfähig ( $\leq 100\%$ )  
=> gering, 15 %; moderat, 30 %; beträchtlich, 65 %; chronisch / schwer erkrankt, 100 %  
Unzureichender klinischer Nutzen: nicht erstattungsfähig (0 %)
- ASMR (Level I–V):  
Mit klinischem Zusatznutzen: I, erheblich; II, beträchtlich; III, moderat; IV, gering  
Ohne klinischen Zusatznutzen: V, nicht vorhanden
- Klinische Effektivität / Primärer klinischer Endpunkt (Efficacité)<sup>36</sup>
- Art der Behandlung (Type de traitement)
- Vergleichstherapien (Compareurs)

---

<sup>35</sup> ACTs, die nach dem Abgleich mit den GBA FNBs und NICE STAs eingeschlossen und zu einem späteren Zeitpunkt, das heißt nach Juni 2018, aufgehoben oder zurückgezogen wurden, sind weiterhin Bestandteil der vergleichenden Untersuchung.

<sup>36</sup> Ergänzend zur Bewertung der klinischen Wirksamkeit (anhand klinisch relevanter Endpunkte) werden in der Regel weitere Bewertungsfaktoren (Vorhandensein relevanter RCTs, Verhältnis Wirksamkeit zu Nebenwirkungen, Krankheitsschweregrad, Vorhandensein therapeutischer Alternativen, Nutzen für die öffentliche Gesundheit) sowie die Rolle des neuen Medikaments innerhalb der therapeutischen Strategie berücksichtigt (HAS, 2020).

*Ergänzende Daten, um die Vergleichbarkeit mit GBA FNBs / NICE STAs zu gewährleisten:*

- **Patientenpopulation (Population):**  
Zielpopulation (Population cible)  
Krankenversicherte Zielpopulation gesamt (Population cible totale de l'assurance maladie)
- **Jährliche Sozialversicherungskosten je Arzneimittel pro Patient (Prix honoraire compris):**  
Berechnung aus dem Preis einer Verpackungseinheit mit der jeweiligen Tageseinheit und / oder der Anzahl je Packung unter Berücksichtigung der Erstattungsrate pro Jahr
- *Plus:* Optionale Bewertung der Kosteneffektivität durch CEESP (Avis d'Efficiencie).

Die HAS gibt die Zielpopulation in der Regel auf Inzidenzebene an und weist Angaben zur theoretischen Gesamtzahl der behandelbaren Patientinnen und Patienten auf Sozialversicherungsebene aus. Demgegenüber werden die Daten für die Berechnung der Kosten nicht explizit in der HAS ACT angegeben, sondern stammen direkt von der Website des französischen Gesundheitsministeriums (Solidarites-sante, 2019). Damit unterliegen die (teilweise eigens) vorgenommenen Kostenberechnungen entsprechenden Unsicherheiten, da die Angaben nicht durchgängig identisch übertragen werden konnten und folglich Anpassungen vorzunehmen waren, um möglichst konkrete (Schätz-)Werte für die Medikamentenkosten pro Jahr zu erhalten.

Alle Medikamente, die von Januar 2011 bis einschließlich Juni 2018 in Form einer ACT bewertet und als relevant eingestuft wurden, sind dabei berücksichtigt, auch wenn die abschließenden Dokumente erst nach dem Untersuchungszeitraum (bis einschließlich Dezember 2018, das heißt bis vor Abschluss der Datenerfassung) veröffentlicht worden sind. Eine Bewertung der Kosteneffektivität durch das zuständige CEESP konnte ausschließlich dann identifiziert werden, wenn in den eingereichten Unterlagen eines pUs eine positive Bewertung (ASMR I–III) beansprucht wurde und darüber hinaus anzunehmen gewesen ist, dass das neue Medikament einen erheblichen Einfluss auf die Gesamtausgaben der Versicherer haben würde.

## **4.2 Vergleichende Untersuchung der HTA-Resultate**

Die vergleichende Untersuchung von GBA, HAS und NICE fokussiert sich ausschließlich auf HTA-Resultate übereinstimmender Behandlungsinterventionen vergleichbarer Indikations- respektive Anwendungsgebiete. Mittels einer tabellarischen Gegenüberstellung der im vorliegenden Untersuchungszeitraum extrahierten GBA FNBs und NICE STA Guidance – sowie da-



ran anschließend der HAS ACTs – konnten die relevanten Interventionen respektive Behandlungsindikationen eindeutig bestimmt werden. Sofern erforderlich wurden zusätzlich Informationen zur möglichen Anwendung eines Medikaments in Kombination mit anderen Arzneimitteln, zu genetischen (Vor-)Bedingungen in der vorliegenden Zielgruppe der Patientinnen und Patienten oder notwendigen Vorbehandlungen (Erst- oder Zweitlinientherapie) berücksichtigt. Für einen einheitlichen, systematischen Abgleich der übereinstimmenden Interventionspaare wurden die vom GBA definierten therapeutischen Gebiete herangezogen, die sich – wiederum vergleichbar zu NICE und HAS – in insgesamt 14 übergeordnete Indikationen unterteilen lassen (vgl. G-BA, 2020): Augenerkrankungen, Hauterkrankungen, Herz-Kreislauf-Erkrankungen, Infektionserkrankungen, Krankheiten des Atmungssystems, Krankheiten des Blutes / der blutbildenden Organe, Krankheiten des Muskel-Skelett-Systems, Krankheiten des Nervensystems, Krankheiten des Urogenitalsystems, Krankheiten des Verdauungssystems, psychische Erkrankungen, Stoffwechselkrankheiten und Sonstiges (allesamt nicht-onkologische Krankheitsbilder), sowie onkologische inklusive hämatologischer Krankheitsbilder (Krebserkrankungen).

Bei der vergleichenden Analyse der HTA-Ergebnisse übereinstimmender Behandlungstechnologien wurde sich primär auf drei übergeordnete (Indikations-)Ebenen konzentriert: Gesamt, onkologisch und nicht-onkologisch. Den wesentlichen Indikator für den Abgleich der Interventionen stellt der Anteil der positiven HTA-Ergebnisse (GBA, mit Zusatznutzen; HAS, mit klinischem Zusatznutzen: ASMR I-IV; NICE, mit Empfehlung: empfohlen / restriktiv) innerhalb der jeweiligen Indikation dar. Um Zusammenhänge zwischen Indikationsgebieten (onkologisch, nicht-onkologisch) und positivem HTA-Outcome der jeweiligen Institution untersuchen zu können, wurde das Chancenverhältnis (Odds Ratio, OR<sup>37</sup>) auf Grundlage von 2 x 2 Kontingenztabellen bestimmt. Dementsprechend wurde die Chance für das Auftreten einer positiven Bewertung durch eine HTA-Agentur mit der Wahrscheinlichkeit einer positiven Empfehlung der jeweils anderen verglichen. Die ORs sind immer mit den zugehörigen 95 %-Konfidenzintervallen (Confidence Intervals, CIs<sup>38</sup>) dargestellt worden, die durch das Potenzieren der loga-

---

<sup>37</sup> Das OR (Chancen- oder Quotenverhältnis) ist ein statistisches Maß, um die Stärke des Zusammenhangs (Effektstärke) von zwei Merkmalen beziehungsweise Variablen zu quantifizieren (ohne dabei auf die Kausalität schließen zu können). Ein  $OR > 1$  bedeutet, dass das Vorhandensein eines Merkmals die Wahrscheinlichkeit für das Vorhandensein des anderen Merkmals erhöht. Ein  $OR < 1$  sagt demnach aus, dass das Vorhandensein einer Variable die Wahrscheinlichkeit für das Vorhandensein der anderen Variable senkt (Sheskin, 2011).

<sup>38</sup> Ein CI (auch Vertrauensintervall genannt) ist ein Bereich, der die Präzision der Lageschätzung eines Parameters angeben soll. Es gibt das Intervall (beispielsweise von 95 %) an, welches mit einer gewissen Wahrscheinlichkeit den Parameter (zum Beispiel einen Mittelwert) der Verteilung einer Zufallsvariable einschließt (Sheskin, 2011).

rhythmischen Grenzen auf eine lineare Skala zurückberechnet wurden. Zudem wurde der entsprechende Signifikanzwert (p-Wert<sup>39</sup>) berechnet, für den  $p < 0,1$  als Grenzwert der statistischen Signifikanz definiert ist. Darüber hinaus ist die Übereinstimmung der HTA-Resultate aller drei Institutionen betrachtet worden, um die Abweichungen auf den verschiedenen Indikationsebenen untersuchen und damit den paarweisen Vergleich unterstützen zu können. Ebenfalls mittels ORs wurden zudem die Outcomes der drei Gebiete (gesamt, onkologisch, nicht-onkologisch) für GBA versus HAS (GBA / HAS), für GBA versus NICE (GBA / NICE) und für HAS versus NICE (HAS / NICE) bestimmt.

Vergleichbar der methodischen Vorgehensweise bei Schaefer et al. (2021) adressieren die übereinstimmenden Interventionspaare dieselbe oder zumindest eine vergleichbare Behandlungsindikation, das heißt es liegt ein vergleichbares Anwendungsgebiet vor. Dies wiederum bedeutet, dass die HTA-Ergebnisse von GBA, HAS und NICE durch die Erfassung der Hauptindikation einer Technologie berücksichtigt wurden, da sich die Behandlungen in Bezug auf die Definitionen der Patientensubgruppen oder ausgewählten Vergleichsintervention(en) unterscheiden können. Sollte mehr als eine Teilpopulation innerhalb der Hauptindikation vorgelegen haben, dann wurde jeweils die Patientengruppe mit dem positivsten Resultat als Indikation für den Abgleich herangezogen.

Darüber hinaus ist die Rolle möglicherweise relevanter Attribute in Bezug auf Onkologika – einschließlich der Therapien für hämatologische Erkrankungen – untersucht worden. Dazu zählen der OD-Status in Deutschland, das (implizite) Abhängigkeitsverhältnis von SMR-Erstattungsraten und ASMR-Empfehlungen in Frankreich sowie die Berücksichtigung von EoL-Kriterien und CDF-Erstattungen in England. Hierzu wurde der Zusammenhang zwischen den genannten Attributen und den jeweiligen HTA-Ergebnissen mit dem Chi-Quadrat-Unabhängigkeitstest<sup>40</sup> oder mit dem Exakten Test nach Fisher<sup>41</sup> auf statistische Signifikanz getestet.

---

<sup>39</sup> Der p-Wert (auch Irrtumswahrscheinlichkeit genannt) drückt bei statistisch signifikantem Ergebnis aus, dass eine angenommene Hypothese (für eine bestimmte Stichprobe) auch auf die Grundgesamtheit zutrifft und dabei über einem festgelegten Niveau (von beispielsweise 10 %) liegt. Demzufolge tritt ein gemessener Zusammenhang zwischen zwei Merkmalen oder Variablen in einer Stichprobe nicht zufällig auf, sondern trifft auch für die Grundgesamtheit zu (Weiß, 2019).

<sup>40</sup> Der Chi-Quadrat-Test beschreibt ein statistisches Testverfahren, das Aussagen über den Zusammenhang zwischen entweder nominal oder ordinal skalierten Variablen treffen kann (Weiß, 2019).

<sup>41</sup> Der Exakte Fisher-Test oder Fisher-Yates-Test ist ebenfalls ein Signifikanztest auf Unabhängigkeit in Kontingenztafeln und stellt folglich eine Alternative zum Chi-Quadrat-Test dar. Während der Chi-Quadrat-Test Voraussetzungen an den Stichprobenumfang stellt ( $N > 5$ ), kann der Fisher-Yates-Test auch bei einer geringen Anzahl von Beobachtungen aussagekräftige Resultate liefern (Weiß, 2019).

### 4.3 Regressionsmodelle zur Analyse der HTA-Kriterien und -Resultate

Die logistischen Regressionsmodelle (Logit-Modelle) zur Analyse der (offiziellen oder empirisch als relevant eingestuften) Bewertungsfaktoren und HTA-Ergebnisse für die GBA-Beschlüsse und IQWiG-Nutzenbewertungen in Deutschland einerseits und die NICE STA Guidance in England andererseits wurden unabhängig voneinander definiert, um die bestehenden Unterschiede und Besonderheiten der jeweiligen HTA-Verfahren und methodischen Vorgehensweisen sowie der verschiedenen BI-Szenarien explizit herausstellen zu können.

#### 4.3.1 GBA-Beschlussfassungen und IQWiG-Dossierbewertungen

Die aus den GBA-Beschlüssen und den IQWiG-Nutzenbewertungen<sup>42</sup> extrahierten Daten basieren auf den jeweils bewerteten Teilpopulation oder der jeweils einzigen evaluierten Indikation, die für einen neuen Wirkstoff identifiziert werden konnten. Neben der primären Analyse auf Teilpopulationsebene wurde zusätzlich die Verfahrensebene (respektive Hauptindikationsebene bei mehreren Subindikationen) analysiert, um die Aussagekraft der Untersuchungsergebnisse auf Subgruppenebene stützen und in einem weitergefassten Kontext darstellen zu können.

Sowohl für GBA als auch IQWiG wurden demnach dieselben Variablen (vgl. auch Regressionsgleichung [1]) verwendet, die in Tabelle 2 und Tabelle 3 ausführlich beschrieben sind. Das Ergebnis der (Zusatz-)Nutzenbewertung nach dem Ausmaß der Bewertungskategorie (mit Zusatznutzen: beträchtlich, erheblich, gering, nicht quantifizierbar; ohne Zusatznutzen: nicht vorhanden, geringer) stellt die abhängige Outcome-Variable dar. Die potenziellen Bewertungsfaktoren beschreiben demgegenüber die unabhängigen beziehungsweise erklärenden Variablen des (ordinalen) Logit-Modells.

Die relevanten Daten zur Patientenpopulation und zu den jährlichen Behandlungskosten pro Patient wurden sowohl für die neu bewertete Therapie als auch für die ZVT extrahiert (Annahme: Bei ODs sind die Kosten der ZVT aufgrund der fehlenden Angaben gleich null). Diese bezogen sich in der Regel auf einen Wertebereich (Minimum, Maximum), woraus jeweils der Mittelwert (Durchschnitt) berechnet wurde. Wenn kein Wertebereich vorlag, wurde der angegebene (oder plausibelste) Einzelwert (anstatt eines Durchschnittswerts) direkt aus den Referenzdokumenten von GBA und IQWiG herangezogen.

---

<sup>42</sup> Die IQWiG-Themengebiete wurden den therapeutischen Gebieten des GBA entsprechend zugeordnet, sodass die Indikationen für die Regressionsanalyse durchgängig übereinstimmten.

**Tabelle 2:** Definition der Regressionsvariablen (1) für GBA und IQWiG  
(Eigene Darstellung)

Variable	Kodierung	Definition
<i>Abhängige Variable:</i>		
Zusatznutzen -Beschluss (GBA) -Empfehlung (IQWiG)	0 = nicht beschlossen/ nicht empfohlen 1 = beschlossen/ empfohlen  2 <sup>#</sup> = eingeschränkt beschlossen / empfohlen	Beschluss/Empfehlung durch GBA/IQWiG, ob für eine bestimmte (Teil-)Population einer Intervention ein (positiver) Zusatznutzen gegenüber der ZVT besteht oder nicht besteht  Zusätzliche Beschluss-/Empfehlungsoption: eingeschränkt (positiver) Zusatznutzen [auf Verfahrensebene]
<i>Unabhängige Variablen:</i>		
OD-Status	0 = nein 1 = ja	Status zur Behandlung seltener Erkrankungen, der ein positives Nutzenergebnis attestiert
Jahr	Numerisch	Jahreszahl der veröffentlichten Beschlussfassung
Indikationsgebiet	Text	Zugeordnetes Indikationsgebiet durch GBA
ZVT <sup>§</sup>	0 = nicht vorhanden 1 = vorhanden	Daten und Informationen zur vom GBA festgelegten ZVT
RCT	0 = nicht vorhanden 1 = vorhanden	Vorlage relevanter RCTs in Dossiers (einzeln oder zusammengefasst), die berücksichtigt sind
Patientenrelevanter Endpunkt	0 = nicht vorhanden 1 = vorhanden	Vorhandensein eines patientenrelevanten Endpunktes (Mortalität, Morbidität, Lebensqualität) mit signifikantem Unterschied zur ZVT
Jahrestherapiekosten*	Numerisch i) <i>Minimum</i> ii) <i>Durchschnitt</i> iii) <i>Maximum</i>	Jährliche Behandlungskosten pro Patient in EUR, wenn angegeben für unteren / oberen Grenzbereich und / oder einen Einzel- / (ermittelten) Durchschnittswert
Patientenzahl* a) Zielpopulation b) Teilpopulation c) Zielpopulation gesamt	Numerisch i) <i>Minimum</i> ii) <i>Durchschnitt</i> iii) <i>Maximum</i>	Wenn angegeben, jeweils für i)–iii): a) GKV-Zielpopulation mit Zusatznutzen b) GKV-Teilpopulation ohne Zusatznutzen c) GKV-Zielpopulation gesamt mit / ohne Zusatznutzen [auf Verfahrensebene]

EUR, Euro; GKV, Gesetzliche Krankenversicherung; OD, Orphan Drug; pU, pharmazeutisches Unternehmen; RCT, Randomized Controlled Trial; ZVT, Zweckmäßige Vergleichstherapie.

<sup>#</sup>Modell-abhängige, dreistufige Variablenkodierung auf Verfahrensebene: Von der Patientenpopulation der (Haupt-)Indikation sind nur einzelne Teilpopulationen mit einem (positiven) Zusatznutzen bewertet worden.

<sup>§</sup>Definitionsgemäß liegen für alle bewerteten Wirkstoffe Angaben zur ZVT vor, außer es handelt sich um Ausnahme- respektive Sonderfälle sowie Arzneimittel mit einem (zum Zeitpunkt der Markteinführung vorliegenden) OD-Status.

\*Kontrollvariable zur Berechnung der unterschiedlichen Budget Impact-Szenarien.

Anmerkung: Vergleichbare Daten, darunter auch Patientenzahlen und Jahrestherapiekosten pro Patient (als Kontrollvariablen) sowie Jahrestherapiekosten insgesamt, wurden aus den jeweiligen pU-Dossiers extrahiert, stellen jedoch lediglich einen ergänzenden Bestandteil der Regressionsanalyse (IQWiG, Teilpopulationsebene) dar.

Basierend auf den jeweiligen Patientenpopulations- und Kostendaten<sup>43</sup> wurden die verschiedenen BI-Szenarien für GBA und IQWiG sowohl auf Teilpopulations- (ex-post Betrachtung) als auch auf Verfahrensebene (ex-ante Betrachtung) definiert: Drei Szenarien (Minimum, Durchschnitt, Maximum; Verfahrensebene: Durchschnitt) für den potenziellen BI sowohl der neuen Therapie als auch der ZVT, und ein Szenario für den inkrementellen BI (durchschnittlicher BI der neuen Therapie minus durchschnittlicher BI der ZVT). Zusätzlich wurden aus den pU-Dossiers die 'Jahrestherapiekosten insgesamt' extrahiert, die einem potenziellen BI entsprechen.

**Tabelle 3:** Definition der Regressionsvariablen (2) – Budget Impact-Szenarien für GBA und IQWiG (Eigene Darstellung)

Variable	Kodierung	Definition
<i>Abhängige Variable:</i>		
Zusatznutzen -Beschluss (GBA) -Empfehlung (IQWiG)	0 = nicht beschlossen/ nicht empfohlen 1 = beschlossen/ empfohlen	Beschluss/Empfehlung durch GBA/IQWiG, ob für eine bestimmte (Teil-)Population einer Intervention ein Zusatznutzen gegenüber der ZVT besteht
	2 <sup>#</sup> = eingeschränkt beschlossen / empfohlen	Zusätzliche Beschluss-/Empfehlungsoption: eingeschränkt (positiver) Zusatznutzen [auf Verfahrensebene]
<i>Unabhängige Variablen:</i>		
Potenzieller BI* –	Numerisch	Berechnung für <u>neue</u> und <u>aktuelle</u> Therapie:
a) Zielpopulation [ex post]	i) <i>Minimum</i>	a) GKV-Zielpopulation [i–iii] × Jahrestherapiekosten [i–iii]
b) Teilpopulation [ex post]	ii) <i>Durchschnitt</i>	b) GKV-Teilpopulation [i–iii] × Jahrestherapiekosten [i–iii]
c) Zielpopulation gesamt [ex ante]	iii) <i>Maximum</i>	c) GKV-Zielpopulation gesamt [i–iii] × Jahrestherapiekosten [i–iii]
Inkrementeller BI* –	Numerisch	Berechnung inkrementeller BI für <u>neue</u> Therapie:
a) Zielpopulation [ex-post]	ii) <i>Durchschnitt</i>	a) BI-GKV neue Therapie [ii] – BI-GKV aktuelle Therapie [ii]
b) Teilpopulation [ex-post]		b) BI-Teilpopulation neue Therapie [ii] – BI-Teilpopulation aktuelle Therapie [ii]
c) Zielpopulation gesamt [ex-ante]		c) BI-Gesamt neue Therapie [ii] – BI-Gesamt aktuelle Therapie [ii]

BI, Budget Impact; GKV, Gesetzliche Krankenversicherung; ZVT, Zweckmäßige Vergleichstherapie.

<sup>#</sup>Modell-abhängige, dreistufige Variablenkodierung auf Verfahrensebene: Von der Patientenpopulation der (Haupt-)Indikation sind nur einzelne Teilpopulationen mit einem (positiven) Zusatznutzen bewertet worden.

\*Szenario-abhängige Variable, um unterschiedliche Projektionen für den BI darzustellen.

<sup>43</sup> Sowohl für die Populationszahlen als auch die Kosten sowie demzufolge für alle BI-Szenarien wurde das arithmetische Mittel mit zugehöriger Standardabweichung (welche die durchschnittliche Abweichung aller gemessenen Werte vom Mittelwert angibt) und der Median berechnet. Alle Werte werden im weiteren Verlauf der Arbeit angegeben. Der Schwerpunkt liegt jedoch auf den prinzipiell präziseren Mittelwerten, deren geringere Robustheit (für die einzelnen Regressionsanalysen) zusätzlich getestet wurde.

Die Rolle respektive der Einfluss der Bewertungskriterien und des BI auf die HTA-Ergebnisse sind mittels logistischer Regressionsanalyse getestet worden (unter Verwendung von *Stata Statistical Software: Release 15*). Dabei wurden die von GBA und IQWiG veröffentlichten FNB-Ergebnisse als eine Funktion des BI und der festgelegten Bewertungsfaktoren unabhängig voneinander modelliert. Die zugehörige Regressionsgleichung sowohl für die GBA-Beschlüsse als auch die IQWiG-Bewertungen lautet entsprechend:

$$\text{Outcome}_{ij} = \beta_0 + \beta_1 * \text{Kriterien}_{ij} + \beta_2 * \text{Budget}_{ij} + \mu_i + \tau_i + \varepsilon_{ij} \quad [1]$$

Auf Teilpopulationsebene (Logit-Modell) stellt  $\text{Outcome}_{ij}$  eine kategoriale Variable (zweistufig) dar, die den Wert 1 zuweist, wenn GBA oder IQWiG die Therapie  $i$  für die Patientengruppe  $j$  mit Zusatznutzen bewerteten und die den Wert 0 zuweist, wenn kein Zusatznutzen durch GBA oder IQWiG angegeben wurde. Auf Verfahrensebene (ordinales Logit-Modell) kann  $\text{Outcome}_{ij}$  als eine dreistufige Variable zusätzlich den Wert 2 zuweisen, wenn GBA oder IQWiG nicht für alle Patientensubgruppen, sondern lediglich für einzelne Teilpopulationen einer Therapie einen positiven Zusatznutzen berichteten.

Die Variable  $\text{Kriterien}_{ij}$  umfasst die relevanten Bewertungsfaktoren, dazu gehören das Vorliegen einer ZVT (Vergleichstherapie) sowie das Vorhandensein relevanter RCTs (klinische Studienevidenz), Unterschiede in den patientenrelevanten Endpunkten (vergleichende klinische Wirksamkeit) und die Informationen zum OD-Status.  $\text{Budget}_{ij}$  setzt sich aus den Variablen zusammen, die die unterschiedlichen BI-Szenarien adressieren: Die Schätzungen zum potenziellen BI (Minimum, Durchschnitt, Maximum) sowie zum inkrementellen BI (Durchschnitt), jeweils gemessen in logarithmischer Form<sup>44</sup>. Erweiternd ist die Variable  $\text{Budget}_{ij}$  als ein quadratischer Term berücksichtigt worden, um die Nichtlinearität zwischen dem potenziellen BI und der Outcome-Variable kontrollieren zu können.  $\beta_s$  beschreibt dabei die jeweiligen Koeffizienten, die geschätzt werden mussten. Während Veröffentlichungsjahr  $\mu_i$  und therapeutisches Gebiet  $\tau_i$  als Indikator- beziehungsweise Dummy-Variablen eingeführt wurden, bezeichnet  $\varepsilon_{ij}$  den Fehlerterm. Schließlich wurde die Robustheit der geschätzten Koeffizienten getestet, indem die obersten fünf Prozent der BI-Schätzwerte unberücksichtigt blieben ('Ausreißerdiagnostik').

---

<sup>44</sup> Die logarithmische Form (oder Log-Form) lautet:  $\log(x + 1)$ . Bevor demnach der Logarithmus gebildet wird, ist entsprechend ein Wert von 1 hinzuzufügen, damit die Nullbeobachtungen (bei fehlenden Informationen zur ZVT) berücksichtigt werden können (wobei der Logarithmus von 0 nicht bestimmt wird).

Die abschließende Gegenüberstellung der GBA-Beschlussfassungen und der IQWiG-Dossierbewertungen basiert auf den extrahierten Daten sowie den jeweiligen Ergebnissen der Regressionsanalysen, die entsprechend ermöglichen sollten, potenziell vorhandene Unterschiede in den Resultaten und in der praktischen Anwendung der (offiziellen oder empirisch als relevant eingestuften) Bewertungsfaktoren aufzuzeigen und direkt vergleichen zu können.

#### **4.3.2 NICE Single Technology Appraisal (STA) Guidance**

Vergleichbar der Vorgehensweise für GBA und IQWiG basieren auch die für NICE extrahierten Daten auf den Subgruppen oder der jeweils einzelnen Indikation einer neuen Gesundheitstechnologie. Ergänzend zur primären Analyse auf Subpopulationsebene wurde für eine weitergefasste Betrachtungsebene ebenfalls die (Haupt-)Indikation untersucht, sofern explizit mehr als eine Subgruppe für eine Behandlungsintervention vorlag.

Für NICE wurden die Variablen aus den STA Guidance verwendet, die in Tabelle 4 und Tabelle 5 detailliert erläutert sind (vgl. auch Regressionsgleichung [2]). Fehlende Informationen wurden dabei, sofern erforderlich, durch die unabhängigen ERG Reports ergänzt. Die abschließende Empfehlung durch NICE (empfohlen, restriktiv, nicht empfohlen) stellt dabei die Outcome-Variable dar. In einer zweiten Modellierung sind die CDF-Erstattungen, die seit dem Relaunch im Juli 2016 einen unmittelbaren Bezug zum Zugang von neuen Krebstechnologien aufweisen, zusätzlich miteinbezogen worden (nicht empfohlen/ erstattet => empfohlen). Als unabhängige, erklärende Variablen wurden die Evaluationsfaktoren berücksichtigt, die möglicherweise einen Einfluss auf die abhängige Outcome-Variable hatten.

Sowohl für die neue Technologie als auch für die Vergleichsintervention wurden die relevanten Daten zur Patientenpopulation und zu den jährlichen Arzneimittelkosten pro Patient berücksichtigt. Sofern angegeben, wurde der angegebene Wertebereich (Minimum, Maximum) einbezogen und daraus der Mittelwert (Durchschnitt) berechnet. Ansonsten wurde ein (plausibelster) Einzelwert (als Durchschnittswert) im zugrunde liegenden Referenzdokument herangezogen. Die verschiedenen BI-Szenarien wurden basierend auf den jeweiligen Patientenpopulations- (NHS-Zielpopulation, ex-post Betrachtung; gesamte theoretisch geeignete/ behandelbare NHS-Population, ex-ante Betrachtung) und Kostendaten<sup>45</sup> berechnet:

---

<sup>45</sup> Sowohl für die Populationszahlen als auch die Kosten sowie demzufolge für alle BI-Szenarien wurde das arithmetische Mittel mit zugehöriger Standardabweichung (welche die durchschnittliche Abweichung aller gemessenen Werte vom Mittelwert angibt) und der Median berechnet, inklusive eines Robustheitstests für die Mittelwerte.

**Tabelle 4:** Definition der Regressionsvariablen (1) für NICE  
(Eigene Darstellung)

Variable	Kodierung	Definition
<i>Abhängige Variable:</i>		
Bewertung (Recommendation)	0 = nicht empfohlen 1 = eingeschränkt empfohlen 2 = empfohlen	Entscheidung durch NICE, ob die Technologie für eine Aufnahme in das NHS empfohlen, optimiert / restriktiv empfohlen oder nicht empfohlen wird
CDF# [Juli 2016 – Juni 2018]	2 = nicht empfohlen – erstattet durch CDF	Die durch NICE negativ bewertete Technologie wird durch den CDF für einen bestimmten Zeitraum erstattet / bereitgestellt
<i>Unabhängige Variablen:</i>		
Jahr	Numerisch	Jahreszahl der veröffentlichten Guidance
Indikationsgebiet	Text	Zugeordnetes Indikationsgebiet durch NICE
Vergleichstherapie	0 = nicht vorhanden 1 = vorhanden	Daten und Informationen zur Vergleichs- oder Standardtherapie
RCT	0 = nicht vorhanden 1 = vorhanden	Vorlage von relevanten RCTs (einzeln oder zusammengefasst), die berücksichtigt sind
Kosteneffektivitäts- verhältnis (ICER per QALY)	0 = N/A 1 < 20.000 GBP 2 = 20.000–30.000 GBP 3 > 30.000 GBP	Kosteneffektivitätsverhältnis der bewerteten Technologie zur Vergleichstherapie; Verwendung des maßgeblichen ICERs; liegt ein Werte-Intervall vor, dient der Mittelwert als Referenz
Klinische Wirksamkeit	0 = schlechter 1 = vergleichbar 2 = besser	Bewertung der klinischen Wirksamkeit der neuen Technologie im Vergleich zur Standardbehandlung (smethode) oder zu einem Placebo
EoL-Kriterien	0 = nicht erfüllt 1 = erfüllt	Berücksichtigung der EoL-Kriterien, wenn eine Technologie einen ICER > 30.000 GBP aufweist
Jährliche Arzneimittel- kosten* (Annual drug acquisition cost)	Numerisch i) <i>Minimum</i> ii) <i>Durchschnitt</i> iii) <i>Maximum</i>	Jährliche Arzneimittelkosten pro Patient in GBP, wenn angegeben für unteren / oberen Grenzbereich und / oder einen Einzel-/ (ermittelten) Durchschnittswert
Patientenzahl* a) Erstattungsfähige Zielpopulation b) Zielpopulation c) Zielpopulation gesamt d) Theoretisch geeignete NHS-Population <sup>§</sup>	Numerisch i) <i>Minimum</i> ii) <i>Durchschnitt</i> iii) <i>Maximum</i>	Wenn angegeben, jeweils für i)–iii): a) Erstattungsfähige NHS-Zielpopulation mit Empfehlung b) NHS-Zielpopulation ohne Empfehlung c) NHS-Zielpopulation mit / ohne Empfehlung [auf Indikationsebene] d) Gesamte theoretisch behandelbare/ geeignete NHS-Population mit vorliegender Indikation

CDF, Cancer Drugs Fund; GBP, Great British Pound; ICER, Incremental Cost-Effectiveness Ratio; NHS, National Health Service; QALY, Quality-Adjusted Life Year; RCT, Randomized Controlled Trial.

#Modell-abhängige Variable, ausschließlich für das Modell inklusive der CDF-Erstattungen berücksichtigt.

\*Kontrollvariable zur Berechnung der verschiedenen Budget Impact-Szenarien.

§Daten zur gesamten theoretisch behandelbaren/ geeigneten NHS-Population mit vorliegender Indikation (Durchschnitt) wurden ergänzend berücksichtigt.



Drei Szenarien (Minimum, Durchschnitt, Maximum) für den potenziellen BI sowohl der neuen Therapie als auch der Vergleichsintervention und ein Szenario für das inkrementelle BI (durchschnittlicher BI der neuen Technologie minus durchschnittlicher BI der Vergleichsintervention). Zudem wurden auf Indikationsebene zwei Szenarien für die gesamte NHS-Zielpopulation (Durchschnitt; ex-ante Betrachtung) gebildet: Ein potenzielles BI-Szenario sowohl für die neue Technologie als auch die Vergleichsintervention sowie ein inkrementelles BI-Szenario. Da die Daten der pUs in der Regel nicht vollständig zugänglich sind, konnten sie nicht als zusätzliches BI-Szenario berücksichtigt werden.

**Tabelle 5:** Definition der Regressionsvariablen (2) – Budget Impact-Szenarien für NICE (Eigene Darstellung)

Variable	Kodierung	Definition
<i>Abhängige Variable:</i>		
Bewertung (Recommendation)	0 = nicht empfohlen 1 = eingeschränkt empfohlen 2 = empfohlen	Entscheidung durch NICE, ob die Technologie für eine Aufnahme in das NHS empfohlen, optimiert / restriktiv empfohlen oder nicht empfohlen wird
CDF# [Jul 2016 – Jun 2018]	2 = nicht empfohlen – erstattet durch CDF	Die durch NICE negativ bewertete Technologie wird durch den CDF für einen bestimmten Zeitraum erstattet / bereitgestellt
<i>Unabhängige Variablen:</i>		
Potenzieller BI* –	Numerisch	Berechnung für <u>neue</u> und aktuelle Therapie:
a) Zielpop. / Erstattung [ex-post]	i) <i>Minimum</i>	a) Erstattungsfähige NHS-Zielpopulation [i]–iii] × Jahrestherapiekosten [i]–iii]
b) Zielpopulation [ex-post]	ii) <i>Durchschnitt</i>	b) NHS-Zielpopulation [i]–iii] × Jahrestherapiekosten [i]–iii]
c) Zielpopulation gesamt [ex-ante]	iii) <i>Maximum</i>	c) NHS-Zielpopulation gesamt [i]–iii] × Jahrestherapiekosten [i]–iii]
d) NHS-Population <sup>§</sup> [ex-ante]		d) NHS-Population [ii] × Jahrestherapiekosten [ii]
Inkrementeller BI* –	Numerisch	Berechnung inkrementeller BI für <u>neue</u> Therapie:
a) Zielpop. / Erstattung [ex-post]	ii) <i>Durchschnitt</i>	a) BI-NHS-Erstattung neue Therapie [ii] – BI-NHS-Erstattung aktuelle Therapie [ii]
b) Zielpopulation [ex-post]		b) BI-NHS-Ziel neue Therapie [ii] – BI-NHS-Ziel aktuelle Therapie [ii]
c) Zielpopulation gesamt [ex-ante]		c) BI-NHS-Ziel gesamt neue Therapie [ii] – BI-NHS-Ziel gesamt aktuelle Therapie [ii]
d) NHS-Population <sup>§</sup> [ex-ante]		d) BI-NHS-Population neue Therapie [ii] – BI-NHS-Population aktuelle Therapie [ii]

BI, Budget Impact; CDF, Cancer Drugs Fund; NHS, National Health Service.

#Modell-abhängige Variable, ausschließlich für das Modell inklusive der CDF-Erstattungen berücksichtigt.

\*Szenario-abhängige Variable, um unterschiedliche Projektionen für den BI darzustellen.

§Projektionen zur gesamten theoretisch behandelbaren/ geeigneten NHS-Population (Durchschnitt) wurden für die einzelnen Szenarien des potenziellen sowie inkrementellen BI ergänzend berechnet.

Wenn keine Informationen zur Vergleichstherapie vorlagen oder Daten zu der von NICE angegebenen Vergleichsintervention fehlten, wurde für die Berechnung der Kostenfolgen wie folgt auf die Vorgehensweise von Mauskopf et al. (2013) zurückgegriffen: Die zur Anpassung der Preisuntergrenze gewählten Werte basieren auf der Annahme, dass der Marktanteil am höchsten und der Preisunterschied zugleich am größten ist, wenn keine vergleichbaren Arzneimittel in derselben Wirkstoffklasse innerhalb des NHS verfügbar sind; demgegenüber ist der Marktanteil als am niedrigsten und der Preisunterschied als am geringsten einzuschätzen, wenn Therapien derselben Klasse innerhalb des NHS zur Verfügung stehen. Dementsprechend wurde der potenzielle BI der neuen Technologie mit 1,0 (keine Vergleichstherapie in Wirkstoffklasse vorhanden => Nullannahme), 0,5 (Vergleichstherapie aus einer anderen Wirkstoffklasse vorhanden) oder 0,33 (Vergleichstherapie aus der gleichen Wirkstoffklasse vorhanden) multipliziert. Der Effekt der (offiziellen oder empirisch als relevant eingestuften) Bewertungskriterien und des BI auf die STA-Empfehlungen durch NICE wurde ebenfalls mittels logistischer Regressionsanalyse gemessen (unter Verwendung von *Stata Statistical Software: Release 15*). Die nachstehende Regressionsgleichung beschreibt das ordinale Logit-Modell, welches sowohl auf Subgruppen- als auch auf Indikationsebene berücksichtigt wurde:

$$\text{Outcome}_{ij} = \beta_0 + \beta_1 * \text{Kriterien}_{ij} + \beta_2 * \text{Budget}_{ij} + \mu_i + \tau_j + \varepsilon_{ij} \quad [2]$$

$\text{Outcome}_{ij}$  ist dabei eine kategoriale Variable (dreistufig), die den Wert 2 annimmt, wenn NICE die Technologie  $i$  für die Subgruppe  $j$  empfiehlt (uneingeschränkt), den Wert 1, wenn es die Subgruppe  $j$  eingeschränkt empfiehlt (restriktiv) und den Wert 0, wenn es die Subgruppe nicht empfiehlt. In einer zusätzlichen Modellspezifikation nimmt die Variable  $\text{Outcome}_{ij}$  (dreistufig) ebenfalls den Wert 2 an, wenn die Technologie  $i$  der Subgruppe  $j$  nicht empfohlen, jedoch (nachträglich) durch den CDF erstattet wird.

Die Variable  $\text{Kriterien}_{ij}$  berücksichtigt die wesentlichen (offiziellen) Bewertungsfaktoren, darunter das Vorhandensein einer Vergleichstechnologie und relevanter RCTs (klinische Studienevidenz), die Kosteneffektivität (ICER/QALY gained), klinische Wirksamkeit sowie die Berücksichtigung der EoL-Kriterien.  $\text{Budget}_{ij}$  adressiert entsprechend die unterschiedlichen BI-Szenarien: Sowohl die potenziellen BI-Schätzungen (Minimum, Durchschnitt, Maximum; Durchschnitt NHS-Population) als auch den inkrementellen BI (Durchschnitt; Durchschnitt

NHS-Population), jeweils gemessen in logarithmischer Form<sup>46</sup>. Analog zu [1] ist  $\text{Budget}_{ij}$  zusätzlich als ein quadratischer Term einbezogen worden, um die Nichtlinearität zwischen den potenziellen BI- und den Outcome-Variablen überprüfen zu können. Koeffizienten, die jeweils geschätzt werden mussten, werden als  $\beta_s$  bezeichnet. Veröffentlichungsjahr  $\mu_i$  und Indikationsgebiet  $\tau_i$  wurden ebenfalls als Dummy-Variablen eingesetzt.  $\varepsilon_{ij}$  beschreibt wiederum den Fehlerterm. Darüber hinaus ist die Robustheit der geschätzten Koeffizienten für die verschiedenen BI-Szenarien getestet worden, indem die obersten fünf Prozent der Schätzwerte (sogenannte Ausreißer oder Extremwerte) ausgeschlossen wurden.

---

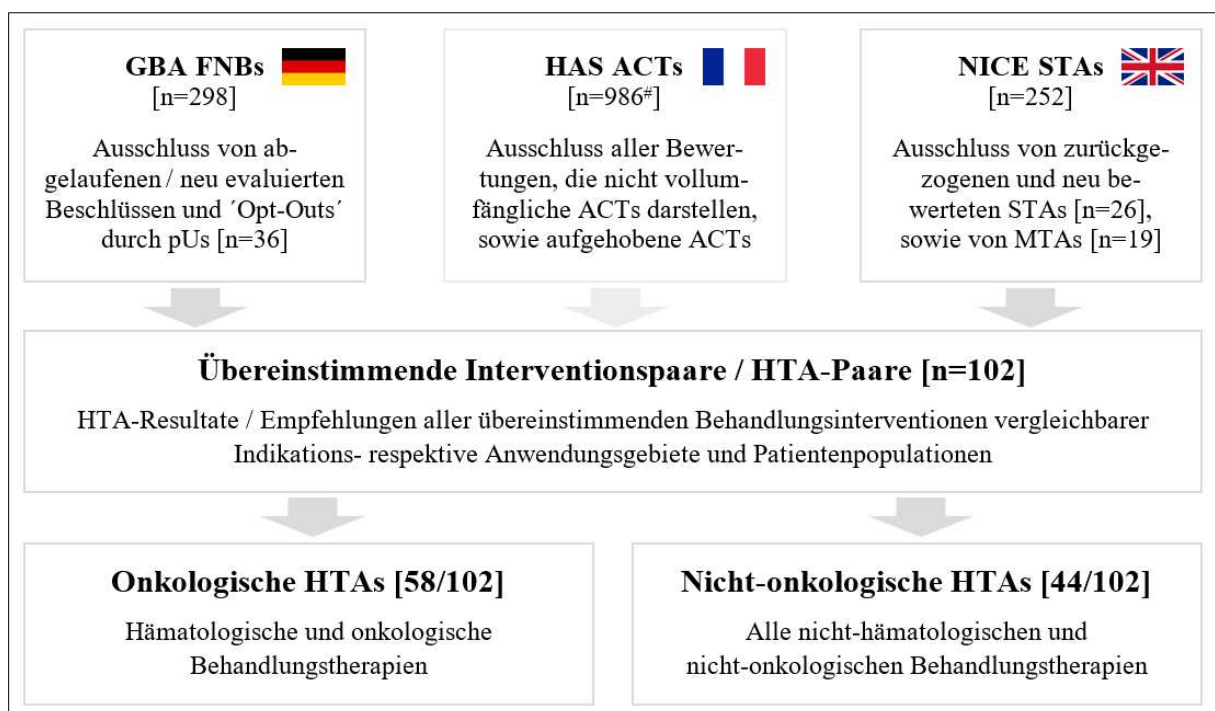
<sup>46</sup> Die logarithmische Form (oder Log-Form) lautet:  $\log(x + 1)$ . Bevor demnach der Logarithmus gebildet wird, ist entsprechend ein Wert von 1 hinzuzufügen, damit die Nullbeobachtungen (bei fehlenden Informationen zur ZVT) berücksichtigt werden können (wobei der Logarithmus von 0 nicht bestimmt wird).

## 5 ERGEBNISSE

Im vorliegenden Kapitel werden die Forschungsergebnisse ausführlich dargelegt, beginnend mit der vergleichenden Untersuchung der offiziellen HTA-Outcomes von GBA, HAS und NICE, bevor im Anschluss daran die Ergebnisse der multivariaten Regressionsanalysen für GBA und IQWiG in Deutschland sowie für NICE in England herausgestellt werden.

### 5.1 Abgleich der identifizierten HTAs von GBA, HAS und NICE

Bei der systematischen Gegenüberstellung wurden alle abgeschlossenen und öffentlich zugänglichen GBA FNBs und NICE STAs sowie (erweiternd) HAS ACTs von Januar 2011 bis einschließlich Juni 2018 berücksichtigt. In Abbildung 5 ist der Auswahlprozess skizzenhaft dargestellt, der sowohl die Kriterien für die Identifizierung der übereinstimmenden Interventionspaare der drei HTA-Institutionen als auch die jeweiligen Ausschlusskriterien zusammenfasst.



**Abbildung 5:** Auswahl der übereinstimmenden HTAs von GBA, HAS und NICE (Januar 2011 – Juni 2018) (Eigene Darstellung; in Anlehnung an: Schaefer et al., 2021)

ACT, Avis de la Commission de la Transparence; FNB, Frühe (Zusatz-)Nutzenbewertung; MTA, Multiple Technology Appraisal; pU, pharmazeutisches Unternehmen; STA, Single Technology Appraisal.

<sup>#</sup>Die Daten der französischen HAS wurden durch die Suche nach allen veröffentlichten (Teil-)Bewertungen von Arzneimitteln beziehungsweise Therapien ermittelt. Daraufhin wurden ausschließlich ACTs mit den in einem vorgelagerten Schritt identifizierten HTA-Paaren von GBA und NICE verglichen.

**Tabelle 6:** Abgleich der identifizierten HTAs von GBA, HAS und NICE (Januar 2011 – Juni 2018)  
(Eigene Darstellung)

Wirkstoff (Kombinationstherapie)	Therapeutisches Gebiet (Anwendung)	GBA/FNB		HAS ACT		NICE STA				
		Zusatznutzen (Beschluss)	OD (Status)	Zusatznutzen (Recommendation)	Erstattungsrate (SMR, in %)	ASMR (Level*)	Empfehlung (Recommendation)	ICER/QALY (in GBP)	SoL- Kriterien	CDF- Erstattung
<i>Onkologische Wirkstoffe:</i>										
Eribulin	Onkologie	+	Nein	+	≤ 65	IV-V	-	69,800	Erfüllt	Erfüllt
Cabazitaxel	Onkologie	+	Nein	+	65	III	+	45,982	Erfüllt	Erfüllt
Abiraterone (Docetaxel)	Onkologie	+	Nein	+	100	III	+	48,400	Erfüllt	Erfüllt
Vemurafenib	Onkologie	+	Nein	+	100	III	+	47,900	Erfüllt	Erfüllt
Ipilimumab	Onkologie (Zweitlinientherapie)	+	Nein	-	≤ 65	V	+	42,200	Erfüllt	Erfüllt
Ipilimumab	Onkologie (Erstlinientherapie)	-	Nein	-	0		+	38,250	Erfüllt	Erfüllt
Axitinib	Onkologie	+	Nein	+	100	IV	+	43,200	Erfüllt	Erfüllt
Ruxolitinib	Hämatologie/Onkologie	+	Nein	+	100	III	+	32,000	Erfüllt	Erfüllt
Crizotinib	Onkologie (Zweitlinientherapie)	+	Nein	+	100	III	-	75,100	Erfüllt	Erstattet
Pixantron	Hämatologie/Onkologie	-	Nein	-	≤ 65	V	+	22,000	N/A	
Aflibercept	Onkologie	+	Nein	-	≤ 65	V	-	47,500	Nicht erfüllt	
Bosutinib	Onkologie	+	Ja	-	100	V	-	51,500	erfüllt	Erstattet
Enzalutamide (Docetaxel)	Onkologie	+	Nein	+	100	III	+	22,600	Nicht erfüllt	
Enzalutamid	Onkologie	+	Nein	+	100	IV	+	30,000	N/A	
Dabrafenib	Onkologie	-	Nein	-	100	V	+	11,000	N/A	
Afatinib	Onkologie	+	Nein	-	100	V	+	N/A	N/A	
Pomalidomid	Hämatologie/Onkologie	+	Nein	+	65	III	+	49,337	Erfüllt	
Nintedanib	Onkologie	+	Nein	-	0		+	51,692	Erfüllt	
Pertuzumab	Onkologie	-	Nein	-	0		+	33,211	N/A	
Pertuzumab (Trastuzumab / Docetaxel)	Onkologie (Kombinationstherapie)	+	Nein	+	≤ 65	III	+	33,211	Erfüllt	
Regorafenib	Onkologie	-	Nein	+	100	IV	+	44,000	Erfüllt	
Idelalisib	Hämatologie/Onkologie	+	Nein	+	100	III	+	40,500	Erfüllt	
Crizotinib	Onkologie (Erstlinientherapie)	+	Nein	+	100	IV	+	47,291	Erfüllt	
Ramucirumab	Onkologie	+	Nein	-	100	V	-	298,150	Nicht erfüllt	
Obinutuzumab	Hämatologie/Onkologie (Follikuläres Lymphom)	+	Ja	-	≤ 65	V	+	30,000	Nicht erfüllt	
Obinutuzumab (Chlorambucil)	Hämatologie/Onkologie (Kombinationstherapie)	+	Ja	+	65	III	+	25,000	Nicht erfüllt	
Brentuximab vedotin	Hämatologie/Onkologie	+	Ja	+	65	III	+	30,000	Nicht erfüllt	
Panobinostat	Hämatologie/Onkologie (Multiples Myelom)	+	Ja	-	≤ 65	V	+	25,000	N/A	
Trifluridine Tipiracil	Onkologie	+	Nein	-	100	V	+	49,392	Erfüllt	

Tabelle 6: (Fortsetzung)

Wirkstoff (Kombinationstherapie)	Therapeutisches Gebiet (Anwendung)	GBA FNB		HAS ACT		NICE STA		CDF- Erstattung	
		Zusatznutzen (Beschluss)	OD (Status)	Zusatznutzen (Recommendation)	Erstattungsrate (SMR, in %)	ASMR (Level*)	Empfehlung (Recommendation)		ICER/QALY (in GBP)
Trametinib	Hämatologie/Onkologie (Myelom)	+	Nein	+	100	III	+	N/A	Erfüllt
Venetoclax	Hämatologie/Onkologie	+	Ja	-	100	V	-	55.000	Erfüllt
Pembrolizumab	Onkologie (vorbehandelter Lungenkrebs)	+	Nein	-	≤ 65	V	+	53.222	Erfüllt
Pembrolizumab	Onkologie (nicht behandelter Lungenkrebs)	+	Nein	+	≤ 65	IV	-	53.830	Erfüllt
Pembrolizumab	Onkologie (nicht behandelter Melanom)	+	Nein	+	≤ 65	IV	+	50.000	Erfüllt
Nivolumab	Onkologie (Melanom, Monotherapie)	+	Nein	+	100	III	+	25.000	Erfüllt
Nivolumab Ipilimumab	Onkologie (Melanom)	-	Nein	-	≤ 65	V	+	30.000	Nicht erfüllt
Osimertinib	Onkologie	+	Nein	+	100	IV	-	65.720	Nicht erfüllt
Ponatinib	Hämatologie/Onkologie	+	Ja	+	100	IV	+	50.000	Erfüllt
Olaparib	Onkologie	+	Ja	+	100	IV	+	46.700	Erfüllt
Blinatumomab	Hämatologie/Onkologie	+	Ja	+	≤ 65	IV	+	49.190	Erfüllt
Olaratumab	Onkologie	+	Ja	-	0		-	60.000	Erfüllt
Lenvatinib (Everolimus)	Onkologie (Kombinationstherapie)	+	Nein	-	0		+	25.000	N/A
Cobimetinib	Onkologie (Melanom)	+	Nein	+	100	III	-	100.000	Erfüllt
Radium 223 Dichlorid	Onkologie	+	Nein	+	≤ 65	IV	-	50.000	Erfüllt
Trastuzumab Emtriansin	Onkologie	+	Nein	+	65	II	-	166.400	Erfüllt
Vismodegib	Onkologie	+	Nein	+	100	IV	-	N/A	N/A
Ibrutinib	Hämatologie/Onkologie	+	Nein	+	100	III	+	50.000	Erfüllt
Ceritinib	Hämatologie/Onkologie (nicht behandeltes NHL)	-	Nein	+	100	IV	+	25.000	N/A
Ceritinib	Hämatologie/Onkologie (vorbehandeltes NHL)	+	Nein	+	100	IV	+	50.000	Erfüllt
Nivolumab	Hämatologie/Onkologie (HL)	-	Nein	-	≤ 65	V	+	30.000	Erfüllt
Nivolumab	Onkologie (Niere)	+	Nein	+	65	III	+	50.000	Erfüllt
Nivolumab	Onkologie (Lunge)	+	Nein	+	65	III	-	49.160	Erfüllt
Nivolumab	Onkologie (Lunge)	+	Nein	+	65	III	-	50.014	Erfüllt
Carfilzomib	Hämatologie/Onkologie (Myelom)	+	Nein	+	≤ 65	IV	+	27.629	Nicht erfüllt
Cabozantinib	Onkologie (Schilddrüse)	+	Ja	+	100	IV	+	25.000	N/A
Cabozantinib	Onkologie (Niere)	+	Nein	+	100	III	+	50.000	Erfüllt
Ribociclib	Onkologie	-	Nein	-	100	V	+	20.000	N/A
Ixazomib	Hämatologie/Onkologie (Multiples Myelom)	+	Ja	-	100	V	-	30.000	Nicht erfüllt

Tabelle 6: (Fortsetzung)

Wirkstoff (Kombinationstherapie)	Therapeutisches Gebiet (Anwendung)	GBA FNB		HAS ACT			NICE STA			
		Zusatznutzen (Beschluss)	OD (Status)	Zusatznutzen (Recommendation)	Erstattungsrate (SMR, in %)	ASMR (Level*)	Empfehlung (Recommendation)	ICER/QALY (in GBP)	EoL- Kriterien	CDF- Erstattung
<i>Nicht-onkologische Wirkstoffe:</i>										
Ticagrelor	Herz-Kreislauf (Akutes Koronarsyndrom)	+	Nein	+	65	IV	+	8.406		
Apixaban	Herz-Kreislauf (VTE, Knie-/Hüftgelenk)	+	Nein	+	65	IV	+	9.272		
Apixaban	Herz-Kreislauf (VTE)	+	Nein	-	65	V	+	12.550		
Apixaban	Herz-Kreislauf (Schlaganfall, Embolie)	+	Nein	-	65	V	+	20.000		
Fingolimod	Nervensystem	+	Nein	+	65	IV	+	30.000		
Pirfenidon	Atmungssystem	+	ja	+	30	IV	+	26.902		
Aflibercept	Augen	-	Nein	-	65	V	+	20.000		
Dapagliflozin	Stoffwechsel	-	Nein	-	30	V	+	20.000		
Ocriplasmin	Augen	+	Nein	+	≤ 65	IV	+	25.700		
Teriflunomid	Nervensystem	-	Nein	-	65	V	+	20.000		
Sofosbuvir	Infektion	+	Nein	+	100	III	+	22.118		
Dimethyl fumarat	Nervensystem	-	Nein	-	65	V	+	27.700		
Mirabegron	Urogenitalsystem	-	Nein	-	0		+	5.270		
Empagliflozin	Stoffwechsel	+	Nein	-	65	V	+	9.068		
Canagliflozin	Stoffwechsel	-	Nein	-	65	V	+	10.839		
Nalmefen	Sonstiges	-	Nein	+	30	IV	+	5.100		
Apremilast	Muskel-Skelett-System	-	Nein	-	15-30	V	+	N/A		
Aflibercept	Augen (Choroidale Neovaskularisation)	-	Nein	+	65	IV	+	14.550		
Nintedanib	Atmungssystem	+	ja	+	30	IV	+	30.000		
Evolocumab	Herz-Kreislauf	-	Nein	+	65	IV	+	33.450		
Alirocumab	Herz-Kreislauf	-	Nein	+	65	IV	+	34.650		
Aflibercept	Augen (Makulaödem / RVV, VAV)	-	Nein	-	65	V	+	30.950		
Aflibercept	Augen (Diabetisches Makulaödem)	-	Nein	+	65	III	+	10.000		
Sacubitril valsartan	Herz-Kreislauf	+	Nein	+	65	IV	+	28.000		
Mepolizumab	Atmungssystem	+	Nein	+	65	IV	+	29.163		
Ticagrelor	Herz-Kreislauf (Atherothrombose)	+	Nein	-	≤ 65	V	+	22.674		
Elbasvir Grazoprevir	Infektion	-	Nein	+	65	IV	+	20.000		
Obeticholic	Verdauungssystem	+	ja	-	65	V	+	33.178		
Ixekizumab	Haut	+	Nein	-	65	V	+	30.000		

Tabelle 6: (Fortsetzung)

Wirkstoff (Kombinationstherapie)	Therapeutisches Gebiet (Anwendung)	GBA FNB		HAS ACT			NICE STA			
		Zusatznutzen (Beschluss)	OD (Status)	Zusatznutzen (Recommendation)	Erstattungsrate (SMR, in %)	ASMR (Level*)	Empfehlung (Recommendation)	ICER/QALY (in GBP)	EoL- Kriterien	CDF- Erstattung
Baricitinib	Muskel-Skelett-System	-	Nein	-	65	V	+	27.910		
Tofacitinib	Muskel-Skelett-System	-	Nein	-	65	V	+	25.000		
Etelcalcetid	Stoffwechsel	-	Nein	-	≤ 65	V	+	20.713		
Belimumab	Muskel-Skelett-System	+	Nein	+	65	IV	+	N/A		
Secukinumab	Haut	+	Nein	+	65	IV	+	30.050		
Secukinumab	Muskel-Skelett-System	-	Nein	-	30-65	V	+	2.234		
Vedolizumab	Verdauungssystem	-	Nein	+	≤ 65	IV-V	+	23.775		
Sarilumab	Muskel-Skelett-System	+	Nein	-	65	V	+	87.813		
Lumacaftor-Ivacaftor	Stoffwechsel	+	Nein	+	65	IV	-	283.500		
Sofosbuvir Velpatasvir Voxilaprevir	Infektion (Hepatitis C)	-	Nein	+	100	IV	+	20.000		
Sofosbuvir Velpatasvir	Infektion (Hepatitis C)	+	Nein	+	65	IV	+	20.000		
Edoxaban	Herz-Kreislauf	+	Nein	-	65	V	+	22.525		
Brodalumab	Haut	+	Nein	-	65	V	+	25.000		
Ombitasvir (Paritaprevir / Ritonavir)	Infektion (Kombinationstherapie)	+	Nein	+	≤ 65	IV	+	20.000		
Glecaprevir (Pibrentasvir)	Infektion (Kombinationstherapie)	-	Nein	+	100	IV	+	20.000		

ACT, Avis de la Commission de la Transparence; ASMR, Amélioration du Service Médical Rendu; CDF, Cancer Drugs Fund; EoL, End-of-Life; FNB, Frühe (Zusatz-)Nutzenbewertung; GBP, British Pound Sterling; HL, Hodgkin-Lymphom; ICER, Incremental Cost-Effectiveness Ratio; N/A, nicht angegeben; NHL, Non-Hodgkin-Lymphom; OD, Orphan Drug; QALY, Quality-Adjusted Life Year; RVV, Retinaler Venenverschluss; SMR, Service Médical Rendu; STA, Single Technology Appraisal; VAV, Venenastverschluss; VTE, Venöse Thromboembolie.

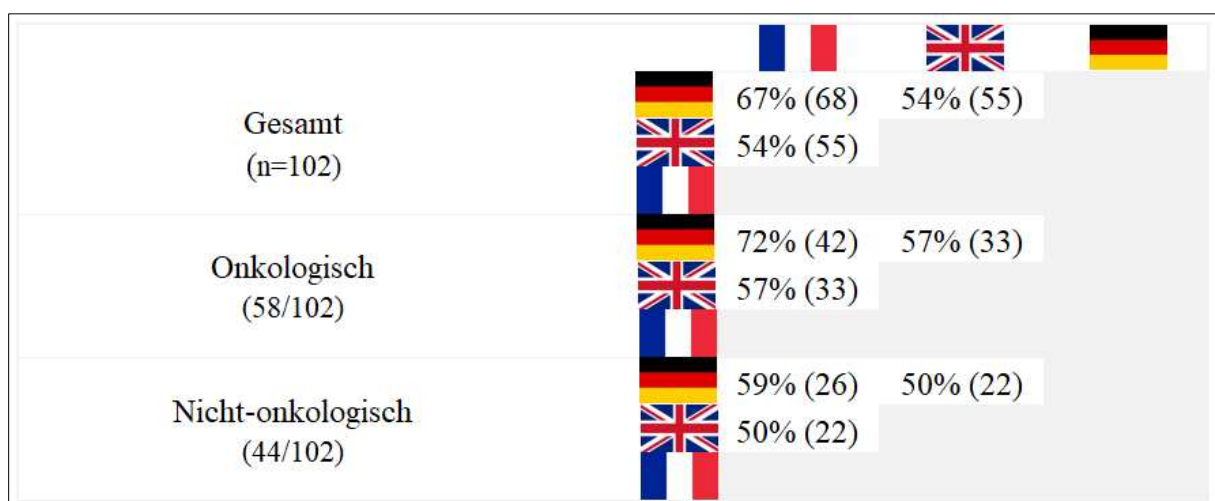
\* ASMR-Level I-V: Mit klinischem Zusatznutzen: I (erheblich), II (beträchtlich), III (moderat), IV (gering); Ohne klinischen Zusatznutzen: V, nicht vorhanden.



Innerhalb des Untersuchungszeitraums konnten insgesamt 102 übereinstimmende Interventionspaare identifiziert werden, wovon 58 für onkologische und 44 für nicht-onkologische Krankheitsbilder indiziert sind. Tabelle 6 gibt einen vollständigen Überblick zu allen berücksichtigten HTAs inklusive der jeweiligen Bewertungsempfehlungen sowie der extrahierten Informationen zu den vordefinierten – und insbesondere für onkologische Therapien potenziell relevanten – Einflussfaktoren von GBA (OD-Status), HAS (SMR beziehungsweise Erstattungsrate, ASMR) und NICE (ICER/QALY, EoL-Kriterien, CDF-Erstattung).

### 5.1.1 Interventionspaare im Vergleich: Eine Übersicht

Der paarweise Abgleich der HTAs zeigt teils erhebliche Unterschiede zwischen den jeweiligen Outcomes. Diskrepanzen in den HTA-Resultaten bestehen dabei offensichtlich beim Vergleich aller drei Agenturen, die sich darüber hinaus je nach Indikationsebene unterschiedlich darstellen. Auf der einen Seite wird erkennbar, dass HTA-Resultate für GBA / HAS (insgesamt, 67 %; onkologisch, 72 %; nicht-onkologisch, 59 %) im Vergleich zu sowohl GBA / NICE als auch HAS / NICE (insgesamt, 54 %; onkologisch, 57 %; nicht-onkologisch, 50 %) eine insgesamt höhere Übereinstimmung aufweisen (vgl. Abbildung 6). Auf der anderen Seite erscheinen die Übereinstimmungen in den HTA-Resultaten onkologischer Therapien (57 – 72 %) durchweg größer als bei den nicht-onkologischen Behandlungen (50 – 59 %).



**Abbildung 6:** Paarweiser Vergleich der Übereinstimmung der HTA-Outcomes von GBA, HAS und NICE (Eigene Darstellung; in Anlehnung an: Schaefer et al., 2021)

Die in Abbildung 6 dargestellten Resultate der paarweisen HTA-Vergleiche von GBA / HAS, GBA / NICE und HAS / NICE werden durch die berechneten ORs weitgehend gestützt. Einerseits zeigt sich ein signifikanter Zusammenhang der HTA-Resultate von GBA / HAS sowohl für die Therapien insgesamt (OR = 1,68, 95 % CI: 0,94–3,00;  $p = 0,08$ ) als auch im Speziellen für die Krebstherapien (OR = 3,09, 95 % CI: 1,27–7,52;  $p = 0,01$ ). Andererseits liegt für die HTA-Outcomes insgesamt (HAS / NICE: OR = 0,29, 95 % CI: 0,15–0,55;  $p < 0,001$ ; GBA / NICE: OR = 0,48, 95 % CI: 0,25–0,94;  $p = 0,03$ ) sowie nicht-onkologischer Therapien (OR = 0,03, 95 % CI: 0,01–0,20;  $p < 0,001$ ) sowohl für GBA / NICE als auch für HAS / NICE ein signifikant niedrigerer Zusammenhang vor (Schaefer et al., 2021).

Folgerichtig verstärken sich die in den paarweisen Abgleichen von GBA / HAS, GBA / NICE und HAS / NICE festgestellten Abweichungen, wenn die Empfehlungen aller drei HTA-Agenturen gegenüberstellend betrachtet werden (vgl. hierzu auch Tabelle 6). Insgesamt knapp zwei Drittel (64/102, 63 %) der HTAs unterscheiden sich dabei grundlegend in den Bewertungsergebnissen. Die HTA-Resultate für Krebstherapien (33/58, 57 %) weichen entsprechend weniger deutlich voneinander ab als die Outcomes für nicht-onkologische Interventionen (31/44, 70 %).

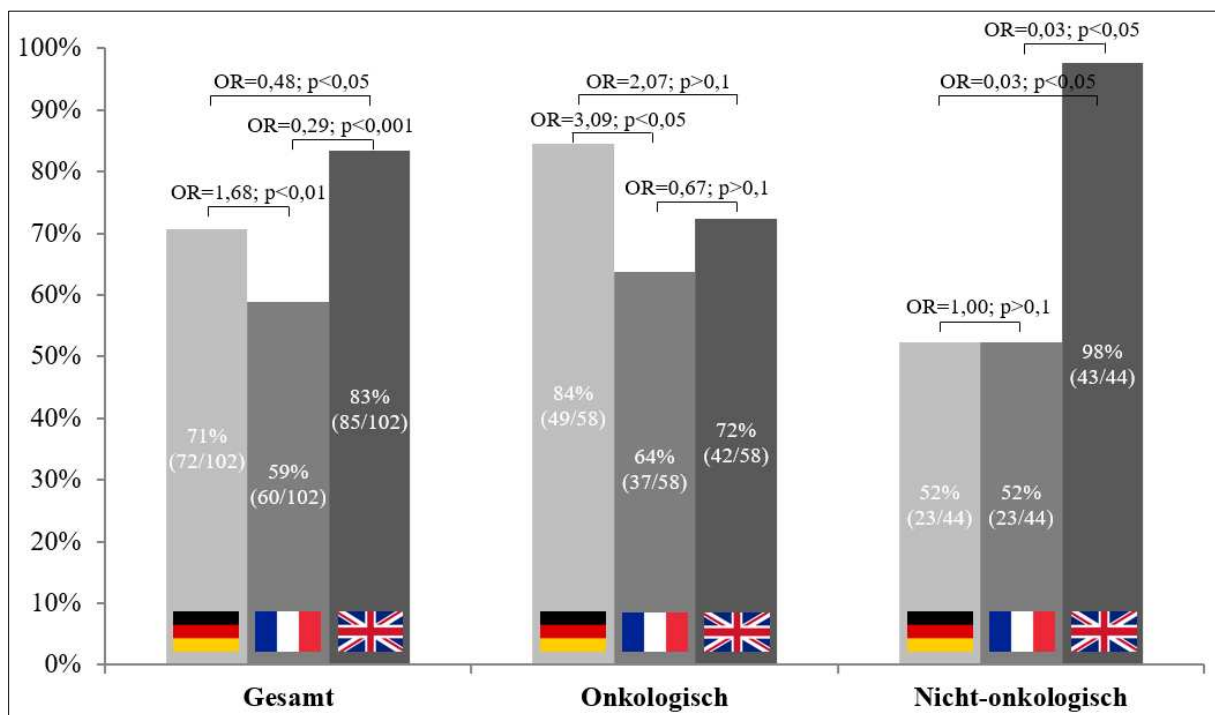
### **5.1.2 Indikationsgebiete im Vergleich: Onkologische versus nicht-onkologische Therapien**

Während NICE 85/102 (83 %) <sup>47</sup> Behandlungen positiv bewertete, berichteten HAS und GBA für 60/102 (59 %) beziehungsweise 72/102 (71 %) Interventionen einen positiven Zusatznutzen (vgl. Abbildung 7). 15/72 Therapien (darunter drei nicht-onkologische Behandlungen), für die der GBA einen positiven Zusatznutzen bestätigte, hatten zum Zeitpunkt der FNB einen OD-Status. Insgesamt bewerteten NICE 73 % (3/3 Nicht-Onkologika, 8/12 Onkologika) und HAS 53 % (2/3 Nicht-Onkologika, 6/12 Onkologika) der ODs positiv, wobei alle durch NICE negativ evaluierten onkologischen ODs durch den CDF berücksichtigt und entsprechend erstattet wurden. Unter den von HAS bewerteten Therapien befanden sich zudem sechs Medikamente (davon ein nicht-onkologisches Arzneimittel), die als nicht-erstattungsfähig eingestuft wurden und demzufolge unter den negativ bewerteten HTA-Ergebnissen abgebildet sind.

---

<sup>47</sup> Mit Cabozantinib (TA516, Behandlung des medullären Schilddrüsenkarzinoms) ist ein Medikament berücksichtigt worden, das als Teil eines MTAs veröffentlicht, jedoch einzeln – vergleichbar einem STA – bewertet wurde. Zudem wurden eine Subgruppe von Ibrutinib (TA491, zusammengefasstes Anwendungsgebiet für die Behandlung maligner Lymphomkrankungen) und eine Teilindikation von Brentuximab Vedotin (TA524, erweitertes Anwendungsgebiet für die Behandlung des Hodgkin-Lymphoms) durch den CDF erstattet; da weitere Subpopulationen jedoch positiv bewertet wurden, sind beide Behandlungen mit einer Empfehlung respektive positiven Bewertung für die vergleichende Untersuchung berücksichtigt worden.

Die HTA-Resultate der Interventionspaare unterscheiden sich teils erheblich, insbesondere beim Vergleich der zugrunde liegenden therapeutischen Gebiete (vgl. Tabelle 7). Während sowohl GBA als auch HAS positive HTA-Ergebnisse für jeweils 23/44 (52 %) nicht-onkologische Interventionen berichteten, empfahl NICE fast alle Nicht-Onkologika (43/44, 98 %; Durchschnittswert ICER, 29.001 GBP/QALY) für eine routinemäßige Anwendung innerhalb des NHS. Exemplarisch können unter anderem die Therapien für muskuloskelettale Erkrankungen betrachtet werden; während NICE alle der sechs identifizierten Behandlungen positiv bewertete, berichteten HAS für Belimumab (das beim Systemischen Lupus erythematoses indiziert ist) und GBA zusätzlich auch für Sarilumab (das bei Rheumatoider Arthritis indiziert ist) einen positiven Zusatznutzen. Bei den Infektionserkrankungen wiederum bewerteten NICE und HAS alle der sechs Wirkstoffe positiv, wohingegen der GBA lediglich für drei der Therapien (Sofosbuvir, Sofosbuvir Velpatasvir und Ombitasvir plus Paritaprevir Ritonavir für die Behandlung chronischer Hepatitis C) einen positiven Zusatznutzen berichtete. Eine Ausnahme unter den Indikationsgebieten bilden die Krankheiten des Atmungssystems, für die alle der drei vorliegenden Therapien sowohl durch NICE als auch GBA und HAS positiv bewertet wurden.



**Abbildung 7:** Positive HTA-Outcomes von GBA, HAS und NICE (Eigene Darstellung; in Anlehnung an: Schaefer et al., 2021)

Anmerkung: Die ergänzenden Daten zeigen die berechneten Odds Ratios (ORs) mit den entsprechenden p-Werten (mit  $p < 0,1$  als statistisch signifikant).

Tatsächlich zeigen nicht-onkologische Bewertungen einen signifikant höheren Zusammenhang gegenüber positiven Empfehlungen durch NICE auf als onkologische STAs (OR = 16,38, 95 % CI: 2,08–129,11; p = 0,008). Im Gegensatz dazu lässt sich für die GBA-Beschlüsse eine signifikant größere Chance für einen positiven Zusatznutzen feststellen, wenn eine Krebstherapie vorliegt (OR = 4,97, 95% CI: 1,97–12,53; p < 0,001). HAS ACTs zeigen wiederum keinen eindeutigen Zusammenhang, weder zu den onkologischen (OR = 1,61, 95 % CI: 0,72–3,57; p = 0,243) noch zu den nicht-onkologischen (OR = 0,62, 95 % CI: 0,28–1,38; p = 0,243) Behandlungsinterventionen (Schaefer et al., 2021).

**Tabelle 7:** Übereinstimmende HTAs nach therapeutischem Gebiet  
(Eigene Darstellung; in Anlehnung an: Schaefer et al., 2021)

Therapeutisches Gebiet	HTA-Paare	GBA FNB		HAS ACT		NICE STA	
	Anzahl (OD*)	Zusatznutzen		ASMR		Empfehlung	
		+	-	+	-	+	-
Onkologische Erkrankungen	58 (12)	49 (12)	9 (0)	37 (6)	21 (6)	42 (8)	16 <sup>#</sup> (4)
Andere Erkrankungen	44 (3)	23 (3)	21 (0)	23 (2)	21 (1)	43 (3)	1 (0)
Atmungssystem	3 (2)	3 (2)	0	3 (2)	0	3 (2)	0
Augenerkrankungen	5	1	4	3	2	5	0
Hauterkrankungen	3	3	0	1	2	3	0
Herz-Kreislauf	9	7	2	5	4	9	0
Infektionen	6	3	3	6	0	6	0
Muskel-Skelett-System	6	2	4	1	5	6	0
Nervensystem	3	1	0	1	0	3	0
Stoffwechselsystem	5	2	3	1	4	4	1
Urogenital	1	0	1	0	1	1	0
Verdauungssystem	2 (1)	1 (1)	1	1	1 (1)	2 (1)	0
Sonstige	1	0	1	1	0	1	0
Gesamt	102 (15)	72 (15)	30 (0)	60 (8)	42 (7)	85 (11)	17 (4)
Relativer Anteil	1,0 (0,15)	0,71 (0,21)	0,29 (0,00)	0,59 (0,13)	0,41 (0,17)	0,83 (0,13)	0,17 (0,24)

ACT, Avis de la Commission de la Transparence; ASMR, Amélioration du Service Médical Rendu; CDF, Cancer Drugs Fund; FNB, Frühe (Zusatz-)Nutzenbewertung; OD, Orphan Drug; STA, Single Technology Appraisal.

\*Anzahl der ODs an den identifizierten Health Technology Assessment-Interventionspaaren nach Indikationsgebiet.

<sup>#</sup>Insgesamt wurden 11/16 (0,69) onkologische Therapien (darunter alle ODs) nachträglich durch den CDF erstattet.

Insgesamt empfahl NICE 42/58 (72 %; Durchschnittswert ICER, 49.474 GBP/QALY) onkologische Behandlungstechnologien. Für mehr als 80 % (47/58) der bewerteten Krebstherapien wurden dabei EoL-Kriterien berücksichtigt; 37/47 (79 %; Durchschnittswert ICER, 54.454 GBP/QALY) Krebstherapien erfüllten die Kriterien, wovon wiederum 70 % (26/37) auch als

EoL-Behandlungen<sup>48</sup> durch NICE empfohlen wurden. Demzufolge scheinen die EoL-Kriterien einen direkten Einfluss auf die Empfehlungen durch NICE zu haben, auch wenn für die vorliegende Stichprobe kein eindeutiger Zusammenhang festgestellt werden konnte ( $p = 0,813$ ; Exakter Fisher-Test). Seit dem Relaunch im Juli 2016 wurden zudem 11/16 (69 %; Durchschnittswert ICER, 64.248 GBP/QALY) der nicht empfohlenen Krebsbehandlungen durch den CDF erstattet. Dennoch konnte – unter der zugrunde liegenden Annahme, dass die Interventionen anschließend für eine routinemäßige Behandlung innerhalb des NHS England empfohlen werden – auch für die CDF-berücksichtigten STAs kein signifikanter Zusammenhang aufgezeigt werden ( $p = 0,407$ ; Exakter Fisher-Test; vgl. Schaefer et al., 2021).

GBA und HAS bewerteten 49/58 (84 %) beziehungsweise 37/58 (64 %) onkologische Therapien mit einem positiven Zusatznutzen. Da mehr als 20 % (12/58) der identifizierten Krebsmedikamente als ODs zugelassen wurden, scheint der OD-Status in Deutschland einen signifikanten Faktor für den vergleichsweise höheren Anteil an positiven GBA-Beschlussfassungen von onkologischen Therapien darzustellen ( $p = 0,003$ ; Chi-Quadrat-Test). Interessanterweise wurde für 21/53 (40 %, unter Ausschluss der fünf nicht-erstattungsfähigen Medikamente) durch die von HAS bewerteten Krebsmedikamente keine fixe Erstattungsrate auf Grundlage des tatsächlichen klinischen Nutzens (SMR) angegeben. Für insgesamt 30/32 (94 %) onkologischen Behandlungen, die in Frankreich entsprechend vollständig erstattet werden, konnte ein signifikanter Zusammenhang zu einem positiven klinischen Zusatznutzen (ASMR) festgestellt werden ( $p = 0,003$ ; Chi-Quadrat-Test; vgl. Schaefer et al., 2021).

## 5.2 Analyse der HTA-Kriterien und GBA-Beschlüsse / IQWiG-Bewertungen

Im Folgenden werden zuerst die deskriptiven Untersuchungsergebnisse für GBA und IQWiG vorgestellt, bevor auf die Resultate der jeweiligen Regressionsanalysen für die Beschlussfassungen respektive Dossierbewertungen eingegangen wird. Abschließend werden die Unterschiede und Gemeinsamkeiten der GBA-Beschlüsse und IQWiG-Bewertungen auf Grundlage der deskriptiven sowie statistischen Ergebnisdarstellungen gegenübergestellt.<sup>49</sup>

---

<sup>48</sup> Bis einschließlich Juni 2018 wurden die EoL-Kriterien ausschließlich für onkologische und hämatologische Therapien angewendet.

<sup>49</sup> Erste Teilergebnisse der Analysen und Untersuchungen zu den GBA-Beschlüssen und IQWiG-Bewertungen wurden auf der 'Virtual ISPOR Europe 2020' vorgestellt (vgl. hierzu bei Schaefer et al., 2020).

### **5.2.1 Deskriptive Untersuchungsergebnisse für GBA und IQWiG**

Die deskriptiven Resultate der Untersuchung zeigen, dass innerhalb des vorliegenden Untersuchungszeitraums 292 FNB-Verfahren veröffentlicht wurden. Davon sind 36 FNBs abgelaufen beziehungsweise nicht mehr aktuell, wobei 24 Wirkstoffe durch Opt-Outs der pUs zurückgezogen oder später vollständig vom Markt zurückgenommen wurden (und bei Abschluss der Untersuchung auch nicht wieder zur Verfügung standen). Für den GBA wurden demnach 262 FNB-Verfahren mit 578 Beschlussentscheiden berücksichtigt, worunter sich 53 ODs mit 68 Beschlüssen befinden. Da es für die Wirkstoffe Ivermectin (Hautkrankheit/ Rosacea) sowie Linagliptin und Gaxilose (Stoffwechselerkrankungen/ Diabetes mellitus und Hypolaktasie) keine Beauftragung durch den GBA gegeben hat, konnten für IQWiG lediglich 259 Dossierbewertungen mit 574 Zusatznutzenempfehlungen berücksichtigt werden.

#### *5.2.1.1 Resultate auf Teilpopulationsebene*

GBA und IQWiG berichteten für 228/578 (40 %) respektive 189/574 (33 %) Teilpopulationen einen positiven Zusatznutzen (vgl. Tabelle 8 und Tabelle 9). Beim Blick auf die FNBs ohne Nachweis für einen positiven Zusatznutzen wird dabei deutlich erkennbar, dass evidenzbasierte Kriterien wie die klinische Studienevidenz (GBA, 95 %; IQWiG, 99 %) und die vergleichende klinische Effektivität (GBA, 99 %; IQWiG, 100 %) in der Regel als nicht vorhanden respektive nicht ausreichend erfüllt eingestuft wurden. Während für das Jahr 2012 interessanterweise mehr GBA-Beschlüsse (71 %) respektive IQWiG-Bewertungen (52 %) mit einem positiven Zusatznutzen vorlagen, wurden für alle weiteren Veröffentlichungsjahre durchgehend mehr FNBs ohne als mit einem positiven Zusatznutzen festgestellt.

**Tabelle 8:** Deskriptive Ergebnisse für GBA auf Teilpopulationsebene  
(Eigene Darstellung)

Variable	GBA-Beschluss		Gesamt (n)
	Ohne Zusatznutzen (%)	Mit Zusatznutzen (%)	
<b>Orphan Drug-Status</b>			
Nein	68,6	31,4	510
Ja	0,0	100,0	68
<b>Zweckmäßige Vergleichstherapie (ZVT)</b>			
Nicht vorhanden	8,8	91,3	80
Vorhanden	68,9	31,1	498
<b>Klinische Studienevidenz: Randomisierte kontrollierte Studie (RCTs als systematischer Review/ Meta-Analyse)</b>			
Nicht vorhanden	95,3	4,7	274
Vorhanden	37,7	62,3	236
<b>Vergleichende klinische Wirksamkeit: Signifikanter Unterschied in den patientenrelevanten Endpunkten</b>			
Nicht vorhanden	98,7	1,3	301
Vorhanden	25,4	74,6	209
<b>Veröffentlichungsjahr</b>			
2011	75,0	25,0	4
2012	29,0	71,0	31
2013	70,6	29,4	51
2014	58,7	41,3	63
2015	57,6	42,4	125
2016	65,8	34,2	149
2017	63,5	36,5	96
2018 (Januar-Juni)	57,6	42,4	59
<b>Indikation / Therapeutisches Gebiet</b>			
Atmungssystem	50,0	50,0	26
Augen	72,7	27,3	11
Blut und blutbildende Organe	71,4	28,6	7
Haut	29,4	70,6	17
Herz-Kreislauf	45,0	55,0	20
Infektionen	55,1	44,9	118
Muskel-Skelett-System	86,4	13,6	22
Nervensystem	57,7	42,3	26
Onkologie	52,3	47,7	176
Psychische Gesundheit	100,0	0,0	7
Stoffwechselsystem	77,6	22,4	125
Urogenital	66,7	33,3	3
Verdauungssystem	62,5	37,5	8
Sonstige	66,7	33,3	12
<b>Gesamt</b>	<b>60,6</b>	<b>39,4</b>	<b>578</b>

**Tabelle 8:** (Fortsetzung)

Variable	GBA-Beschluss		Gesamt
	Ohne Zusatznutzen	Mit Zusatznutzen	
Jahrestherapiekosten pro Patient (EUR, in Tsd.)			
Mittelwert	354,4	134,6	267,3
Standardabweichung	2.819,5	290,8	2.199,3
Median	15,4	74,0	41,2
Zielpopulation innerhalb der Gesetzlichen Krankenversicherung (in Tsd.)			
Mittelwert	141,6	68,3	112,6
Standardabweichung	495,2	282,9	425,5
Median	16,3	2,9	6,2
Potenzieller Budget Impact: Zielpopulation (EUR, in Mio.)			
Mittelwert	1.620	472	1.160
Standardabweichung	12.100	911	9.460
Median	143	161	150
Inkrementeller Budget Impact: Zielpopulation (EUR, in Mio.)			
Mittelwert	1.340	262	922
Standardabweichung	12.100	596	9.440
Median	22	77	43

Das größte Indikationsgebiet stellen mit insgesamt 176 FNBs (30 %) die onkologischen Therapien dar. Der GBA bestätigte für 84/176 (48 %, inklusive 28 ODs) Krebstherapien einen positiven Zusatznutzen, das IQWiG stellte für 76/176 (43 %) Therapien eine positive Bewertungsempfehlung aus. Demgegenüber wurden durch den GBA 144/402 (36 %, inklusive 40 ODs) und durch das IQWiG 113/398 (28 %) nicht-onkologische Behandlungsinterventionen positiv bewertet. Darunter sind die Therapien für Stoffwechsel- (GBA, 125; IQWiG, 122) und Infektionserkrankungen (118) am häufigsten vertreten. Während 45 % respektive 28 % der Infektionsbehandlungen durch GBA beziehungsweise IQWiG positiv bewertet wurden, sind nur etwa ein Fünftel der Stoffwechseltherapien (GBA, 22 %; IQWiG, 21 %) mit einem Zusatznutzen belegt worden. Lediglich Therapien von Muskel-Skelett-Erkrankungen (14 %) sowie psychischen Leiden (0 %) wurden noch seltener positiv bewertet, letztere blieben für den vorliegenden Untersuchungszeitraum sogar ganz ohne positive Zusatznutzenbewertung. Wirkstoffe, die dem Indikationsgebiet der Hautkrankheiten zugeordnet werden können, wurden vergleichsweise am häufigsten positiv bewertet (GBA, 71 %; IQWiG, 69 %).



**Tabelle 9:** Deskriptive Ergebnisse für IQWiG auf Teilpopulationsebene  
(Eigene Darstellung)

Variable	IQWiG-Bewertung		Gesamt (n)
	Ohne Zusatznutzen (%)	Mit Zusatznutzen (%)	
<b>Orphan Drug-Status</b>			
Nein	76,1	23,9	506
Ja	0,0	100,0	68
<b>Zweckmäßige Vergleichstherapie (ZVT)</b>			
Nicht vorhanden	31,1	68,9	106
Vorhanden	75,2	24,8	468
<b>Klinische Studienevidenz: Randomisierte kontrollierte Studie (RCTs als systematischer Review/ Meta-Analyse)</b>			
Nicht vorhanden	98,7	1,3	307
Vorhanden	41,0	59,0	200
<b>Vergleichende klinische Wirksamkeit: Signifikanter Unterschied in den patientenrelevanten Endpunkten</b>			
Nicht vorhanden	99,7	0,3	338
Vorhanden	28,4	71,6	169
<b>Veröffentlichungsjahr</b>			
2011	75,0	25,0	4
2012	48,4	51,6	31
2013	69,4	30,6	49
2014	68,3	31,7	63
2015	69,4	30,6	124
2016	71,6	28,4	148
2017	66,7	33,3	96
2018 (Januar-Juni)	57,6	42,4	59
<b>Indikation / Therapeutisches Gebiet</b>			
Atmungssystem	61,5	38,5	26
Augen	72,7	27,3	11
Blut und blutbildende Organe	71,4	28,6	7
Haut	31,3	68,8	16
Herz-Kreislauf	50,0	50,0	20
Infektionen	72,0	28,0	118
Muskel-Skelett-System	86,4	13,6	22
Nervensystem	65,4	34,6	26
Onkologie	56,8	43,2	176
Psychische Gesundheit	100,0	0,0	7
Stoffwechselsystem	79,5	20,5	122
Urogenital	66,7	33,3	3
Verdauungssystem	62,5	37,5	8
Sonstige	75,0	25,0	12
<b>Gesamt</b>	<b>67,9</b>	<b>32,1</b>	<b>574</b>

**Tabelle 9:** (Fortsetzung)

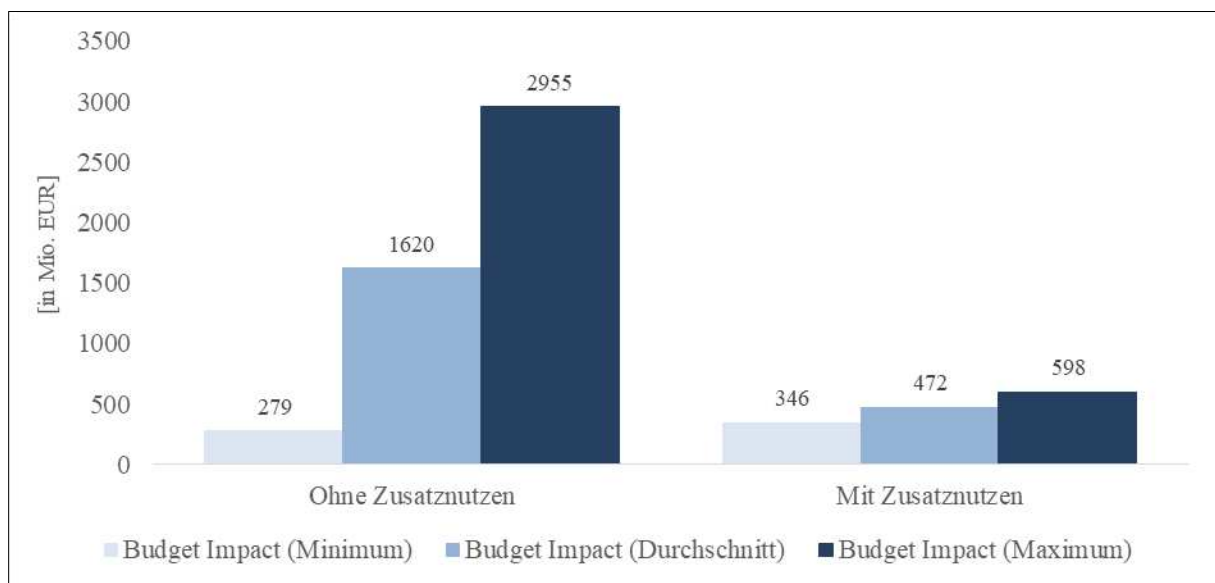
Variable	IQWiG-Bewertung		Gesamt
	Ohne Zusatznutzen	Mit Zusatznutzen	
Jahrestherapiekosten pro Patient (EUR, in Tsd.)			
Mittelwert	48,8	140,5	79,4
Standardabweichung	74,2	296,0	186,1
Median	19,7	64,9	39,0
Zielpopulation innerhalb der Gesetzlichen Krankenversicherung (in Tsd.)			
Mittelwert	140,0	82,6	120,7
Standardabweichung	487,9	315,1	438,1
Median	11,4	3,0	6,4
Potenzieller Budget Impact: Zielpopulation (EUR, in Mio.)			
Mittelwert	543	618	568
Standardabweichung	1.470	1.350	1.430
Median	144	207	158
Inkrementeller Budget Impact: Zielpopulation (EUR, in Mio.)			
Mittelwert	190	317	233
Standardabweichung	1.120	891	1.050
Median	23	79	41

Die Angaben zu den Jahrestherapiekosten pro Patient sind für den GBA (Mittelwert; GBA, 267.279 EUR) im Durchschnitt erheblich höher als für das IQWiG (Mittelwert; IQWiG, 79.400 EUR). Interessanterweise berichteten die GBA-Beschlussfassungen höhere Kosten für Therapien ohne einen festgestellten Zusatznutzen gegenüber Therapien mit positivem Zusatznutzen (Mittelwert<sup>50</sup>; 354.445 EUR versus 134.618 EUR). Demgegenüber finden sich in den IQWiG-Bewertungen höhere Kosten für Wirkstoffe mit positiver Empfehlung im Vergleich zu Therapien ohne Zusatznutzenempfehlung (Mittelwert; 140.532 EUR versus 48.834 EUR). Die durchschnittlichen Angaben zu den Patientenpopulationen stimmen für GBA und IQWiG weitgehend überein (Mittelwert; 112.644 versus 120.731). Sowohl in den GBA-Beschlüssen (Mittelwert; 141.600 versus 68.300) als auch in den IQWiG-Bewertungen (Mittelwert; 140.000 versus 82.600) liegen die Angaben für die Patientenpopulationen ohne Zusatznutzen deutlich über den durchschnittlichen Mittelwerten mit einem positiven Zusatznutzen.

Entsprechend wurden auch die Schätzungen zum potenziellen BI (Mittelwert; GBA, 1.160 Mio. EUR versus IQWiG, 568 Mio. EUR) und inkrementellen BI (Mittelwert; GBA, 922 Mio. EUR

<sup>50</sup> Die Medianwerte (GBA) für die Jahrestherapiekosten pro Patient liegen deutlich unter den Mittelwerten (Durchschnitt, 41.200 EUR) und sind für Therapien ohne Zusatznutzen (15.400 EUR) niedriger als für Interventionen mit Zusatznutzen (74.000 EUR).

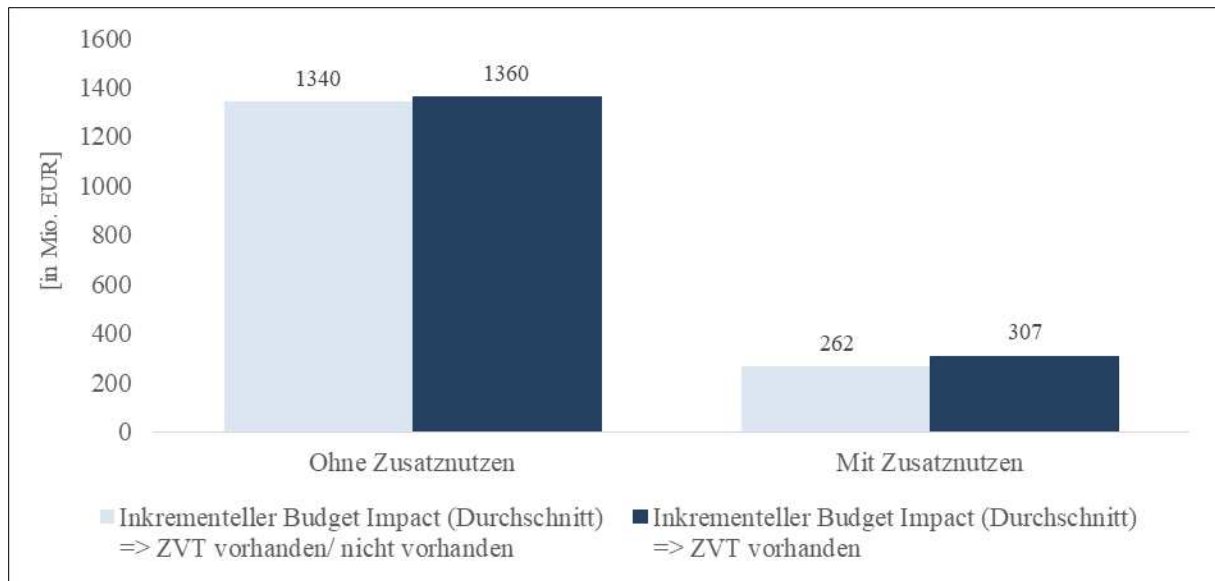
versus IQWiG, 233 Mio. EUR) von den beiden Variablen Patientenzahl und Therapiekosten beeinflusst. Für den GBA sind sowohl der potenzielle (Mittelwert<sup>51</sup>; 1.620 Mio. EUR versus 472 Mio. EUR) als auch der inkrementelle (Mittelwert<sup>52</sup>; 1.340 Mio. EUR versus 262 Mio. EUR) BI für Therapien ohne festgestellten Zusatznutzen im Durchschnitt höher als für Therapien mit einem positiven Zusatznutzen (vgl. Abbildung 8 und Abbildung 9). Die durchschnittlichen Schätzungen zum potenziellen BI (Mittelwert; 543 Mio. EUR versus 618 Mio. EUR) und inkrementellen BI (Mittelwert; 190 Mio. EUR versus 317 Mio. EUR) für IQWiG zeigen wiederum niedrigere Werte für Therapien ohne gegenüber Behandlungen mit positivem Zusatznutzen auf (vgl. Abbildung 10 und Abbildung 11). Die unterschiedlichen Szenarien zum potenziellen BI stützen die dargelegten Mittelwerte für GBA und IQWiG, auch wenn sie für die GBA-Beschlussfassungen ohne positiven Zusatznutzen eine große Spannweite ausweisen (Minimum, 279 Mio. EUR versus Maximum, 2.955 Mio. EUR).



**Abbildung 8:** Budget Impact-Szenarien – GBA-Beschlüsse auf Teilpopulationsebene (Eigene Darstellung)

<sup>51</sup> Die Medianwerte (GBA) für den potenziellen BI liegen deutlich unter den Mittelwerten (Durchschnitt, 150 Mio. EUR) und sind für Therapien ohne Zusatznutzen (143 Mio. EUR) etwas niedriger als für Interventionen mit Zusatznutzen (161 Mio. EUR), was insbesondere auf die Jahrestherapiekosten zurückzuführen ist.

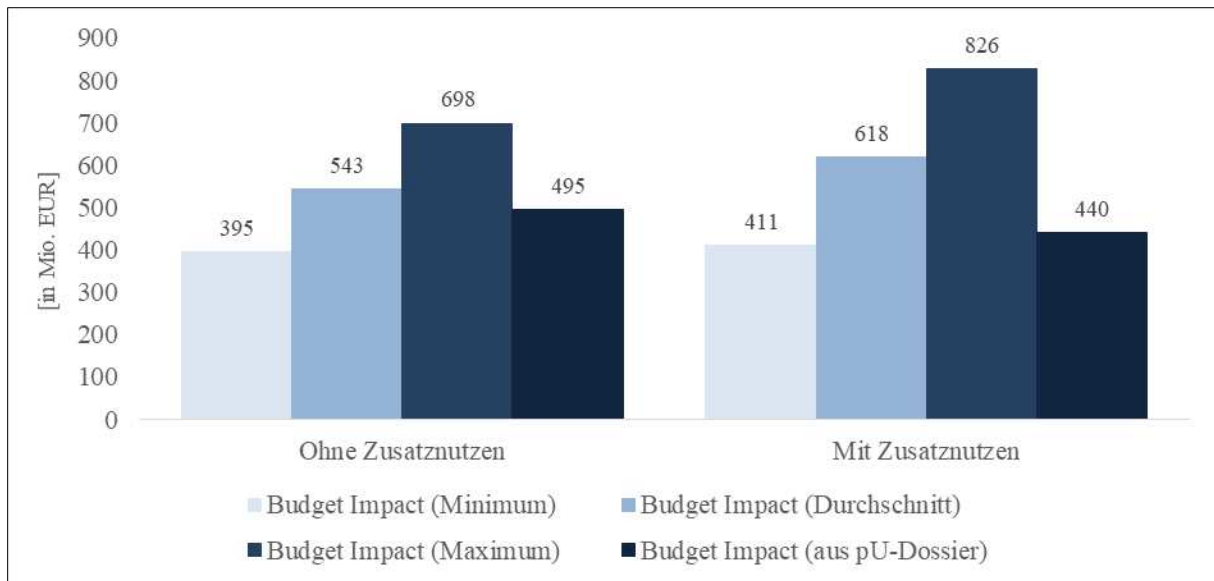
<sup>52</sup> Die Medianwerte (GBA) für den inkrementellen BI liegen ebenfalls unter den Mittelwerten (Durchschnitt, 43 Mio. EUR) und sind für Therapien ohne Zusatznutzen (22 Mio. EUR) niedriger als für Interventionen mit Zusatznutzen (77 Mio. EUR).



**Abbildung 9:** Inkrementelle Budget Impact-Szenarien – GBA-Beschlüsse auf Teilpopulationsebene (Eigene Darstellung)

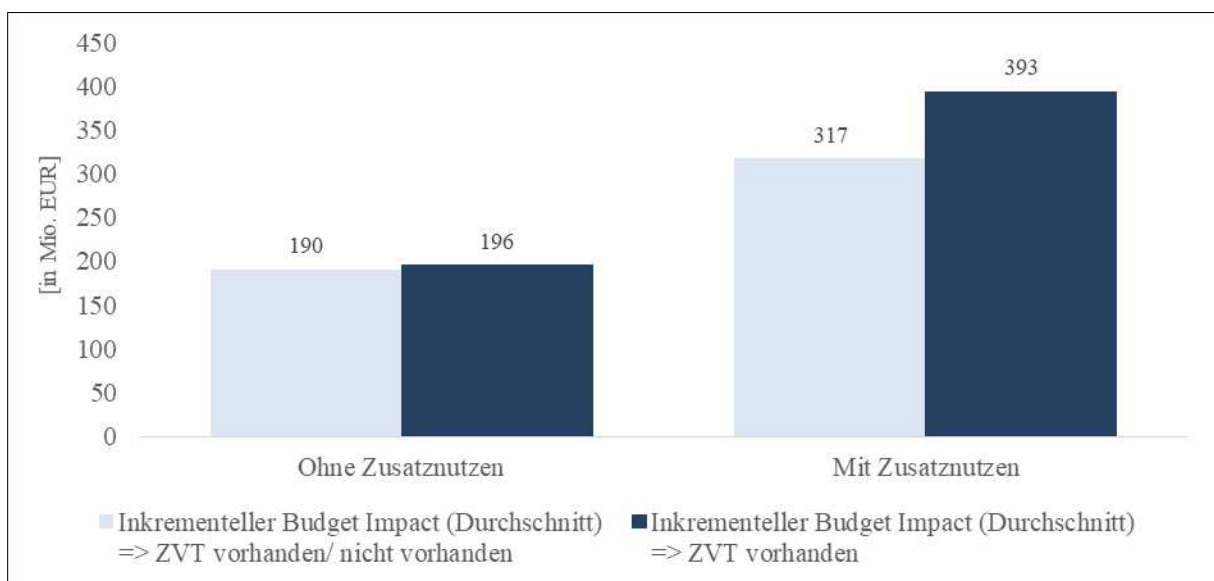
ZVT, Zweckmäßige Vergleichstherapie.

Während die in den jeweiligen pU-Dossiers angegebenen Behandlungskosten (Mittelwert, 80.275 EUR) vergleichbar sind, konnten im Vergleich zu den IQWiG-Nutzenbewertungen durchschnittlich größere Patientenzahlen (Mittelwert, 156.218) bei einer insgesamt kleineren Anzahl an Teilpopulationen je eingereichter Indikation festgestellt werden. Demnach ist es nicht weiter überraschend, dass die in den pU-Dossiers angegebenen Werte zu den gesamten Jahrestherapiekosten (Mittelwert, 469 Mio. EUR), die entsprechend dem potenziellen BI gegenübergestellt werden können, niedriger sind als die durchschnittlichen Schätzungen für das IQWiG. Im Gegensatz zu den IQWiG-Bewertungen – und demnach in Übereinstimmung mit den Berechnungen für den GBA – liegen die durchschnittlichen Jahrestherapiekosten insgesamt für Subgruppen ohne Zusatznutzen (495 Mio. EUR) über den Angaben für Subgruppen mit Zusatznutzen (440 Mio. EUR).



**Abbildung 10:** Budget Impact-Szenarien – IQWiG-Bewertungen und pU-Dossiers auf Teilpopulationsebene (Eigene Darstellung)

pU, pharmazeutisches Unternehmen.



**Abbildung 11:** Inkrementelle Budget Impact-Szenarien – IQWiG-Bewertungen auf Teilpopulationsebene (Eigene Darstellung)

ZVT, Zweckmäßige Vergleichstherapie.

### 5.2.1.2 *Resultate auf Verfahrensebene*

Die Ergebnisse auf Teilpopulationsebene werden durch die Untersuchungsergebnisse auf Verfahrensebene sowohl für den GBA als auch IQWiG weitestgehend gestützt. Insgesamt fallen die FNB-Resultate auf Indikationsebene dabei positiver (mit Zusatznutzen: GBA, 151/262, 58 %; IQWiG, 143/259, 55 %) aus, was auf die zusätzliche Outcome-Variable 'restriktiv positiver Zusatznutzen' (GBA, 56/262, 21 %; IQWiG, 57/259, 22 %) zurückzuführen ist (vgl. Tabelle 10 und Tabelle 11).

Dies lässt sich prinzipiell an allen berücksichtigten (potenziellen) Bewertungsfaktoren sowohl für die GBA-Beschlussfassungen als auch die IQWiG-Bewertungen aufzeigen. Zum Beispiel liegt für die beiden Veröffentlichungsjahre 2015 und 2017 der Anteil der (restriktiv) positiven Resultate von GBA (61 % und 56 %) und IQWiG (58 % und 56 %) im Gegensatz zur Untersuchung auf Subgruppenebene anteilig über den negativen Ergebnissen. Für die Bewertungsfaktoren klinische Studienevidenz (GBA, 19 %; IQWiG, 17 %) und die vergleichende klinische Wirksamkeit (GBA, 19 %; IQWiG, 16 %) zeigt sich, dass sich der relative Anteil positiver FNBS auch bei Nicht-Vorhandensein der beiden Kriterien deutlich erhöht. Bei den therapeutischen Gebieten, exemplarisch an den Wirkstoffen für die Behandlung von Krebserkrankungen (GBA, 74 %; IQWiG, 68 %), kardiovaskulären Krankheitsbildern (GBA, 73 %; IQWiG, 64 %) oder Atemwegserkrankungen (GBA, 82 %; IQWiG, 64 %) zu betrachten, lässt sich ebenfalls ein deutlicher Anstieg der positiven FNBS gegenüber den Ergebnissen auf Teilpopulationsebene erkennen.

Die abweichenden Erkenntnisse auf Verfahrensebene gelten ausschließlich für die Behandlungsinterventionen, die sowohl positive als auch negative Bewertungen respektive Beschlüsse für die verschiedenen Teilpopulationen innerhalb eines Anwendungsgebiets einer Therapie aufweisen und demzufolge auch ein restriktiv positiver Zusatznutzen vorliegen konnte.

**Tabelle 10:** Deskriptive Ergebnisse für GBA auf Verfahrensebene  
(Eigene Darstellung)

Variable	GBA-Beschluss			Gesamt (n)
	Ohne Zusatznutzen (%)	Restriktiver Zusatznutzen (%)	Mit Zusatznutzen (%)	
<b>Orphan Drug-Status</b>				
Nein	53,1	26,8	20,1	209
Ja	0,0	0,0	100,0	53
<b>Zweckmäßige Vergleichstherapie (ZVT)</b>				
Nicht vorhanden	6,5	1,6	91,9	62
Vorhanden	53,5	27,5	19,0	200
<b>Klinische Studienevidenz: Randomisierte kontrollierte Studie (RCTs als systematischer Review/ Meta-Analyse)</b>				
Nicht vorhanden	81,0	16,2	2,9	105
Vorhanden	25,0	37,5	37,5	104
<b>Vergleichende klinische Wirksamkeit: Signifikanter Unterschied in den patientenrelevanten Endpunkten</b>				
Nicht vorhanden	81,3	17,9	0,9	112
Vorhanden	20,6	37,1	42,3	97
<b>Veröffentlichungsjahr</b>				
2011	0,0	100,0	0,0	1
2012	18,8	18,8	62,5	16
2013	56,0	12,0	32,0	25
2014	51,9	22,2	25,9	27
2015	38,8	26,5	34,7	49
2016	44,1	25,0	30,9	68
2017	44,4	17,8	37,8	45
2018 (Januar-Juni)	35,5	16,1	48,4	31
<b>Indikation / Therapeutisches Gebiet</b>				
Atmungssystem	18,2	45,5	36,4	11
Augen	66,7	11,1	22,2	9
Blut und blutbildende Organe	71,4	0,0	28,6	7
Haut	33,3	11,1	55,6	9
Herz-Kreislauf	27,3	36,4	36,4	11
Infektionen	42,9	35,7	21,4	28
Muskel-Skelett-System	62,5	12,5	25,0	8
Nervensystem	64,7	5,9	29,4	17
Onkologie	26,3	29,3	44,4	99
Psychische Gesundheit	100,0	0,0	0,0	4
Stoffwechselsystem	57,8	8,9	33,3	45
Urogenital	66,7	0,0	33,3	3
Verdauungssystem	40,0	0,0	60,0	5
Sonstige	66,7	0,0	33,3	6
<b>Gesamt</b>	<b>42,4</b>	<b>21,4</b>	<b>36,3</b>	<b>262</b>

**Tabelle 10:** (Fortsetzung)

Variable	GBA-Beschluss			Gesamt
	Ohne Zusatznutzen	Restriktiver Zusatznutzen	Mit Zusatznutzen	
Jahrestherapiekosten pro Patient (EUR, in Tsd.)				
Mittelwert	299,3	63,9	166,0	199,9
Standardabweichung	2.527,3	51,6	229,7	1.640,7
Median	9,5	65,7	88,9	47,4
Zielpopulation innerhalb der Gesetzlichen Krankenversicherung (in Tsd.)				
Mittelwert	378,3	311,3	62,9	247,2
Standardabweichung	906,3	697,8	297,3	707,5
Median	56,8	17,6	1,4	7,8
Potenzieller Budget Impact: Zielpopulation (EUR, in Mio.)				
Mittelwert	4.720	1.590	640	2.560
Standardabweichung	40.700	2.410	1.360	26.400
Median	447	529	132	344
Inkrementeller Budget Impact: Zielpopulation (EUR, in Mio.)				
Mittelwert	4.110	742	402	2.030
Standardabweichung	40.700	1.780	815	26.300
Median	69	220	96	107

Interessanterweise weisen die Resultate auf Indikationsebene sowohl für GBA als auch IQWiG durchschnittlich geringere Jahrestherapiekosten<sup>53</sup> pro Patient auf, wenn man die positiv restriktiven (Mittelwert; GBA, 63.900 EUR versus IQWiG, 53.000 EUR) mit den positiven (Mittelwert; GBA, 166.000 EUR versus IQWiG, 190.800 EUR) und negativen (Mittelwert; GBA, 299.300 EUR versus IQWiG, 54.300 EUR) FNB-Ergebnissen vergleicht. Während die durchschnittliche Größe der Patientengruppen bei negativen FNBs (Mittelwert; GBA, 378.300 versus IQWiG, 382.500) im Vergleich zu den restriktiv positiven (Mittelwert; GBA, 311.300 versus IQWiG, 318.200) und positiven (Mittelwert; GBA, 62.900 versus IQWiG, 64.500) Resultaten ebenfalls höher ist, zeigen sich für die BI-Projektionen teils deutliche Unterschiede gegenüber den Szenarien auf Teilpopulationsebene (vgl. Abbildung 12 und Abbildung 13).

<sup>53</sup> Die Medianwerte für die Jahrestherapiekosten pro Patient liegen sowohl für GBA (Durchschnitt, 47.400 EUR) als auch IQWiG (Durchschnitt, 46.400 EUR) unter den Mittelwerten. Für den GBA liegen die Medianwerte im Gegensatz zu den Mittelwerten für Therapien ohne Zusatznutzen (9.500 EUR) unter den Interventionen mit Zusatznutzen (eingeschränkt, 65.700 EUR; uneingeschränkt, 88.900 EUR).



**Tabelle 11:** Deskriptive Ergebnisse für IQWiG auf Verfahrensebene  
(Eigene Darstellung)

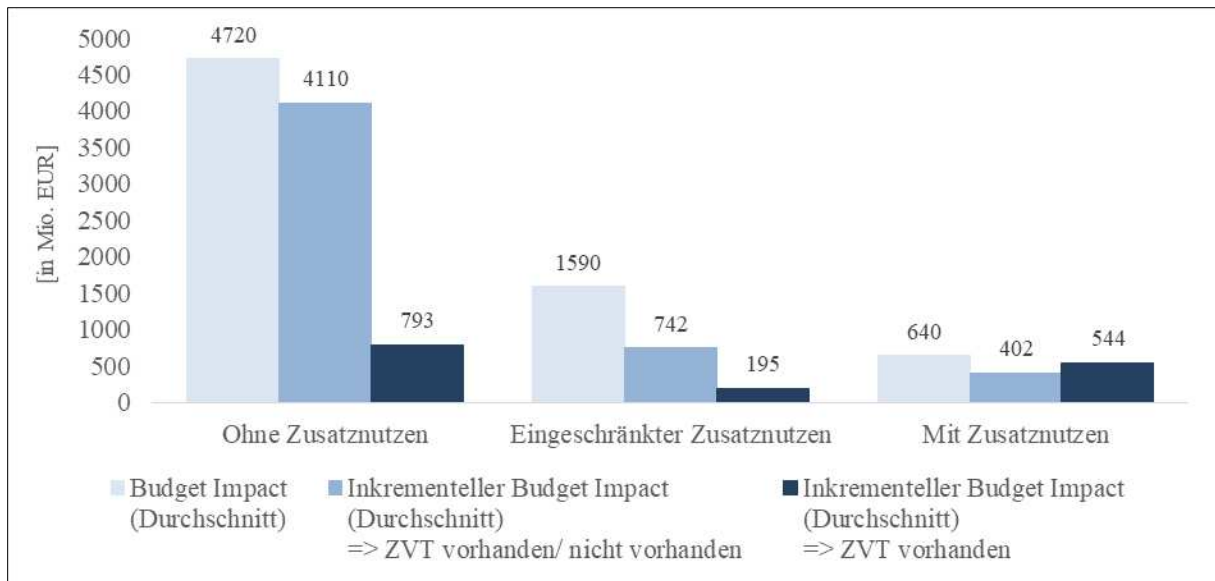
Variable	IQWiG-Bewertung			Gesamt (n)
	Ohne Zusatznutzen (%)	Restriktiver Zusatznutzen (%)	Mit Zusatznutzen (%)	
<b>Orphan Drug-Status</b>				
Nein	56,3	27,7	16,0	206
Ja	0,0	0,0	100,0	53
<b>Zweckmäßige Vergleichstherapie (ZVT)</b>				
Nicht vorhanden	10,9	1,6	87,5	64
Vorhanden	55,9	28,7	15,4	195
<b>Klinische Studienevidenz: Randomisierte kontrollierte Studie (RCTs als systematischer Review/ Meta-Analyse)</b>				
Nicht vorhanden	83,3	16,7	0,0	114
Vorhanden	22,8	41,3	35,9	92
<b>Vergleichende klinische Wirksamkeit: Signifikanter Unterschied in den patientenrelevanten Endpunkten</b>				
Nicht vorhanden	84,4	15,6	0,0	122
Vorhanden	15,5	45,2	39,3	84
<b>Veröffentlichungsjahr</b>				
2011	0,0	100,0	0,0	1
2012	37,5	25,0	37,5	16
2013	54,2	16,7	29,2	24
2014	51,9	22,2	25,9	27
2015	41,7	29,2	29,2	48
2016	49,3	16,4	34,3	67
2017	44,4	22,2	33,3	45
2018 (Januar-Juni)	32,3	22,6	45,2	31
<b>Indikation / Therapeutisches Gebiet</b>				
Atmungssystem	36,4	36,4	27,3	11
Augen	66,7	11,1	22,2	9
Blut und blutbildende Organe	71,4	0,0	28,6	7
Haut	25,0	25,0	50,0	8
Herz-Kreislauf	36,4	36,4	27,3	11
Infektionen	42,9	46,4	10,7	28
Muskel-Skelett-System	62,5	25,0	12,5	8
Nervensystem	70,6	5,9	23,5	17
Onkologie	32,3	23,2	44,4	99
Psychische Gesundheit	100,0	0,0	0,0	4
Stoffwechselsystem	53,5	11,6	34,9	43
Urogenital	66,7	0,0	33,3	3
Verdauungssystem	40,0	0,0	60,0	5
Sonstige	50,0	33,3	16,7	6
<b>Gesamt</b>	<b>44,8</b>	<b>22,0</b>	<b>33,2</b>	<b>259</b>

**Tabelle 11:** (Fortsetzung)

Variable	IQWiG-Bewertung			Gesamt
	Ohne Zusatznutzen	Restriktiver Zusatznutzen	Mit Zusatznutzen	
Jahrestherapiekosten pro Patient (EUR, in Tsd.)				
Mittelwert	54,3	53,0	190,8	99,4
Standardabweichung	103,8	50,7	329,3	212,3
Median	13,2	39,0	104,2	46,4
Zielpopulation innerhalb der Gesetzlichen Krankenversicherung (in Tsd.)				
Mittelwert	382,5	318,2	64,5	263,1
Standardabweichung	930,1	674,6	318,2	733,6
Median	40,2	26,5	1,5	7,4
Potenzieller Budget Impact: Zielpopulation (EUR, in Mio.)				
Mittelwert	1.200	2.340	531	1.230
Standardabweichung	3.520	4.710	957	3.330
Median	341	680	251	335
Inkrementeller Budget Impact: Zielpopulation (EUR, in Mio.)				
Mittelwert	405	908	375	507
Standardabweichung	2.940	2.770	603	2.390
Median	50	199	133	97

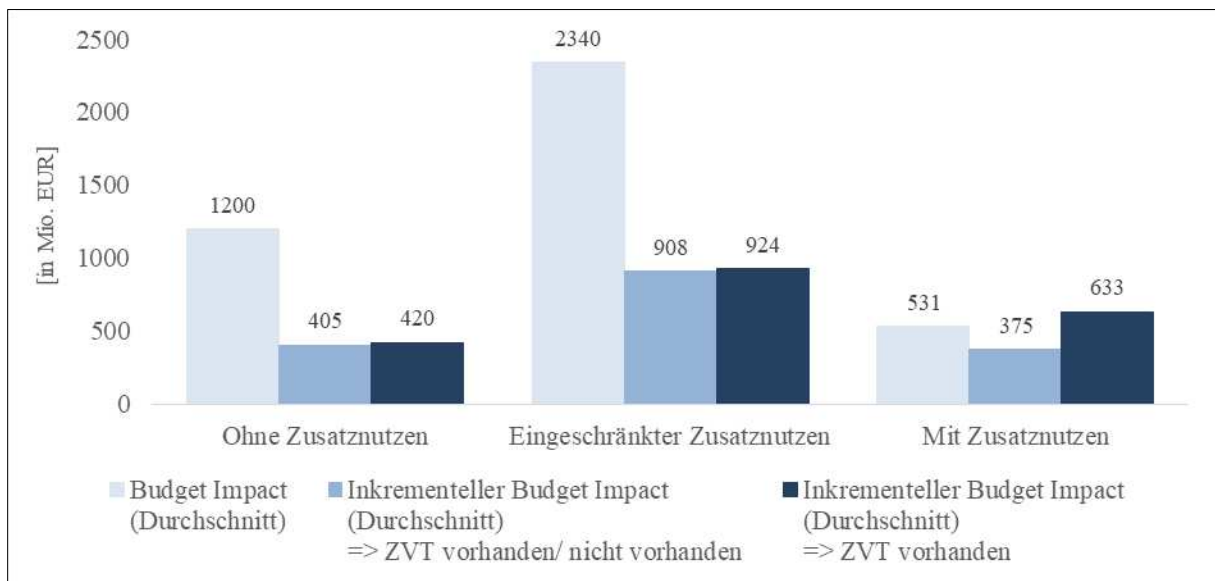
Unverkennbare Differenzen bestehen dabei allen voran bei den IQWiG-Nutzenbewertungen (vgl. hierzu Abbildung 13 gegenüber den Abbildungen 10 und 11); sowohl für die potenziellen BI-Szenarien (Mittelwert, 1.230 Mio. EUR) als auch für die inkrementellen BI-Szenarien (Mittelwert, 507 Mio. EUR) sind die durchschnittlichen Schätzungen für die restriktiv positiven Resultate (Mittelwert; 2.340 Mio. EUR respektive 908 Mio. EUR) deutlich am höchsten. Die potenziellen und inkrementellen BI-Szenarien<sup>54</sup> für den GBA zeigen, dass diese vergleichbar der Untersuchung auf Teilpopulationsebene für negative Beschlüsse (Mittelwert; 4.720 Mio. EUR und 4.110 Mio. EUR) ebenfalls höher sind als für positive (Mittelwert; 640 Mio. EUR und 402 Mio. EUR) beziehungsweise positiv restriktive (Mittelwert; 1.590 Mio. EUR und 742 Mio. EUR) Entscheidungen.

<sup>54</sup> Die Medianwerte (GBA) liegen sowohl für den potenziellen (344 Mio. EUR) als auch den inkrementellen (107 Mio. EUR) BI unter den Mittelwerten. Im Gegensatz zu den Mittelwerten sind die Medianwerte sowohl für den potenziellen (447 Mio. EUR versus 529 Mio. EUR) als auch den inkrementellen (69 Mio. EUR versus 220 Mio. EUR) BI ohne Zusatznutzen niedriger als bei eingeschränkt positivem Zusatznutzen.



**Abbildung 12:** [Inkrementelle] Budget Impact-Szenarien – GBA-Beschlüsse auf Verfahrensebene (Eigene Darstellung)

ZVT, Zweckmäßige Vergleichstherapie.



**Abbildung 13:** [Inkrementelle] Budget Impact-Szenarien – IQWiG-Bewertungen auf Verfahrensebene (Eigene Darstellung)

ZVT, Zweckmäßige Vergleichstherapie.

## 5.2.2 Ergebnisse der Regressionsanalysen für GBA und IQWiG

Die deskriptiven Untersuchungsergebnisse bilden sowohl für GBA als auch IQWiG die grundlegende Evidenz für die (möglichen) Unterschiede bei der Berücksichtigung der einzelnen Bewertungskriterien ab. Ihr jeweiliger Einfluss beziehungsweise Effekt auf die Outcome-Variable in den Beschlussfassungen und Dossierbewertungen wurde entsprechend der Regressionsgleichung [1] auf statistische Signifikanz getestet. Da bei Wirkstoffen mit OD-Status – per Definition – ein positiver Beschluss respektive eine positive Bewertung vorliegt, konnte ein direkter (kausaler) Zusammenhang zwischen den FNB-Outcomes und den (offiziellen oder empirisch als relevant eingestuften) FNB-Kriterien nicht untersucht werden. Demnach sind für ODs ausschließlich die zur Verfügung stehenden Daten in die jeweiligen Regressionsmodelle von GBA und IQWiG eingeschlossen worden.

### 5.2.2.1 Logit-Modell auf Teilpopulationsebene

Die Resultate auf Teilpopulationsebene zeigen demzufolge (vgl. Tabelle 12 und Tabelle 13), dass signifikante Unterschiede in den patientenrelevanten Endpunkten gegenüber der ZVT (vergleichende klinische Wirksamkeit) als ein zentrales Kriterium der FNB einen signifikant positiven Einfluss sowohl auf die GBA-Beschlüsse ( $p < 0,01$ ) als auch die IQWiG-Bewertungsempfehlungen ( $p < 0,01$ ) haben. Die klinische Studienevidenz zeigt einen eindeutig positiven Effekt auf die Bewertungen durch IQWiG ( $p < 0,01$ ), jedoch nicht auf die Beschlussfassungen durch den GBA ( $p > 0,1$ ). Offensichtlich hat das Vorhandensein einer ZVT weder auf die GBA-Beschlüsse ( $p > 0,1$ ) noch auf die IQWiG-Bewertungen ( $p > 0,1$ ) einen Einfluss, wobei zu berücksichtigen bleibt, dass der zuvor genannte Evaluationsfaktor (vergleichende klinische Wirksamkeit) die Berücksichtigung der durch den GBA bestimmten ZVT bereits (implizit) voraussetzt beziehungsweise umfasst.

Beim Betrachten der im Logit-Modell ergänzend berücksichtigten Indikatorvariablen zeigt sich darüber hinaus, dass bei den IQWiG-Bewertungen weder für die Variable Veröffentlichungsjahr noch für das therapeutische Gebiet robuste Ergebnisse vorliegen. Lediglich mit den Veröffentlichungsjahren 2014, 2015, 2016 und 2017 ( $p < 0,1$ ) sowie den Behandlungen für Hauterkrankungen ( $p < 0,1$ ) scheinen die Empfehlungen durch IQWiG zu korrelieren, auch wenn die Resultate nicht durchgängig als robust einzustufen gewesen sind. Die GBA-Beschlüsse weisen einen eindeutig positiven Zusammenhang zu den Therapien für Hauterkrankungen ( $p < 0,05$ ) auf, wohingegen für die Outcomes zu Onkologika ( $p < 0,1$ ) wiederum keine durchgängig robusten Ergebnisse vorliegen (vgl. Anhang, Tabelle A1 und Tabelle A2).

**Tabelle 12: Resultate der Regressionsanalyse für GBA auf Teilpopulationsebene**  
(Eigene Darstellung)

Variable	GBA / Logit-Modell								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Zweckmäßige Vergleichstherapie (Dummy)	-0,677	-0,691	-0,774	-0,699	-0,763	-0,706	-0,756	--	--
Studienevidenz: Randomisierte kontrollierte Studie / RCT (Dummy)	1,143	1,142	1,147	1,140	1,138	1,139	1,134	--	--
Vergleichende klinische Wirksamkeit (Dummy)	0,534	0,416	0,388	0,402	0,366	0,392	0,357	0,458	0,397
Orphan Drug-Status (Dummy)	0,561	0,609	0,612	0,607	0,611	0,606	0,611	0,608	0,627
Potenzieller Budget Impact: Minimum (Log)	5,810***	5,845***	5,900***	5,844***	5,909***	5,844***	5,905***	5,694***	5,779***
Potenzieller Budget Impact: Minimum (Log) <sup>2</sup>	0,716	0,738	0,749	0,738	0,75	0,738	0,749	0,733	0,746
Potenzieller Budget Impact: Minimum (Log)	--	--	--	--	--	--	--	--	--
Potenzieller Budget Impact: Minimum (Log) <sup>2</sup>	--	--	--	--	--	--	--	--	--
Potenzieller Budget Impact: Durchschnitt (Log)				0,012	0,664				0,860
Potenzieller Budget Impact: Durchschnitt (Log) <sup>2</sup>				0,099	1,042				1,075
Potenzieller Budget Impact: Maximum (Log)					-0,018				-0,024
Potenzieller Budget Impact: Maximum (Log) <sup>2</sup>					0,029				0,030
Inkrementeller Budget Impact: Durchschnitt (Log)						0,021	0,687		
Konstante	-5,445	-5,443	-9,64	-5,627	-11,442	-5,767	-11,781	-6,034**	-13,812
Veröffentlichungsjahr (Dummies)	3,457	3,677	8,845	3,709	9,981	3,717	10,388	3,014	10,254
Indikation / Therapeutisches Gebiet (Dummies)	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Beobachtungen	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja
	493	486	486	486	486	486	486	474	474

Regressionsergebnisse sind statistisch signifikant bei: \*\*\* p < 0,01; \*\* p < 0,05; \* p < 0,1.

**Tabelle 13:** Resultate der Regressionsanalyse für IQWiG auf Teilpopulationsebene  
(Eigene Darstellung)

Variable	IQWiG / Logit-Modell								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Zweckmäßige Vergleichstherapie (Dummy)	-0,185	-0,266	-0,232	-0,126	-0,098	-0,094	-0,081	--	--
Studienevidenz: Randomisierte kontrollierte Studie / RCT (Dummy)	1,105	1,243	1,241	1,248	1,246	1,248	1,248	--	--
Vergleichende klinische Wirksamkeit (Dummy)	2,810***	2,549***	2,503***	2,738***	2,717***	2,800***	2,789***	2,935***	2,901***
Orphan Drug-Status (Dummy)	0,811	0,788	0,794	0,799	0,802	0,805	0,808	0,833	0,830
Potenzieller Budget Impact: Minimum (Log)	6,175***	6,449***	6,467***	6,326***	6,318***	6,301***	6,296***	5,858***	6,263***
Potenzieller Budget Impact: Minimum (Log) <sup>2</sup>	1,159	1,233	1,231	1,210	1,206	1,207	1,205	1,153	1,227
Potenzieller Budget Impact: Durchschnitt (Log)	--	--	--	--	--	--	--	--	--
Potenzieller Budget Impact: Maximum (Log)	--	0,254**	0,776	--	--	--	--	--	--
Potenzieller Budget Impact: Durchschnitt (Log) <sup>2</sup>		0,116	1,158						
Potenzieller Budget Impact: Maximum (Log) <sup>2</sup>		-0,015	0,032						
Inkrementeller Budget Impact: Durchschnitt (Log)				0,278**	0,714	0,279**	0,501		0,418
				0,129	1,377	0,130	1,419		1,392
					-0,012		-0,006		-0,002
					0,037		0,038		0,038
Konstante	-6,739	-11,071***	-15,728	-12,068***	-16,091	-12,257***	-14,331	-7,678***	-14,087
Veröffentlichungsjahr (Dummies)	5,288	3,017	10,731	3,363	13,120	3,452	13,638	2,152	13,094
Indikation / Therapeutisches Gebiet (Dummies)	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Beobachtungen	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja
	488	470	470	471	471	471	471	443	443

Regressionsergebnisse sind statistisch signifikant bei: \*\*\* p < 0,01; \*\* p < 0,05; \* p < 0,1.

Während die unterschiedlichen potenziellen BI-Szenarien<sup>55</sup> keinen Zusammenhang zu den GBA-Beschlussfassungen aufzeigen ( $p > 0,1$ ), korrelieren alle drei potenziellen BI-Szenarien (Minimum; Durchschnitt; Maximum) mit den IQWiG-Dossierbewertungen ( $p < 0,05$ ; positiver linearer Zusammenhang des potenziellen BI und der positiven Bewertungsempfehlungen). Dabei scheint die Wahrscheinlichkeit einer positiven Empfehlung durch IQWiG ab einem bestimmten (nicht eindeutig quantifizierbaren) Wertebereich abzunehmen (sogenannter U-Kurven-Effekt), jedoch stellten sich die Ergebnisse für keines der drei potenziellen BI-Szenarien als signifikant heraus. Demgegenüber konnte beim inkrementellen BI-Szenario weder für die GBA-Beschlüsse ( $p > 0,1$ ) noch für die IQWiG-Bewertungen ( $p > 0,1$ ) ein eindeutiger Zusammenhang festgestellt werden. Die vorliegenden Resultate beider Regressionsmodelle erweisen sich darüber hinaus als robust, da sie bei Ausschluss der jeweiligen Extremwerte, das heißt der obersten fünf Prozent der BI-Schätzwerte, keine signifikanten Abweichungen aufzeigten.

#### 5.2.2.2 Ordinales Logit-Modell auf Verfahrensebene

Die deskriptiven Analyseergebnisse auf Indikationsebene weisen sowohl für GBA als auch für IQWiG weitgehend vergleichbare Ergebnisse aus. Beim Betrachten der deskriptiven Evidenz für die Regressionsanalyse sind insbesondere die Unterschiede zwischen den eingeschränkt positiven und den uneingeschränkt positiven FNBs explizit herauszustellen. Während bei den uneingeschränkt positiven Resultaten sowohl für die GBA-Beschlüsse als auch die IQWiG-Bewertungen (fast) ausnahmslos die Kriterien klinische Studienevidenz (GBA, 93 %; IQWiG, 100 %) sowie die vergleichende klinische Wirksamkeit gegenüber der ZVT (GBA, 98 %; IQWiG, 100 %) erfüllt sind, zeigen die uneingeschränkt positiven FNB-Ergebnisse einen deutlich geringeren prozentualen Anteil (GBA, 70 % und 64 %; IQWiG, 67 % und 67 %) auf.

Im Unterschied zur Analyse der Teilpopulationen zeigt das ordinale Logit-Modell für IQWiG, dass das Vorhandensein einer ZVT einen Effekt auf eine (eingeschränkt) positive Bewertung zu haben scheint ( $p < 0,1$ ), der sich jedoch ähnlich wie für die klinische Studienevidenz ( $p < 0,1$ ) als nicht robust herausstellte. Für die GBA-Beschlussfassungen zeigt sich ebenfalls, dass das Vorhandensein relevanter RCTs oder einer ZVT einen stärkeren Einfluss auf eine positive

---

<sup>55</sup> Während die im Logit-Modell zugrunde liegende Kontrollvariable Jahrestherapiekosten pro Patient sowohl für GBA ( $p < 0,001$ ; zunehmende Kosten und negativer Beschluss) als auch IQWiG ( $p < 0,001$ ; zunehmende Kosten und positive Bewertung) signifikante lineare Zusammenhänge für alle drei Szenarien (Minimum; Durchschnitt; Maximum) aufweist, zeigt sich für die Patientenpopulation keine eindeutige Korrelation zu den Beschlussfassungen respektive Empfehlungen (vgl. hierzu auch Anhang, Tabelle A1 und Tabelle A2). Eine gesonderte Darstellung der Variable Jahrestherapiekosten insgesamt, die aus den jeweils zugehörigen pU-Dossiers (der IQWiG-Nutzenbewertungen) stammen, befindet sich im Anhang (vgl. Tabelle A2).

FNB zu haben scheint als auf Subgruppenebene, jedoch erwiesen sich die Regressionsergebnisse als nicht signifikant. Die vergleichende klinische Wirksamkeit zeigt wiederum sowohl zu den (uneingeschränkt) positiven GBA-Beschlüssen ( $p < 0,01$ ) als auch den IQWiG-Bewertungen ( $p < 0,01$ ) einen signifikant positiven Zusammenhang (vgl. Tabelle 14 und Tabelle 15).

Für die nachfolgend nicht explizit dargestellten Indikatorvariablen lässt sich einerseits feststellen, dass im Vergleich zur Subgruppenanalyse keines der Veröffentlichungsjahre einen signifikanten Effekt auf die Dossierbewertungen durch IQWiG aufweist. Andererseits zeigt sich bei den therapeutischen Gebieten für die Behandlungen von Hauterkrankungen ( $p < 0,1$ ) ein positiver und für Onkologika ( $p < 0,05$ ) ein signifikant positiver Zusammenhang zu den GBA-Beschlussfassungen. Ein signifikanter Einfluss auf die IQWiG-Nutzenbewertungen konnte wiederum für Hauterkrankungen ( $p < 0,05$ ) festgestellt werden. Herz-Kreislauf-Erkrankungen ( $p < 0,1$ ) zeigen sowohl für GBA-Beschlüsse als auch IQWiG-Empfehlungen einen Zusammenhang auf, der sich jedoch als nicht robust herausstellte (vgl. Anhang, Tabelle A3).

**Tabelle 14:** Resultate der Regressionsanalyse für GBA auf Verfahrensebene (Eigene Darstellung)

Variable	GBA / Ordinales Logit-Modell				
	(1)	(2)	(3)	(4)	(5)
Zweckmäßige Vergleichstherapie (Dummy)	-1,107 0,776	-1,256 0,811	-1,229 0,817	-- --	-- --
Evidenz: Randomisierte kontrollierte Studie / RCT (Dummy)	0,649 0,567	0,581 0,576	0,589 0,578	0,763 0,583	0,719 0,593
Vergleichende klinische Wirksamkeit (Dummy)	2,519*** 0,571	2,536*** 0,576	2,517*** 0,582	2,127*** 0,596	2,150*** 0,610
Orphan Drug-Status (Dummy)	-- --	-- --	-- --	-- --	-- --
Potenzieller Budget Impact: Durchschnitt (Log)		0,065 0,097	-0,180 0,979		0,105 1,099
Potenzieller Budget Impact: Durchschnitt (Log) <sup>2</sup>			0,007 0,027		-0,001 0,029
Inkrementeller Budget Impact: Durchschnitt (Log)				0,469** 0,216	0,459** 0,217
Konstante 1	3,658 2,347	4,603* 2,741	2,356 9,348	4,129* 2,285	5,564 10,671
Konstante 2	5,915** 2,385	6,862** 2,778	4,616 9,356	6,570*** 2,333	8,012 10,685
Veröffentlichungsjahr (Dummies)	Ja	Ja	Ja	Ja	Ja
Indikation / Therapeutisches Gebiet (Dummies)	Ja	Ja	Ja	Ja	Ja
Beobachtungen	209	207	207	198	198

Regressionsergebnisse sind statistisch signifikant bei: \*\*\*  $p < 0,01$ ; \*\*  $p < 0,05$ ; \*  $p < 0,1$ .



**Tabelle 15:** Resultate der Regressionsanalyse für IQWiG auf Verfahrensebene  
(Eigene Darstellung)

Variable	IQWiG / Ordinales Logit-Modell				
	(1)	(2)	(3)	(4)	(5)
Zweckmäßige Vergleichstherapie (Dummy)	0,961	0,487	0,498	--	--
	1,019	1,184	1,259	--	--
Evidenz: Randomisierte kontrollierte Studie / RCT (Dummy)	0,766	0,578	0,616	0,893	0,768
	0,642	0,675	0,661	0,665	0,682
Vergleichende klinische Wirksamkeit (Dummy)	3,453***	3,622***	3,762***	3,280***	3,724***
	0,650	0,685	0,688	0,676	0,720
Orphan Drug-Status (Dummy)	--	--	--	--	--
	--	--	--	--	--
Potenzieller Budget Impact: Durchschnitt (Log)		0,394***	4,841**		4,795**
		0,137	2,127		2,158
Potenzieller Budget Impact: Durchschnitt (Log) <sup>2</sup>			-0,112**		-0,110**
			0,053		0,054
Inkrementeller Budget Impact: Durchschnitt (Log)				0,125	-0,016
				0,201	0,209
Konstante 1	5,954**	12,162***	55,984***	5,091**	55,326**
	2,639	3,527	21,410	2,490	21,681
Konstante 2	8,761***	15,106***	59,033***	8,025***	58,545***
	2,712	3,627	21,498	2,571	21,786
Veröffentlichungsjahr (Dummies)	Ja	Ja	Ja	Ja	Ja
Indikation / Therapeutisches Gebiet (Dummies)	Ja	Ja	Ja	Ja	Ja
Beobachtungen	205	204	204	194	194

Regressionsergebnisse sind statistisch signifikant bei: \*\*\*  $p < 0,01$ ; \*\*  $p < 0,05$ ; \*  $p < 0,1$ .

Das potenzielle BI-Szenario<sup>56</sup> weist einen positiven Zusammenhang zu den Empfehlungen durch IQWiG ( $p < 0,05$ ) auf, der bis zu einem bestimmten Grenzwert (von schätzungsweise 655 Mio. EUR) zu beobachten ist, bevor sich die Wahrscheinlichkeit für eine negative Bewertung zu erhöhen scheint ( $\Rightarrow$  inverser U-Kurven-Effekt). Demgegenüber zeigt sich kein Effekt für die GBA-Beschlüsse ( $p > 0,1$ ). Im Unterschied zur Subgruppenanalyse zeigt das Szenario für den inkrementellen BI einen signifikanten Zusammenhang zu den GBA-Beschlussfassungen auf ( $p < 0,05$ ) – jedoch nicht für die Bewertungen durch IQWiG ( $p > 0,1$ ). Demnach scheint die positiv lineare Assoziation in unmittelbarem Zusammenhang mit dem BI der ZVTs zu stehen. Vergleichbar den Ergebnissen auf Teilpopulationsebene zeigten sich auch die Resultate der Regressionsmodelle auf Verfahrensebene bei Ausschluss der Extremwerte als robust.

<sup>56</sup> Die im ordinalen Logit-Modell vorliegende Kontrollvariable Jahrestherapiekosten weist dabei sowohl für GBA ( $p < 0,01$ ; zunehmende Kosten und negativer Beschluss) als auch IQWiG ( $p < 0,01$ ; zunehmende Kosten und positive Bewertung) einen signifikanten Einfluss auf die Outcome-Variablen auf (vgl. auch Anhang, Tabelle A3).

### 5.2.3 *GBA-Beschlussfassungen versus IQWiG-Dossierbewertungen*

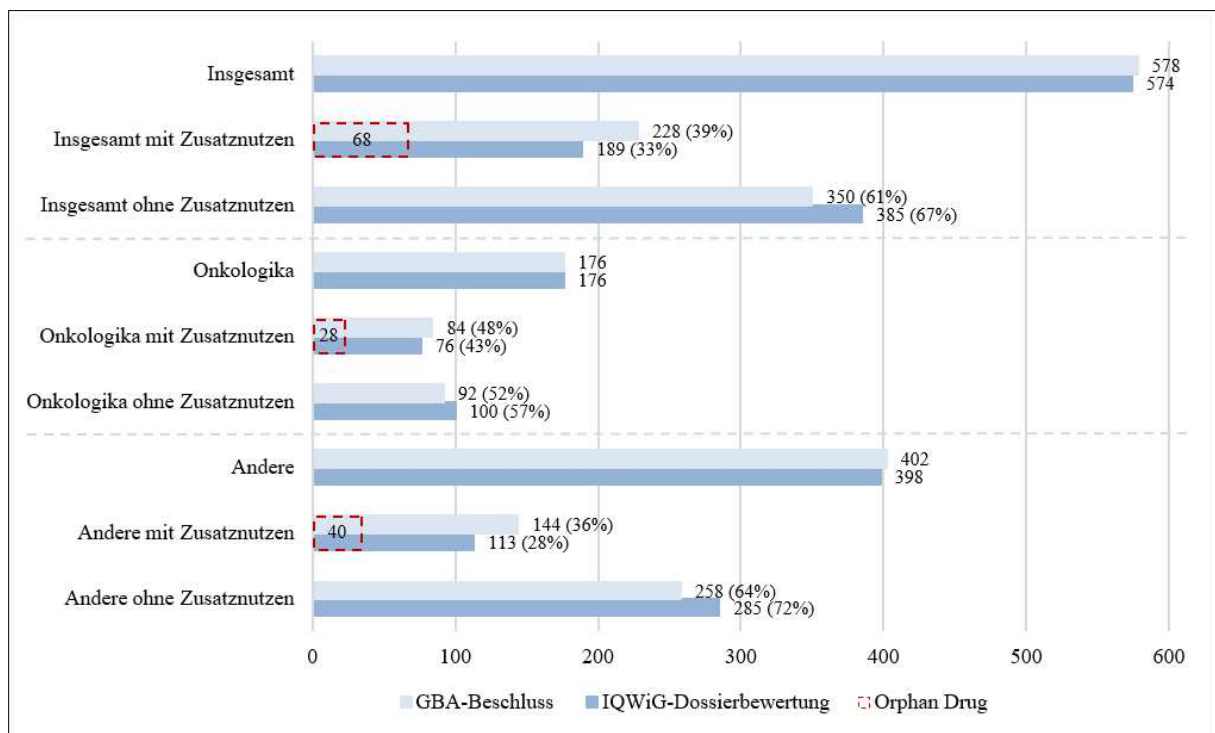
Beim direkten Vergleich der GBA-Beschlüsse und IQWiG-Empfehlungen wird deutlich, dass die offiziellen HTA-Resultate sowohl auf Teilpopulationsebene (87 %, 501/574) als auch auf Verfahrensebene (84 %, 218/259) mehrheitlich übereinstimmen (vgl. bei Tabellen 8–11). Bei den abweichenden HTA-Resultaten berichtete der GBA für 56/73 (77 %; auf Verfahrensebene: 27/41, 66 %) und das IQWiG für 17/73 (23 %; auf Verfahrensebene: 14/41, 34 %) der Wirkstoffe einen positiven Zusatznutzen. Die vorliegenden Einschätzungen zu den Outcomes werden durch die zuvor für GBA und IQWiG analysierten Evaluationsfaktoren bestätigt (vgl. Tabellen 12–15); die Berücksichtigung der klinischen Studienevidenz stimmt in 89 % (448/505) der bewerteten Subgruppen überein, das Kriterium der vergleichenden klinischen Effektivität hinsichtlich signifikanter Unterschiede in mindestens einem der patientenrelevanten Endpunkte gegenüber der ZVT in 87 % (437/505) der Bewertungsfälle. Die Regressionsergebnisse unterstreichen die Bedeutung der evidenzbasierten Kriterien – primär in Form der vergleichenden klinischen Wirksamkeit – sowohl für GBA als auch IQWiG, wobei der für die IQWiG-Bewertungen als signifikant herausgestellte Einfluss der klinischen Studienlage die vorliegenden Abweichungen in den FNB-Resultaten (zumindest teilweise) erklärt. Darüber hinaus sind die festgestellten Diskrepanzen auch darauf zurückzuführen, dass die in den pU-Dossiers berücksichtigten ZVTs in etwa fünf Prozent (28/574) der von GBA und IQWiG evaluierten Teilpopulationen voneinander abweichen und demnach nicht durchgängig akzeptiert wurden.

Ein deutlicher Unterschied zwischen den GBA-Beschlüssen und den IQWiG-Bewertungen besteht zudem im Hinblick auf die angegebenen Jahrestherapiekosten pro Patient (Teilpopulationsebene). Während sich die Kostenangaben in den GBA-Beschlusstexten im Durchschnitt zwischen 39.728 EUR und 669.162 EUR (Minimum und Maximum, ohne Zusatznutzen) respektive 98.565 EUR und 170.671 EUR (Minimum und Maximum, mit Zusatznutzen) bewegen, liegen sie in den IQWiG-Dossierbewertungen zwischen 41.088 EUR und 56.580 EUR (Minimum und Maximum, ohne Zusatznutzen) beziehungsweise 105.013 EUR und 176.051 EUR (Minimum und Maximum, mit Zusatznutzen). Die Abweichungen spiegeln sich – bei insgesamt vergleichbaren Zahlen zu den durchschnittlichen Patientenpopulationen<sup>57</sup> – in den einzelnen BI-Szenarien von GBA und IQWiG wider, weshalb die Schätzwerte im Mittel auch vergleichbare Tendenzen wie die Jahrestherapiekosten aufweisen.

---

<sup>57</sup> An dieser Stelle ist nochmals darauf hinzuweisen, dass bestehende Unterschiede in der Bildung von Subgruppen durch GBA und IQWiG, sowie insbesondere auch durch die pUs, für die vorliegenden Untersuchungen bereinigt, das heißt angelehnt an die in den GBA-Beschlusstexten berücksichtigten Teilpopulationen, verwendet wurden.

Damit lässt sich festhalten, dass sowohl in den GBA-Beschlussfassungen als auch IQWiG-Nutzenbewertungen die Anwendung der Prinzipien der EbM weitgehend konsistent befolgt werden, wobei IQWiG den eigenen Methoden unter Berücksichtigung der zur Verfügung gestellten Evidenz 'rigoroser' zu folgen scheint als der GBA (vgl. Tabellen 12–15). Den Ergebnissen der Regressionsanalysen zufolge weist der GBA auf Grundlage der Bewertungsempfehlungen durch IQWiG vergleichsweise positivere respektive weniger strikte Beschlüsse aus (vgl. Abbildung 14). Neben den zuvor bereits ausführlich dargelegten Evaluationskriterien weist darauf auch die erhöhte Akzeptanz des GBA gegenüber der Einreichung klinischer Evidenz auf Grundlage von Surrogat- anstelle patientenrelevanter Endpunkte hin. Exemplarisch lässt sich dies unter anderem an den Beschlusstexten von Crizotinib (FNB A15-59 / Behandlung des fortgeschrittenen, nicht-kleinzelligen Lungenkarzinoms) und Sipuleucel-T (FNB A14-38 / Behandlung des Prostatakarzinoms) aufzeigen, da bei den Bewertungen durch IQWiG die vom pU eingereichten Surrogatparameter nicht akzeptiert wurden und demzufolge eine negative Nutzenempfehlung vorlag. Weitere potenziell einflussnehmende Prozessmerkmale, darunter die Einbindung von Stakeholdern während der Anhörung durch den GBA sowie, wenn vorhanden, ergänzende Analysen durch IQWiG in Form von Addenda zu den durchgeführten Nutzenbewertungen, sind mittels der vorliegenden Datenbasis nicht nachweisbar gewesen.



**Abbildung 14:** GBA-Beschlüsse versus IQWiG-Bewertungen auf Teilpopulationsebene (Eigene Darstellung)

### 5.3 Analyse der STA-Kriterien und NICE Guidance

Sowohl die deskriptiven Untersuchungen als auch die Regressionsanalysen der NICE STA Guidance lassen sich anhand der zwei definierten Modelle unterteilen. Einerseits in eine Analyse unter nicht expliziter Berücksichtigung der (nachträglichen) Erstattungen durch den CDF, die als Basismodell verstanden werden kann und andererseits in eine ergänzende Untersuchung unter spezifischer Heranziehung des CDF als (zusätzliche) Outcome-Variable.<sup>58</sup>

#### 5.3.1 Deskriptive Untersuchungsergebnisse für NICE

Für den Untersuchungszeitraum konnten auf der NICE-Website 252 abgeschlossene TAs identifiziert werden. Darunter fallen 19 MTAs und 233 STAs, wobei von letzteren bereits 18 STA Guidance ersetzt und weitere acht zurückgezogen wurden. Demnach sind für NICE insgesamt 207 STA Guidance mit 267 einzelnen Empfehlungen berücksichtigt worden. Unter den berücksichtigten STAs befinden sich mit Aflibercept TA486 (Augenerkrankung/ Neovaskularisation), Golimumab TA497 (muskuloskelettale Erkrankung/ Arthritis) und Guselkumab TA521 (Hauterkrankung, Psoriasis) zudem drei FTAs, die vergleichbar den STA Guidance analysiert wurden. Mit Cabozantinib TA516 (Krebs/ medulläres Schilddrüsenkarzinom) wurde darüber hinaus auch ein Wirkstoff miteinbezogen, der – wie bei einem MTA – in Kombination (mit Vandetanib) bewertet wurde.

##### 5.3.1.1 Resultate auf Subgruppenebene

Von den 267 STA-Empfehlungen auf Subgruppenebene wurden 217 (81 %) durch NICE positiv bewertet, wobei 59/217 (22 %) eine uneingeschränkt und 158/217 (59 %) eine restriktiv positive Empfehlung erhielten (vgl. Tabelle 16).

---

<sup>58</sup> Erste Teilergebnisse der Analysen und Untersuchungen zu den NICE STAs wurden auf der 'Virtual ISPOR Europe 2020' vorgestellt (vgl. hierzu bei Schaefer et al., 2020).

**Tabelle 16:** Deskriptive Ergebnisse für NICE auf Subgruppenebene  
(Eigene Darstellung)

Variable	NICE Guidance			Gesamt (n)
	Keine Empfehlung (%)	Restriktive Empfehlung (%)	Empfehlung (%)	
<b>Vergleichstherapie</b>				
Nicht vorhanden	60,0	30,0	10,0	20
Vorhanden	15,4	61,5	23,1	247
<b>Klinische Studienevidenz: Randomisierte kontrollierte Studie (RCTs als systematischer Review/ Meta-Analyse)</b>				
Nicht vorhanden	42,9	42,9	14,3	14
Vorhanden	17,4	60,1	22,5	253
<b>Kosteneffektivitäts-Verhältnis (ICER/QALY gained): Einzelwert oder Wertebereich</b>				
< 20.000 GBP/QALY	1,3	69,7	28,9	76
20.000 - 30.000 GBP/QALY	3,8	73,1	23,1	78
> 30.000 GBP/QALY	43,6	40,6	15,8	101
<b>Klinische Wirksamkeit gegenüber Vergleichstherapie (basierend auf der vorliegenden Studienevidenz)</b>				
Schlechter	0,0	100,0	0,0	2
Vergleichbar	19,6	60,8	19,6	158
Besser	17,8	56,1	26,2	107
<b>End-of-Life-Kriterien (sofern berücksichtigt)</b>				
Nicht erfüllt	53,6	25,0	21,4	28
Erfüllt	33,9	45,8	20,3	59
<b>Veröffentlichungsjahr</b>				
2011	13,8	65,5	20,7	29
2012	5,9	58,8	35,3	17
2013	22,2	55,6	22,2	18
2014	5,3	42,1	52,6	19
2015	8,7	69,6	21,7	46
2016	26,1	50,0	23,9	46
2017	25,0	65,6	9,4	64
2018 (Januar-Juni)	28,6	50,0	21,4	28
<b>Indikation / Therapeutisches Gebiet</b>				
Atmungssystem	0,0	100,0	0,0	5
Augen	0,0	68,8	31,3	16
Blut und blutbildende Organe	14,3	71,4	14,3	7
Haut	0,0	100,0	0,0	8
Herz-Kreislauf	0,0	52,0	48,0	25
Infektionen	10,0	86,7	3,3	30
Muskel-Skelett-System	4,5	90,9	4,5	22
Nervensystem	0,0	50,0	50,0	6
Onkologie	36,0	38,6	25,4	114
Psychische Gesundheit	12,5	75,0	12,5	16
Stoffwechselsystem	0,0	50,0	50,0	2
Urogenital	0,0	100,0	0,0	1
Verdauungssystem	0,0	83,3	16,7	6
Sonstige	22,2	44,4	33,3	9
<b>Gesamt</b>	<b>18,7</b>	<b>59,2</b>	<b>22,1</b>	<b>267</b>

**Tabelle 16** (Fortsetzung):

Variable	NICE Guidance			Gesamt
	Keine Empfehlung	Restriktive Empfehlung	Empfehlung	
Jährliche Arzneimittelkosten pro Patient (GBP, in Tsd.)				
Mittelwert	43,3	26,0	28,8	29,8
Standardabweichung	28,4	27,8	35,8	30,5
Median	43,1	14,5	15,8	20,4
Geeignete (erstattungsfähige) Zielpopulation entsprechend der NICE Guidance (in Tsd.)				
Mittelwert	335,0	110,8	38,6	136,8
Standardabweichung	2.261,4	921,7	155,0	1.207,7
Median	0,7	1,7	1,3	1,4
NHS-Population: Gesamte theoretisch behandelbare / geeignete Patientenpopulation innerhalb des NHS (in Tsd.)				
Mittelwert	341,1	176,8	61,3	182,0
Standardabweichung	2.260,6	1.286,3	221,6	1.391,7
Median	1,9	4,1	3,0	2,9
Potenzieller Budget Impact: Zielpopulation (GBP, in Mio.)				
Mittelwert	332	125	62	149
Standardabweichung	2.000	612	152	978
Median	24	27	19	24
Potenzieller Budget Impact: NHS-Population (GBP, in Mio.)				
Mittelwert	453	287	120	280
Standardabweichung	2.060	1.230	186	1.300
Median	60	69	48	61
Inkrementeller Budget Impact: Zielpopulation (GBP, in Mio.)				
Mittelwert	170	3	1	33
Standardabweichung	997	188	113	457
Median	13	4	5	6
Inkrementeller Budget Impact: NHS-Population (GBP, in Mio.)				
Mittelwert	234	63	35	88
Standardabweichung	1.030	496	145	589
Median	30	11	12	13

ICER, Incremental Cost-Effectiveness Ratio; NHS, National Health Service; QALY, Quality-Adjusted Life Year.

Beim Vergleich der STAs nach Veröffentlichungsjahren ist zum einen das Jahr 2014 mit 19 Bewertungen hervorzuheben, da mehr als die Hälfte der Technologien (53 %) uneingeschränkt positiv empfohlen wurden. Demgegenüber erhielten nur etwas mehr als fünf Prozent der bewerteten Technologien keine Empfehlung für eine routinemäßige Anwendung innerhalb des NHS. Zum anderen ist das Jahr 2017 mit insgesamt 64 veröffentlichten STAs anzuführen, indem etwas mehr als 9 % der Technologien uneingeschränkt sowie 66 % restriktiv positiv bewertet wurden und demnach für 25 % der Bewertungen kein positives Empfehlungsergebnis durch NICE vorgelegen hatte.

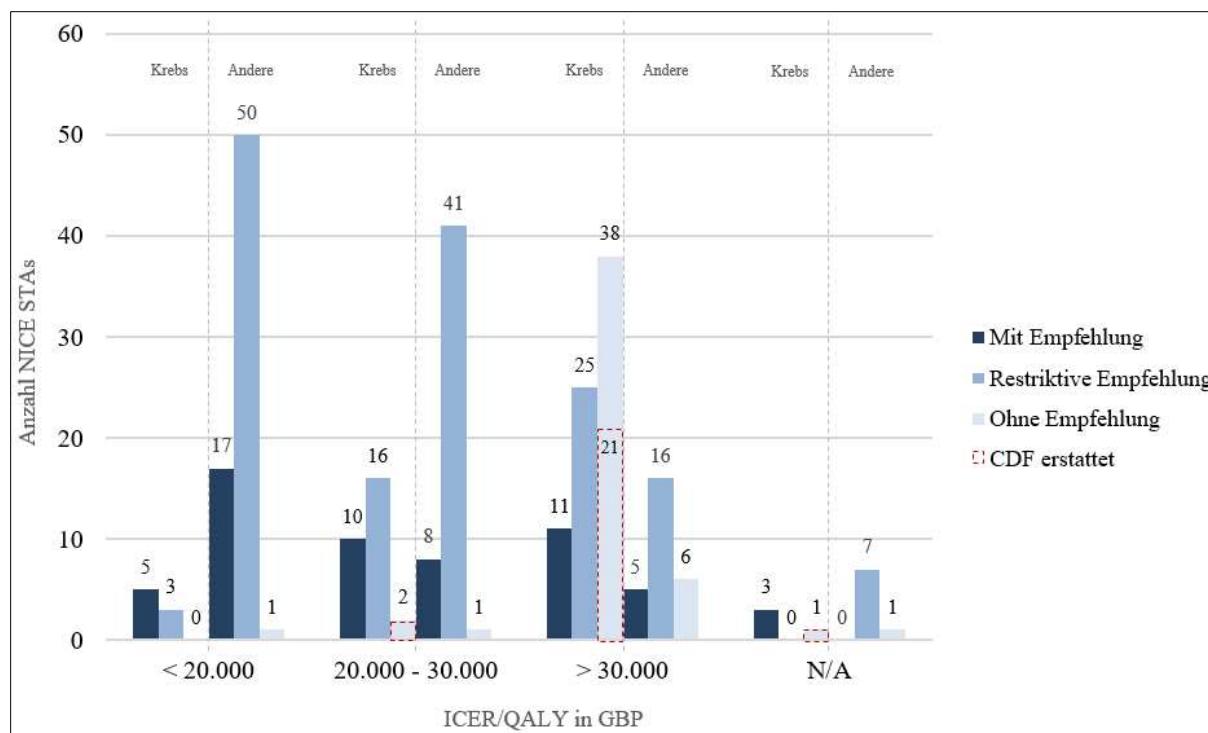
Unter Berücksichtigung der definierten Bewertungsfaktoren wird ersichtlich, dass bei Vorhandensein einer Vergleichstherapie (85 %) sowie adäquater klinischer Studienevidenz (83 %) der Anteil (eingeschränkt sowie uneingeschränkt) positiver Empfehlungen deutlich über den Evaluationen mit einer negativen Empfehlung liegt. Die von NICE angegebenen Bewertungszusammenfassungen zur klinischen Effektivität gegenüber den Vergleichstherapien unterstreichen die Einschätzungen für (eingeschränkt und uneingeschränkt) positive Empfehlungen (bei vergleichbarer Wirksamkeit, 80 %; bei besserer Wirksamkeit, 82 %), auch wenn eine komplett unabhängige Betrachtung vom Kosteneffektivitäts-Verhältnis nur bedingt Aussagekraft hätte.

Für den Bewertungsfaktor Kosteneffektivität lässt sich eindeutig erkennen, dass Technologien mit einem ICER < 20.000 GBP/QALY oder 20.000 – 30.000 GBP/QALY in der Regel eine positive (29 % oder 23 %) beziehungsweise eingeschränkt positive (70 % oder 73 %) Empfehlung aufweisen. Mit Tocilizumab TA238<sup>59</sup> (muskuloskelettale Erkrankung/ Arthritis) wurde lediglich eine Subgruppe einer Technologie mit einem ICER < 20.000 GBP/QALY negativ bewertet. Im Vergleich liegt bei einem ICER > 30.000 GBP/QALY mit fast 44 % ein wesentlich höherer Anteil an negativen Bewertungen durch NICE vor. Eine ausführliche Darstellung der NICE-Empfehlungen und ICER-Schwellenwerte befindet sich in Abbildung 15, welche sich nach onkologischen – unter ergänzender Berücksichtigung des CDF – und nicht-onkologischen STAs unterteilt.

Die Krebsbehandlungen stellen mit 43 % (114/267) das anteilig größte Indikationsgebiet dar. Insgesamt wurden 73/114 (64 %) Krebstechnologien positiv bewertet, worunter 44/73 (60 %) eingeschränkt positive Empfehlungen fallen. Demnach liegt für die Krebstechnologien mit etwa 36 % (41/114) der größte Anteil an negativen Bewertungen aller Indikationsgebiete vor. Die EoL-Kriterien, deren Verwendung im vorliegenden Untersuchungszeitraum ausschließlich für onkologische und hämatologische Interventionen identifiziert werden konnte, wurden in 87/114 (76 %) Bewertungen einbezogen. In 59/87 STAs wurden die EoL-Kriterien wiederum berücksichtigt, mit (eingeschränkt und uneingeschränkt) positiven Empfehlungen in knapp zwei Drittel der Fälle (66 %). Darüber hinaus wurden 24/41 (59 %) Krebstherapien, die nicht für eine routinemäßige Anwendung innerhalb des NHS empfohlen waren, durch den CDF erstattet (vgl. hierzu die ausführliche Darstellung in Kapitel 5.3.2.1).

---

<sup>59</sup> Die Bewertung von Tocilizumab (TA238, Behandlung der systemischen juvenilen idiopathischen Arthritis), die sich primär auf Kinder und Jugendliche bezieht, weist erhebliche Unsicherheiten in Bezug auf die Definition der Subgruppen und demnach auch für den plausibelsten ICER/QALY-Schätzwert auf, weshalb sich eine eindeutige Zuordnung innerhalb der beiden Subpopulationen – mit je einer positiven und einer negativen Empfehlung durch NICE – als äußerst schwierig darstellte.



**Abbildung 15:** STA-Empfehlungen zugeordnet nach ICER-Schwellenwerten auf Subgruppenebene (Eigene Darstellung)

CDF, Cancer Drugs Fund; GBP, Great British Pound; ICER, Incremental Cost-Effectiveness Ratio; QALY, Quality-Adjusted Life Year; STA, Single Technology Appraisal.

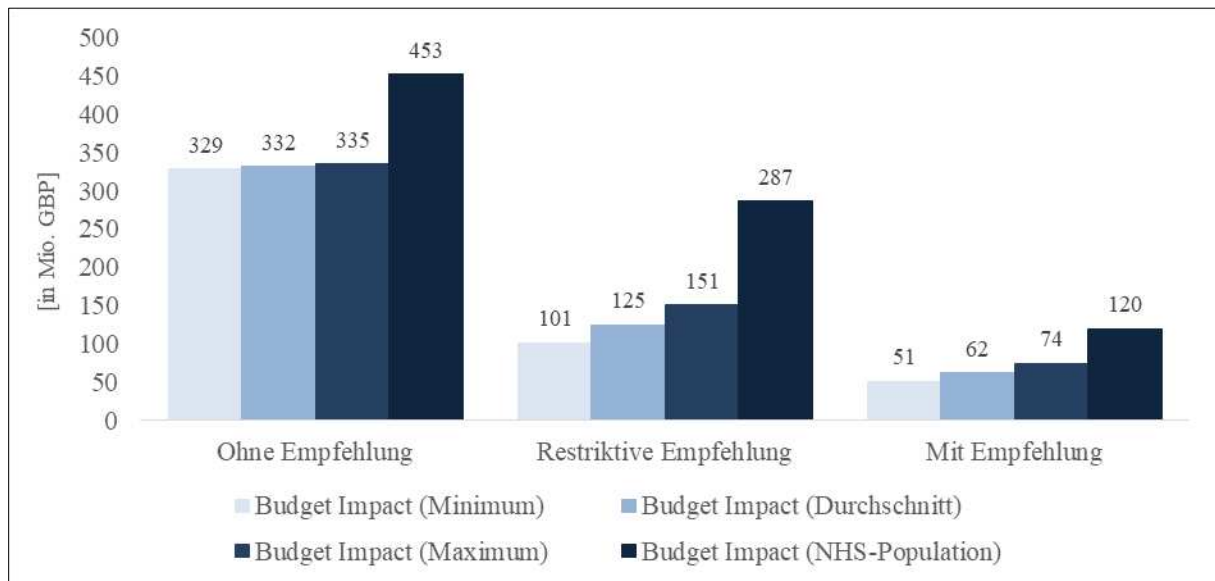
Im Gegensatz dazu wurden 144/153 (94 %) nicht-onkologische Interventionen positiv bewertet, darunter 114/144 (79 %) STAs positiv restriktiv. Die therapeutischen Gebiete mit den meisten Bewertungen stellen abnehmend die Technologien für Infektionskrankheiten (30 STAs), Herz-Kreislauf-Erkrankungen (25 STAs) sowie muskuloskelettale Erkrankungen (22 STAs) dar. Während zehn respektive knapp fünf Prozent der Wirkstoffe für Infektions- beziehungsweise Muskel-Skelett-Erkrankungen negativ bewertet wurden, liegen für Behandlungen kardiovaskulärer Erkrankungen keine negativen Empfehlungen durch NICE vor.

Die jährlichen Arzneimittelkosten (Mittelwert, 29.795 GBP) und die Größe der Patientenpopulation (Mittelwert<sup>60</sup>, 136.806) weisen im Durchschnitt höhere Angaben für Technologien ohne Empfehlung (Mittelwert, 43.306 GBP; 334.954 Patientinnen und Patienten) auf als die von NICE für den Einsatz im NHS eingeschränkt positiv (Mittelwert, 25.974 GBP; 110.763 Patien-

<sup>60</sup> Die Medianwerte für die durchschnittliche Größe der Patientenpopulation (Zielpopulation, 1.400 Personen) liegen unter den Mittelwerten. Zudem sind sie im Gegensatz zu den Mittelwerten für STAs ohne Empfehlung (700 Personen) niedriger als für STAs mit Empfehlung (restriktiv, 1.700 Personen; uneingeschränkt, 1.300 Personen).



tinnen und Patienten) oder uneingeschränkt positiv (Mittelwert, 28.809 GBP; 38.626 Patientinnen und Patienten) empfohlenen STAs. Vergleichbar den Resultaten für den GBA in Deutschland zeigen folglich auch die potenziellen sowie inkrementellen BI-Szenarien (Mittelwerte<sup>61</sup>) die eindeutige Tendenz, dass für STAs ohne Empfehlung höhere Schätzwerte als für STAs mit (eingeschränkter oder uneingeschränkter) positiver Empfehlung vorliegen (vgl. Abbildung 16 und Abbildung 17).



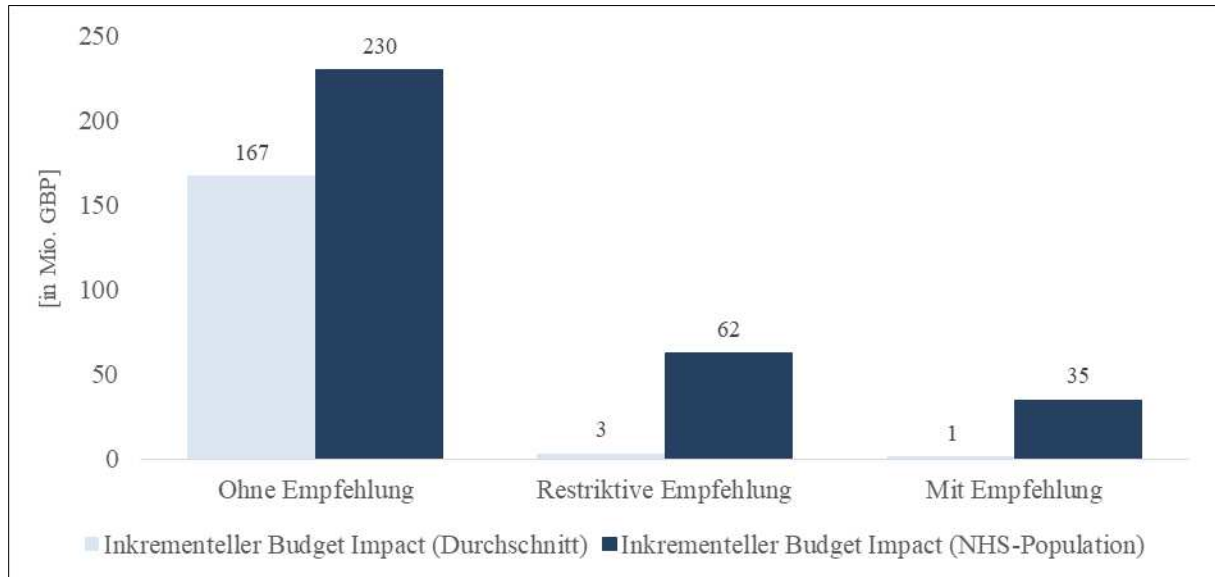
**Abbildung 16:** Budget Impact-Szenarien für NICE auf Subgruppenebene (Eigene Darstellung)

Der durchschnittliche BI für uneingeschränkt empfohlene STAs (potenziell, 62 Mio. GBP; inkrementell, 1 Mio. GBP) ist dabei niedriger als der für restriktiv positive Empfehlungen (potenziell, 125 Mio. GBP; inkrementell, 3 Mio. GBP) und der BI für STAs ohne Empfehlung ist im Durchschnitt wesentlich höher (potenziell, 332 Mio. GBP; inkrementell, 167 Mio. GBP). Sowohl die Minimum- als auch Maximum-Szenarien stützen dabei die Schätzwerte für den durchschnittlichen potenziellen BI. Darüber hinaus zeigen auch die potenziellen (Mittelwert<sup>62</sup>,

<sup>61</sup> Die durchschnittlichen Medianwerte liegen sowohl für den potenziellen (24 Mio. GBP) als auch den inkrementellen (6 Mio. GBP) BI unter den Mittelwerten. Im Gegensatz zu den Mittelwerten ist der potenzielle BI (Median) für die negativ bewerteten STAs (24 Mio. GBP) zudem niedriger als für die restriktiv positiv bewerteten STAs (27 Mio. GBP).

<sup>62</sup> Der Median für den potenziellen BI (61 Mio. GBP) liegt durchschnittlich unter dem Mittelwert. Zudem ist er im Gegensatz zum arithmetischen Mittel für negative Empfehlungen (60 Mio. GBP) niedriger als für positiv restriktiv bewertete STAs (69 Mio. GBP).

280 Mio. GBP) und inkrementellen (Mittelwert, 88 Mio. GBP) BI-Schätzungen für das NHS-Populations-Szenario eine vergleichbare Tendenz auf, wobei die Werte aufgrund der durchschnittlich größeren Patientenzahlen (Mittelwert<sup>63</sup>, 182.000) insgesamt höher sind.



**Abbildung 17:** Inkrementelle Budget Impact-Szenarien für NICE auf Subgruppenebene (Eigene Darstellung)

### 5.3.1.2 Resultate auf Indikationsebene

Die deskriptiven Resultate auf Subgruppenebene stützen die nachfolgenden Ergebnisdarstellungen auf Indikationsebene weitgehend. Genau genommen fallen die NICE STAs auf Indikationsebene prinzipiell vergleichbar aus (vgl. Tabelle 17), da die einzelnen Guidance für gewöhnlich in nur wenige Subgruppenentscheidungen mit entsprechend eigenen Bewertungsempfehlungen unterteilt werden. Darüber hinaus liegen sowohl für die Untersuchung auf Subpopulations- als auch auf Indikationsebene drei Outcome-Variablen vor, die (mögliche) Abweichungen in den Resultaten als sehr viel unwahrscheinlicher darstellen. Während – vergleichbar zur Analyse auf Subpopulationsebene – insgesamt 80 % (164/207) der NICE STAs positiv bewertet wurden, fallen etwas mehr als die Hälfte der STAs (105/207, 51 %) auf Indikationsebene unter die positiv restriktiven Bewertungsempfehlungen.

<sup>63</sup> Die Medianwerte für die Durchschnittsgröße der Patientenpopulation (NHS-Population, 2.900 Personen) liegen unter den Mittelwerten. Zudem sind sie im Gegensatz zu den Mittelwerten für STAs ohne Empfehlung (1.900 Personen) niedriger als für STAs mit Empfehlung (restriktiv, 4.100 Personen; uneingeschränkt, 3.000 Personen).

**Tabelle 17:** Deskriptive Ergebnisse für NICE auf Indikationsebene  
(Eigene Darstellung)

Variable	NICE Guidance			Gesamt (n)
	Keine Empfehlung (%)	Restriktive Empfehlung (%)	Empfehlung (%)	
<b>Vergleichstherapie</b>				
Nicht vorhanden	61,1	27,8	11,1	18
Vorhanden	16,9	52,9	30,2	189
<b>Klinische Studienevidenz: Randomisierte kontrollierte Studie (RCTs als systematischer Review/ Meta-Analyse)</b>				
Nicht vorhanden	45,5	36,4	18,2	11
Vorhanden	19,4	51,5	29,1	196
<b>Kosteneffektivitäts-Verhältnis (ICER/QALY gained): Einzelwert oder Wertebereich</b>				
< 20.000 GBP/QALY	0,0	52,2	47,8	46
20.000 - 30.000 GBP/QALY	4,5	68,2	27,3	66
> 30.000 GBP/QALY	45,3	36,0	18,6	86
<b>Klinische Wirksamkeit gegenüber Vergleichstherapie (basierend auf der vorliegenden Studienevidenz)</b>				
Schlechter	0,0	100,0	0,0	2
Vergleichbar	21,5	52,9	25,6	121
Besser	20,2	46,4	33,3	84
<b>End-of-Life-Kriterien (sofern berücksichtigt)</b>				
Nicht erfüllt	55,6	22,2	22,2	27
Erfüllt	33,3	44,4	22,2	54
<b>Veröffentlichungsjahr</b>				
2011	16,7	50,0	33,3	18
2012	7,1	50,0	42,9	14
2013	21,4	50,0	28,6	14
2014	6,3	31,3	62,5	16
2015	3,6	60,7	35,7	28
2016	27,3	47,7	25,0	44
2017	32,6	54,3	13,0	46
2018 (Januar-Juni)	25,9	51,9	22,2	27
<b>Indikation / Therapeutisches Gebiet</b>				
Atmungssystem	0,0	100,0	0,0	5
Augen	0,0	58,3	41,7	12
Blut und blutbildende Organe	16,7	66,7	16,7	6
Haut	0,0	100,0	0,0	8
Herz-Kreislauf	0,0	40,0	60,0	20
Infektionen	0,0	83,3	16,7	6
Muskel-Skelett-System	0,0	92,3	7,7	13
Nervensystem	0,0	50,0	50,0	6
Onkologie	36,4	36,4	27,1	107
Psychische Gesundheit	0,0	50,0	50,0	2
Stoffwechselsystem	12,5	62,5	25,0	8
Urogenital	0,0	100,0	0,0	1
Verdauungssystem	0,0	83,3	16,7	6
Sonstige	28,6	28,6	42,9	7
<b>Gesamt</b>	<b>20,8</b>	<b>50,7</b>	<b>28,5</b>	<b>207</b>

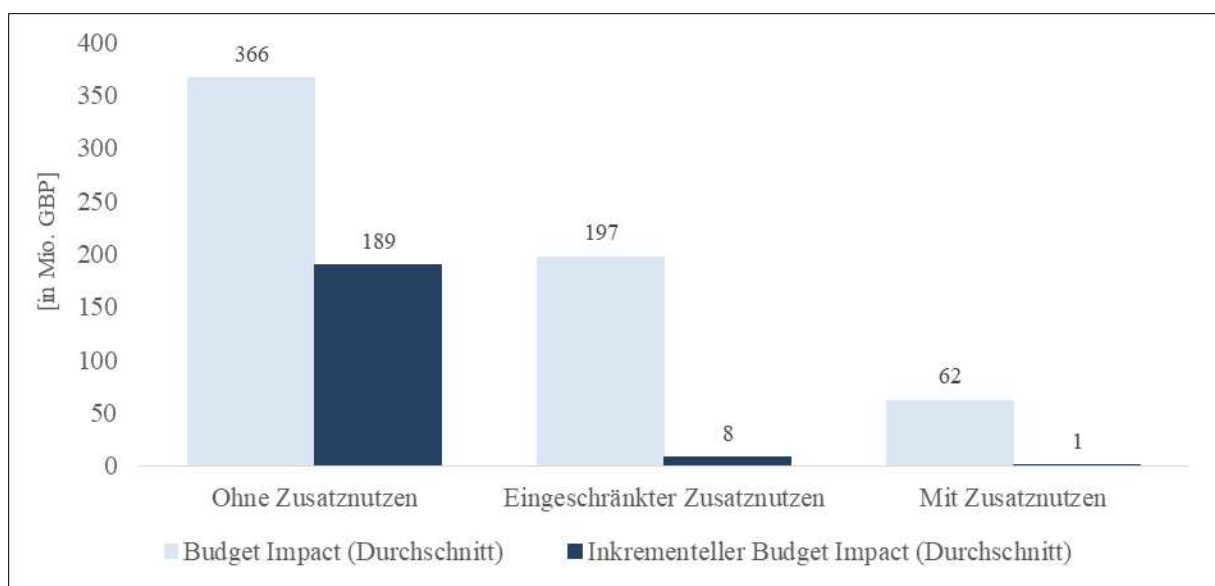
**Tabelle 17** (Fortsetzung):

Variable	NICE Guidance			Gesamt
	Keine Empfehlung	Restriktive Empfehlung	Empfehlung	
Jährliche Arzneimittelkosten pro Patient (GBP, in Tsd.)				
Mittelwert	42,4	24,6	28,8	29,4
Standardabweichung	28,4	26,4	35,8	30,4
Median	38,4	15,5	15,8	19,8
Geeignete (erstattungsfähige) Zielpopulation entsprechend der NICE Guidance (in Tsd.)				
Mittelwert	372,8	173,5	38,6	176,5
Standardabweichung	2.438,3	1.148,1	155,0	1.377,6
Median	0,7	1,6	1,3	1,2
Potenzieller Budget Impact: Zielpopulation (GBP, in Mio.)				
Mittelwert	366	197	62	193
Standardabweichung	2.160	779	152	1.120
Median	23	25	19	23
Inkrementeller Budget Impact: Zielpopulation (GBP, in Mio.)				
Mittelwert	189	8	1	43
Standardabweichung	1.080	264	113	527
Median	13	5	5	7

ICER, Incremental Cost-Effectiveness Ratio; NHS, National Health Service; QALY, Quality-Adjusted Life Year.

Erkennbare Unterschiede zu den Resultaten auf Subpopulationsebene lassen sich primär für die Interventionen zur Behandlung von Infektionserkrankungen (6/207, 3 %) und muskuloskelettaler Krankheitsbilder (13/207, 6 %) erkennen, da für beide therapeutischen Gebiete auf Indikationsebene keine negativen Empfehlungen durch NICE vorliegen. Demgegenüber sind die Bewertungsempfehlungen für Krebstherapien (107/207, 52 %) auf Indikations- und Subgruppenebene nahezu identisch. Für das Veröffentlichungsjahr 2015 liegen wiederum weniger negative Empfehlungen auf Indikationsebene vor (4 %), wohingegen die Resultate für das Jahr 2017 mit fast einem Drittel (33 %) negativer Empfehlungen durch NICE im Vergleich zur Subgruppenanalyse ein genau umgekehrtes Bild aufzeigen. Bei der Kosteneffektivität sind die STA-Empfehlungen vergleichbar zu den Resultaten der Subgruppenanalyse, wobei für den ICER-Schwellenwert < 20.000 GBP/QALY keine negative Empfehlung mehr vorliegt.

Im Wesentlichen entsprechen sowohl die jährlichen Arzneimittelkosten pro Patient (Mittelwert, 29.400 GBP) als auch die Patientenzahlen (Mittelwert<sup>64</sup>; Zielpopulation, 176.500) im Durchschnitt den Werten der Untersuchung auf Subgruppenebene. Dasselbe gilt entsprechend für die potenziellen sowie inkrementellen BI-Szenarien der Zielpopulation (Mittelwert, 192 Mio. GBP und 43 Mio. GBP), die auf Indikationsebene durchschnittlich höher sind als für die Subgruppen. Demzufolge weisen STAs mit negativen Bewertungen auf Indikationsebene durchschnittlich höhere (potenzielle und inkrementelle) BI-Schätzungen (Mittelwerte<sup>65</sup>) auf als die restriktiv und uneingeschränkt positiven Empfehlungen durch NICE (vgl. Abbildung 18).



**Abbildung 18:** [Inkrementelle] Budget Impact-Szenarien für NICE auf Indikationsebene (Eigene Darstellung)

<sup>64</sup> Die Medianwerte für die durchschnittliche Größe der Patientenpopulation (1.200 Personen) liegen unter den Mittelwerten. Zudem sind sie im Gegensatz zu den Mittelwerten für STAs ohne Empfehlung (700 Personen) niedriger als für STAs mit Empfehlung (restriktiv, 1.600 Personen; uneingeschränkt, 1.300 Personen).

<sup>65</sup> Der Median für den potenziellen BI (23 Mio. GBP) liegt durchschnittlich unter dem arithmetischen Mittel. Zudem ist er im Gegensatz zum Mittelwert für negative Empfehlungen (23 Mio. GBP) etwas niedriger als für positiv restriktiv bewertete STAs (25 Mio. GBP).

### **5.3.2 Deskriptive Untersuchungsergebnisse für NICE unter Berücksichtigung des Cancer Drugs Fund (CDF)**

Unter Heranziehung des CDF (als eine zusätzliche Outcome-Variable mit Empfehlung durch NICE), der seit Juli 2016 die abgeschlossenen Guidance von Gesundheitstechnologien zur Behandlung onkologischer und hämatologischer Erkrankungen nachträglich beeinflussen kann, zeigen die deskriptiven Untersuchungsergebnisse auf Subgruppen- sowie Indikationsebene teils deutliche Unterschiede für die berücksichtigten Krebstherapien auf.

#### *5.3.2.1 Resultate auf Subgruppenebene – inklusive CDF*

Bei ergänzender Berücksichtigung des CDF wird offensichtlich, dass sich der Anteil negativer STAs (26/267, 10 %; ohne CDF, 19 %) zugunsten der uneingeschränkt positiven Empfehlungen (83/267, 31 %; ohne CDF, 22 %) durch NICE verschiebt (vgl. Tabelle 18). Zum Beispiel ist für die NICE STAs nach Veröffentlichungsjahren sehr deutlich erkennbar, dass die Empfehlungen von 2016 bis 2018 im Durchschnitt positiver ausfallen als ohne Berücksichtigung des CDF. Für das Jahr 2016 wurden unter CDF-Berücksichtigung 87 % (ohne CDF, 74 %), für 2017 95 % (ohne CDF, 75 %) und für 2018 89 % (ohne CDF, 71 %) der bewerteten Technologien positiv bewertet.

Bei einem gesonderten Fokus auf die evidenzbasierten Bewertungskriterien wird erkennbar, dass die NICE STAs bei Nicht-Vorhandensein einer Vergleichstherapie (90 %; ohne CDF, 40 %) sowie Fehlen adäquater klinischer Studienevidenz (93 %; ohne CDF, 57 %) insgesamt positiver ausfallen. Für die Bewertung der klinischen Effektivität lässt sich wiederum festhalten, dass der Anteil positiver Empfehlungen bei vergleichbarer (86 %; ohne CDF, 80 %) oder besserer (96 %; ohne CDF, 82 %) klinischer Wirksamkeit gegenüber der vorliegenden Standard- oder Vergleichstherapie zunimmt. Bei Betrachtung des Kosteneffektivitäts-Kriteriums wird offensichtlich, dass sich insbesondere der Anteil positiver Empfehlungen für STAs mit einem ICER > 30.000 GBP/QALY deutlich erhöht (77 %; ohne CDF, 56 %).

Die Bewertungen für Krebstherapien zeigen darüber hinaus einen deutlichen Anstieg der uneingeschränkt positiven Empfehlungen durch NICE (47 %; ohne CDF, 25 %). Dies scheint sich auch mit Blick auf die Kriterien für EoL-Behandlungen zu bestätigen; sowohl bei Ablehnung (64 %; ohne CDF, 46 %) als auch bei Akzeptanz (95 %; ohne CDF, 66 %) der EoL-Kriterien erhöht sich der Anteil uneingeschränkt positiver Bewertungen.

**Tabelle 18:** Deskriptive Ergebnisse für NICE auf Subgruppenebene – inklusive CDF  
(Eigene Darstellung)

Variable	NICE Guidance unter Berücksichtigung des Cancer Drugs Fund (CDF)			Gesamt (n)
	Keine Empfehlung (%)	Restriktive Empfehlung (%)	Empfehlung (%)	
<b>Vergleichstherapie</b>				
Nicht vorhanden	10,0	30,0	60,0	20
Vorhanden	9,7	61,5	28,7	247
<b>Klinische Studienevidenz: Randomisierte kontrollierte Studie (RCTs als systematischer Review/ Meta-Analyse)</b>				
Nicht vorhanden	7,1	42,9	50,0	14
Vorhanden	9,9	60,1	30,0	253
<b>Kosteneffektivitäts-Verhältnis (ICER/QALY gained): Einzelwert oder Wertebereich</b>				
< 20.000 GBP/QALY	1,3	69,7	28,9	76
20.000 - 30.000 GBP/QALY	1,3	73,1	25,6	78
> 30.000 GBP/QALY	22,8	40,6	36,6	101
<b>Klinische Wirksamkeit gegenüber Vergleichstherapie (basierend auf der vorliegenden Studienevidenz)</b>				
Schlechter	0,0	100,0	0,0	2
Vergleichbar	13,9	60,8	25,3	158
Besser	3,7	56,1	40,2	107
<b>End-of-Life-Kriterien (sofern berücksichtigt)</b>				
Nicht erfüllt	35,7	25,0	39,3	28
Erfüllt	5,1	45,8	49,2	59
<b>Veröffentlichungsjahr</b>				
2011	13,8	65,5	20,7	29
2012	5,9	58,8	35,3	17
2013	22,2	55,6	22,2	18
2014	5,3	42,1	52,6	19
2015	8,7	69,6	21,7	46
2016	13,0	50,0	37,0	46
2017	4,7	65,6	29,7	64
2018 (Januar-Juni)	10,7	50,0	39,3	28
<b>Indikation / Therapeutisches Gebiet</b>				
Atmungssystem	0,0	100,0	0,0	5
Augen	0,0	68,8	31,3	16
Blut und blutbildende Organe	14,3	71,4	14,3	7
Haut	0,0	100,0	0,0	8
Herz-Kreislauf	0,0	52,0	48,0	25
Infektionen	10,0	86,7	3,3	30
Muskel-Skelett-System	4,5	90,9	4,5	22
Nervensystem	0,0	50,0	50,0	6
Onkologie	14,9	38,6	46,5	114
Psychische Gesundheit	12,5	75,0	12,5	16
Stoffwechselsystem	0,0	50,0	50,0	2
Urogenital	0,0	100,0	0,0	1
Verdauungssystem	0,0	83,3	16,7	6
Sonstige	22,2	44,4	33,3	9
<b>Gesamt</b>	<b>9,7</b>	<b>59,2</b>	<b>31,1</b>	<b>267</b>

**Tabelle 18** (Fortsetzung):

Variable	NICE Guidance unter Berücksichtigung des Cancer Drugs Fund (CDF)			Gesamt
	Keine Empfehlung	Restriktive Empfehlung	Empfehlung	
Jährliche Arzneimittelkosten pro Patient (GBP, in Tsd.)				
Mittelwert	31,5	26,0	36,6	29,8
Standardabweichung	28,1	27,8	34,9	30,5
Median	28,7	14,5	30,8	20,4
Geeignete (erstattungsfähige) Zielpopulation entsprechend der NICE Guidance (in Tsd.)				
Mittelwert	643,7	110,8	27,6	136,8
Standardabweichung	3.133,1	921,7	131,5	1.207,7
Median	1,6	1,7	0,7	1,4
NHS-Population: Gesamte theoretisch behandelbare / geeignete Patientenpopulation innerhalb des NHS (in Tsd.)				
Mittelwert	655,1	176,8	43,8	182,0
Standardabweichung	3.130,9	1.286,3	188,4	1.391,7
Median	2,8	4,1	1,8	2,9
Potenzieller Budget Impact: Zielpopulation (GBP, in Mio.)				
Mittelwert	601	125	52	149
Standardabweichung	2.740	612	131	978
Median	31	27	19	24
Potenzieller Budget Impact: NHS-Population (GBP, in Mio.)				
Mittelwert	815	287	98	280
Standardabweichung	2.800	1.230	163	1.300
Median	102	69	41	61
Inkrementeller Budget Impact: Zielpopulation (GBP, in Mio.)				
Mittelwert	304	3	6	33
Standardabweichung	1.370	188	97	457
Median	11	4	10	6
Inkrementeller Budget Impact: NHS-Population (GBP, in Mio.)				
Mittelwert	414	63	34	88
Standardabweichung	1.400	496	125	589
Median	54	11	16	13

ICER, Incremental Cost-Effectiveness Ratio; NHS, National Health Service; QALY, Quality-Adjusted Life Year.

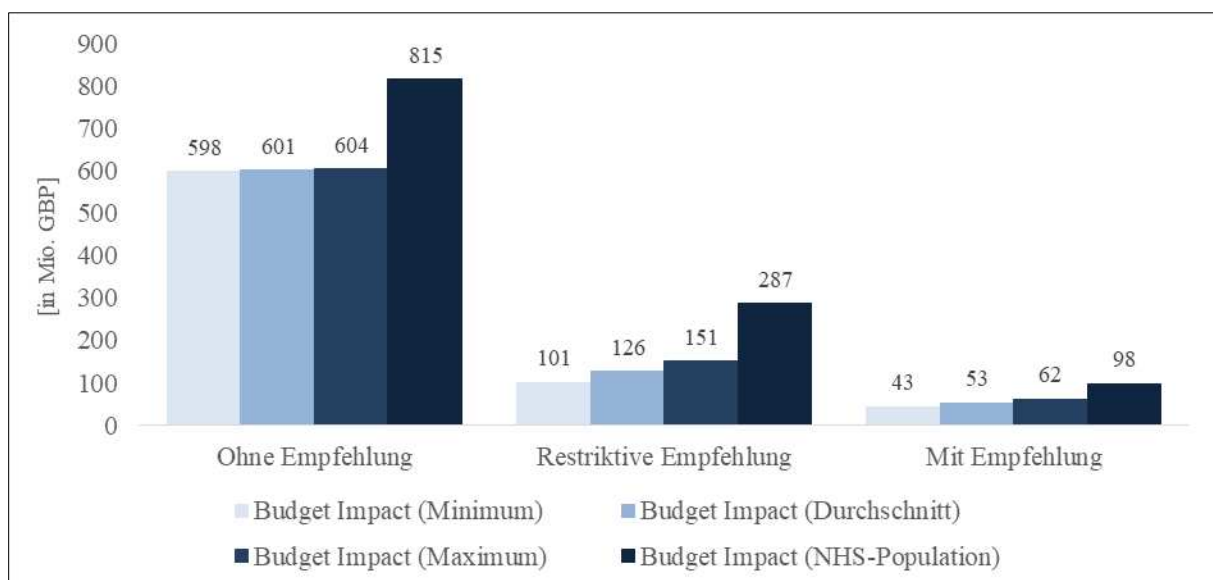
Offensichtlich liegt die Größe der durchschnittlichen Populationszahl (Mittelwert<sup>66</sup>) bei negativer Bewertung durch NICE mit 643.700 Patientinnen und Patienten deutlich über der Anzahl ohne Berücksichtigung des CDF (334.953), während uneingeschränkt positive Empfehlungen

<sup>66</sup> Die Medianwerte für die Größe der Patientenpopulation (Zielpopulation, 1.400 Personen) liegen im Durchschnitt unter den Mittelwerten. Zudem sind sie im Gegensatz zu den Mittelwerten für STAs ohne Empfehlung (1.600 Personen) niedriger als für STAs mit restriktiv positiver Empfehlung (1.700 Personen). Vergleichbar zeigen sich die Medianwerte für die NHS-Population; sie sind durchschnittlich (2.900 Personen) niedriger als die Mittelwerte und sind für STAs ohne Empfehlung (2.800 Personen) ebenfalls niedriger als für STAs mit positiv restriktiver Bewertung (4.100 Personen).

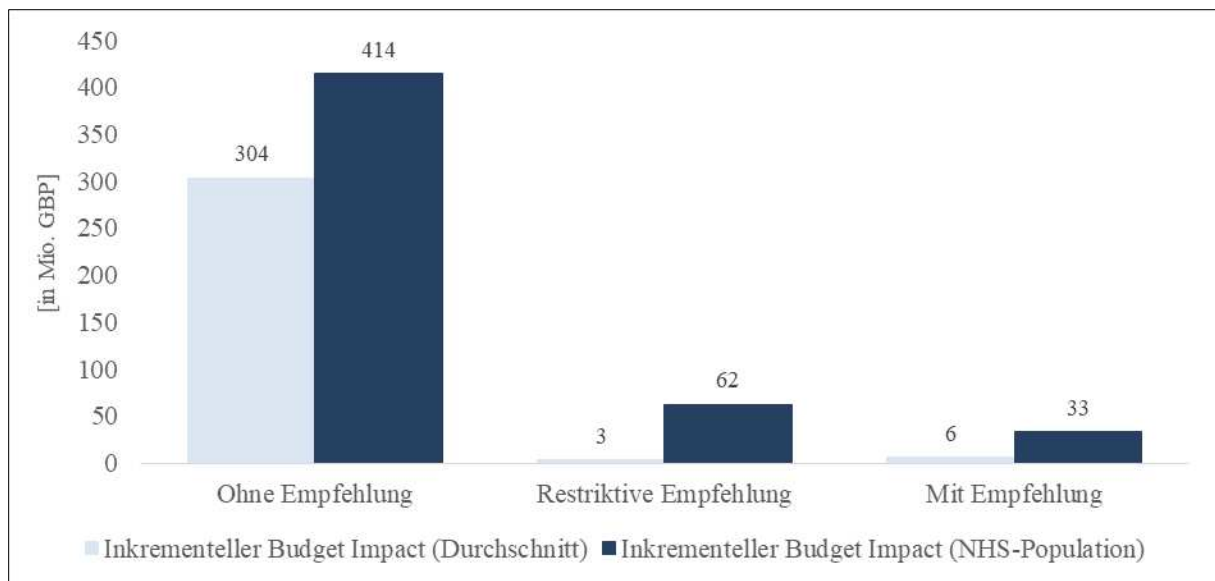


mit 27.595 Patientinnen und Patienten im Vergleich niedriger sind (ohne CDF, 38.626). Demgegenüber zeigen die durchschnittlichen Kosten pro Patient (Mittelwert) für uneingeschränkt positive Empfehlungen eine Zunahme (36.634 GBP; ohne CDF, 28.809 GBP) und für negative Empfehlungen eine Abnahme (31.451 GBP; ohne CDF, 43.306 GBP) gegenüber den Angaben ohne CDF-Berücksichtigung. Demzufolge wird auch ersichtlich, dass mit Berücksichtigung der CDF-Variable die jährlichen Arzneimittelkosten pro Patient (Mittelwert) für uneingeschränkt positiv bewertete Technologien durchschnittlich am höchsten sind.

Für die unterschiedlichen BI-Szenarien zeichnet sich das folgende Bild ab (vgl. Abbildung 19 und Abbildung 20); während der potenzielle BI (Mittelwert) bei negativen Bewertungen durch NICE mit durchschnittlich 601 Mio. GBP (ohne CDF, 332 Mio. GBP) höher einzuschätzen ist, liegt er bei uneingeschränkt positiven Empfehlungen mit 53 Mio. GBP (ohne CDF, 62 Mio. GBP) im Durchschnitt leicht darunter. Die Schätzungen für den inkrementellen BI (Mittelwert) sind im Durchschnitt sowohl für uneingeschränkt positive (6 Mio. GBP; ohne CDF, 1 Mio. GBP) als auch für negative (304 Mio. GBP; ohne CDF, 170 Mio. GBP) Bewertungen höher. Die Szenarien für die NHS-Population zeigen ebenfalls eine deutliche Zunahme (beispielhaft für den inkrementellen BI, Mittelwert: 414 Mio. GBP gegenüber ohne CDF, 234 Mio. GBP) bei einer negativen Empfehlung durch NICE, jedoch vergleichbare Schätzwerte sowohl für uneingeschränkt als auch restriktiv positive Bewertungen.



**Abbildung 19:** Budget Impact-Szenarien für NICE auf Subgruppenebene – inklusive CDF (Eigene Darstellung)



**Abbildung 20:** Inkrementelle Budget Impact-Szenarien für NICE auf Subgruppenebene – inklusive CDF (Eigene Darstellung)

### 5.3.2.2 Resultate auf Indikationsebene – inklusive CDF

Vergleichbar den Untersuchungen auf Subgruppenebene zeigen die Resultate auf Indikationsebene, dass für einzelne Bewertungsfaktoren durch die Berücksichtigung des CDF teilweise deutliche Unterschiede bestehen (vgl. Tabelle 19). Insgesamt liegen dabei mit etwa 90 % (186/207) der STA Guidance mehr positive Bewertungsergebnisse vor als ohne den CDF als zusätzlich berücksichtigte Outcome-Variable. Zurückzuführen sind die Veränderungen auf das Indikationsgebiet der onkologischen Erkrankungen, deren Bewertungen mit insgesamt 90/107 (84 %; ohne CDF, 64 %) positiv bewerteten STAs wesentlich besser ausfallen, insbesondere bei Erfüllung der EoL-Kriterien (in insgesamt 51/54 Krebstherapiebewertungen).

Bei Betrachtung der Veröffentlichungsjahre 2016 (86 %; ohne CDF, 73 %), 2017 (94 %; ohne CDF, 67 %) und 2018 (89 %; ohne CDF, 74 %) wird ebenfalls deutlich erkennbar, dass der Anteil der positiven Empfehlungen zunimmt. Entsprechend den Ergebnissen für die Subgruppen zeigen auch die evidenzbasierten Bewertungsfaktoren eine starke Tendenz dahingehend, dass der Anteil (uneingeschränkt und restriktiv) positiver Empfehlungen zunimmt, wenn keine Vergleichstherapie (89 %; ohne CDF, 39 %) oder klinische Studienevidenz in Form von RCTs (91 %; ohne CDF, 55 %) vorliegt. Für das Kosteneffektivitäts-Kriterium lässt sich dabei ebenfalls feststellen, dass insbesondere STAs mit einem ICER > 30.000 GBP/QALY positiver bewertet werden (77 %; ohne CDF, 55 %).

**Tabelle 19:** Deskriptive Ergebnisse für NICE auf Indikationsebene – inklusive CDF  
(Eigene Darstellung)

Variable	NICE Guidance unter Berücksichtigung des Cancer Drugs Fund (CDF)			Gesamt (n)
	Keine Empfehlung (%)	Restriktive Empfehlung (%)	Empfehlung (%)	
<b>Vergleichstherapie</b>				
Nicht vorhanden	11,1	22,2	66,7	18
Vorhanden	10,1	52,9	37,0	189
<b>Klinische Studienevidenz: Randomisierte kontrollierte Studie (RCTs als systematischer Review/ Meta-Analyse)</b>				
Nicht vorhanden	9,1	27,3	63,6	11
Vorhanden	10,2	51,5	38,3	196
<b>Kosteneffektivitäts-Verhältnis (ICER/QALY gained): Einzelwert oder Wertebereich</b>				
< 20.000 GBP/QALY	0,0	52,2	47,8	46
20.000 - 30.000 GBP/QALY	1,5	68,2	30,3	66
> 30.000 GBP/QALY	23,3	34,9	41,9	86
<b>Klinische Wirksamkeit gegenüber Vergleichstherapie (basierend auf der vorliegenden Studienevidenz)</b>				
Schlechter	0,0	100,0	0,0	2
Vergleichbar	14,9	52,1	33,1	121
Besser	3,6	46,4	50,0	84
<b>End-of-Life-Kriterien (sofern berücksichtigt)</b>				
Nicht erfüllt	37,0	22,2	40,7	27
Erfüllt	5,6	42,6	51,9	54
<b>Veröffentlichungsjahr</b>				
2011	16,7	50,0	33,3	18
2012	7,1	50,0	42,9	14
2013	21,4	50,0	28,6	14
2014	6,3	31,3	62,5	16
2015	3,6	60,7	35,7	28
2016	13,6	47,7	38,6	44
2017	6,5	54,3	39,1	46
2018 (Januar-Juni)	11,1	48,1	40,7	27
<b>Indikation / Therapeutisches Gebiet</b>				
Atmungssystem	0,0	100,0	0,0	5
Augen	0,0	58,3	41,7	12
Blut und blutbildende Organe	16,7	66,7	16,7	6
Haut	0,0	100,0	0,0	8
Herz-Kreislauf	0,0	40,0	60,0	20
Infektionen	0,0	83,3	16,7	6
Muskel-Skelett-System	0,0	92,3	7,7	13
Nervensystem	0,0	50,0	50,0	6
Onkologie	15,9	35,5	48,6	107
Psychische Gesundheit	0,0	50,0	50,0	2
Stoffwechselsystem	12,5	62,5	25,0	8
Urogenital	0,0	100,0	0,0	1
Verdauungssystem	0,0	83,3	16,7	6
Sonstige	28,6	28,6	42,9	7
<b>Gesamt</b>	<b>10,1</b>	<b>50,2</b>	<b>39,6</b>	<b>207</b>

**Tabelle 19** (Fortsetzung):

Variable	NICE Guidance unter Berücksichtigung des Cancer Drugs Fund (CDF)			Gesamt
	Keine Empfehlung	Restriktive Empfehlung	Empfehlung	
Jährliche Arzneimittelkosten pro Patient (GBP, in Tsd.)				
Mittelwert	28,5	24,2	36,4	29,4
Standardabweichung	26,1	26,2	35,0	30,4
Median	28,5	15,4	29,5	19,8
Geeignete (erstattungsfähige) Zielpopulation entsprechend der NICE Guidance (in Tsd.)				
Mittelwert	762,9	175,2	27,9	176,5
Standardabweichung	3.489,0	1.153,6	132,3	1.377,6
Median	1,2	1,6	0,7	1,2
Potenzieller Budget Impact: Zielpopulation (GBP, in Mio.)				
Mittelwert	702	199	53	193
Standardabweichung	3.050	783	131	1.120
Median	24	27	19	23
Inkrementeller Budget Impact: Zielpopulation (GBP, in Mio.)				
Mittelwert	358	8	6	43
Standardabweichung	1.520	265	98	527
Median	10	5	11	7

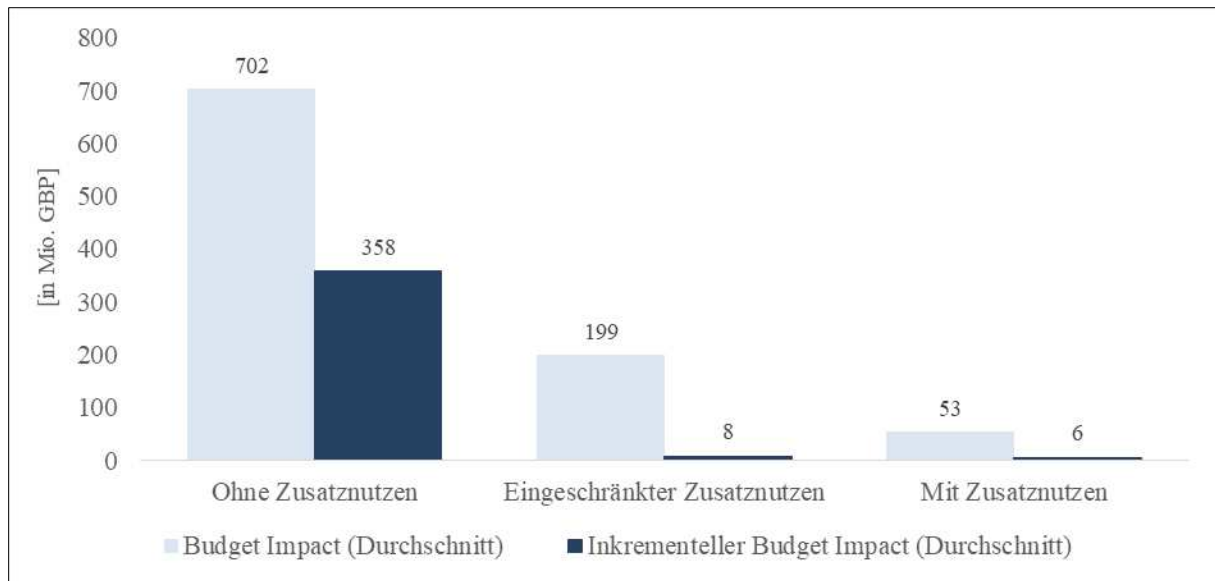
ICER, Incremental Cost-Effectiveness Ratio; NHS, National Health Service; QALY, Quality-Adjusted Life Year.

Während sich für die jährlichen durchschnittlichen Arzneimittelkosten pro Patient ein leichter Anstieg bei den positiven Empfehlungen (Mittelwert, 35.600 GBP) zeigt, erhöht sich die durchschnittliche Anzahl der Patientinnen und Patienten bei negativen Empfehlungen (Mittelwert<sup>67</sup>; Zielpopulation, 800.800) durch NICE erheblich. Sowohl für die potenziellen (Mittelwert<sup>68</sup>; Zielpopulation, 358 Mio. GBP) als auch die inkrementellen (Mittelwert<sup>69</sup>; Zielpopulation, 185 Mio. GBP) BI-Szenarien wird ersichtlich, dass die Werte bei einer Nicht-Empfehlung durch NICE im Durchschnitt wesentlich höher liegen als wenn der CDF unberücksichtigt bleibt beziehungsweise nicht erweiternd herangezogen wird (vgl. Abbildung 21).

<sup>67</sup> Die Medianwerte für die Größe der Patientenpopulation (1.200 Personen) liegen im Durchschnitt unter den Mittelwerten. Zudem sind sie im Gegensatz zu den Mittelwerten für STAs ohne Empfehlung (1.200 Personen) niedriger als für STAs mit restriktiv positiver Empfehlung (1.600 Personen).

<sup>68</sup> Der Median für den potenziellen BI (23 Mio. GBP) liegt durchschnittlich unter dem arithmetischen Mittel. Zudem ist er im Gegensatz zum Mittelwert für negative Empfehlungen (24 Mio. GBP) niedriger als für positiv restriktiv bewertete STAs (27 Mio. GBP).

<sup>69</sup> Der Median für den inkrementellen BI (7 Mio. GBP) liegt durchschnittlich unter dem arithmetischen Mittel. Zudem ist er im Gegensatz zum Mittelwert für negative Empfehlungen (10 Mio. GBP) niedriger als für uneingeschränkt positiv bewertete STAs (11 Mio. GBP).



**Abbildung 21:** [Inkrementelle] Budget Impact-Szenarien für NICE auf Indikationsebene – inklusive CDF (Eigene Darstellung)

### 5.3.3 Resultate der Regressionsanalyse für NICE

Die deskriptiven Untersuchungsergebnisse stellen die zugrunde liegende Evidenz für die vorliegenden Unterschiede der einzelnen Attribute dar, die mittels logistischer Regressionsanalyse auf statistische Signifikanz getestet wurden. Entsprechend sind die Evaluationskriterien aus Regressionsgleichung [2] dahingehend analysiert worden, welchen Effekt respektive Einfluss sie auf die NICE STA Outcomes haben.

#### 5.3.3.1 Ordinales Logit-Modell auf Subgruppenebene

Auf Subgruppenebene zeigen die Resultate des ordinalen Logit-Modells (vgl. Tabelle 20), dass zwischen den positiven NICE STAs und dem Vorhandensein einer Vergleichstherapie ( $p < 0,05$ ) ein signifikanter Zusammenhang besteht, wohingegen das Vorliegen von relevanten RCTs ( $p > 0,1$ ) keinen eindeutigen Effekt auf die Empfehlungen durch NICE zu haben scheint. Die klinische Effektivität ( $p > 0,1$ ) zeigt ebenfalls keine signifikante Korrelation zu den NICE STA Guidance, wobei eine isolierte Betrachtung ohne Berücksichtigung der Kosteneffektivität nur bedingt aussagekräftig wäre (vgl. hierzu auch Kapitel 5.3.1.1).

Die Ergebnisse für die ICER-Variable lassen eindeutig erkennen, dass NICE STAs primär von der plausibelsten Schätzung für das Kosteneffektivitäts-Verhältnis beeinflusst werden. Dabei

wird offensichtlich, dass mit zunehmendem ICER ( $> 30.000$  GBP/QALY) die Wahrscheinlichkeit für eine Ablehnung respektive negative Bewertung durch NICE signifikant steigt ( $p < 0,01$ ). Bei den Krebstherapien, die (in den meisten Fällen auch) einen ICER über dem oberen Schwellenwert von  $30.000$  GBP/QALY aufweisen, weist zudem die Berücksichtigung der EoL-Kriterien einen signifikant positiven Einfluss auf die Empfehlungen durch NICE auf ( $p < 0,01$ ). Für die im Modell ergänzend berücksichtigte Indikatorvariable Veröffentlichungsjahr konnte ein positiver Zusammenhang der NICE STA Guidance zu den Jahren 2014 und 2015 ( $p < 0,1$ ) festgestellt werden. Für das ebenfalls als Dummy-Variable kodierte, aber nicht gesondert dargestellte Indikationsgebiet zeigte sich bei Haut- und Infektionskrankheiten ( $p < 0,05$ ) sowie den Stoffwechselerkrankungen ( $p < 0,1$ ) und onkologischen Krankheitsbildern ( $p < 0,1$ ; ohne zusätzliche Berücksichtigung der EoL-Variable) ein negativer Zusammenhang zu den Empfehlungen durch NICE (vgl. Anhang, Tabelle A4).

Der potenzielle BI<sup>70</sup> hat offensichtlich einen Effekt auf die Empfehlungen durch NICE ( $p < 0,1$ ), wie für zwei der drei Szenarien (Durchschnitt; Maximum) aufgezeigt werden konnte. Dies zeigt sich insbesondere dadurch, dass NICE in seiner Vorgehensweise äußerst strikt gegenüber Behandlungen mit sehr hohen Kosten und damit (in der Regel auch) gegenüber Therapien mit einem sehr hohen potenziellen BI zu sein scheint ( $p < 0,1$ ; inverse U-Kurve zwischen dem potenziellen BI und den NICE STA Outcomes). Das bedeutet, dass sich die Wahrscheinlichkeit einer positiven Empfehlung durch NICE bis zu einem bestimmten Wertebereich zu erhöhen scheint und demnach Behandlungstechnologien mit einem (potenziellen) BI, die über diesem Schwellenwert (geschätztes Mittel:  $150$  Mio. GBP; mit einem entsprechenden Median von etwa  $28,1$  Mio. GBP) liegen, deutlich seltener positiv bewertet werden. Demgegenüber zeigte der inkrementelle BI vergleichbar zu den Ergebnissen für den potenziellen und inkrementellen BI der NHS-Population keinen signifikanten Zusammenhang ( $p > 0,1$ ) zu den Empfehlungen durch NICE auf. Abschließend bleibt festzuhalten, dass sich die Resultate der Regressionsanalyse als weitgehend robust darstellten, nachdem die obersten fünf Prozent der BI-Schätzungen ausgeschlossen wurden.

---

<sup>70</sup> Für die statistisch signifikanten Szenarien ist erkennbar, dass die in der Regression berücksichtigte Kontrollvariable Arzneimittelkosten pro Patient einen signifikanten Einfluss auf die Outcome-Variable ( $p < 0,1$ ; inverser U-Kurven-Effekt bei zunehmenden Kosten und positiver Empfehlung) aufweist (vgl. hierzu auch Anhang, Tabelle A4).

**Tabelle 20:** Resultate der Regressionsanalyse für NICE auf Subgruppenebene  
(Eigene Darstellung)

Ordinales Logit-Modell														
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Vergleichstherapie (Dummy)	1,469**	1,464**	1,382**	1,437**	1,403**	1,502**	1,401**	1,495**	1,404**	1,513**	--	--	--	--
Klinische Studienevidenz: RCT (Dummy)	0,584	0,710	0,596	0,601	0,592	0,600	0,592	0,600	0,592	0,602	--	--	--	--
ICER: 20.000 - 30.000 GBP/QALY (Dummy)	-0,285	-0,878	-0,127	-0,090	-0,224	-0,172	-0,276	-0,220	-0,239	-0,297	0,458	0,462	0,549	0,494
ICER: > 30.000 GBP/QALY (Dummy)	0,690	0,799	0,692	0,693	0,697	0,700	0,699	0,702	0,699	0,703	0,865	0,873	0,869	0,887
Klinische Wirksamkeit: Vergleichbar (Dummy)	-0,390	0,087	-0,468	-0,448	-0,406	-0,381	-0,370	-0,356	-0,395	-0,416	-0,385	-0,360	-0,398	-0,427
Klinische Wirksamkeit: Besser (Dummy)	0,402	1,901	0,409	0,409	0,407	0,407	0,408	0,407	0,406	0,405	0,415	0,421	0,419	0,423
End-of-Life-Kriterien (Dummy)	-2,341**	-3,166	-2,473**	-2,457**	-2,344**	-2,345**	-2,324**	-2,335**	-2,338**	-2,357**	-2,373**	-2,387**	-2,317**	-2,348**
Potenzieller Budget Impact: Minimum (Log)	0,459	2,017	0,470	0,470	0,461	0,462	0,460	0,461	0,460	0,461	0,472	0,475	0,473	0,476
Potenzieller Budget Impact: Minimum (Log) <sup>2</sup>	0,133	--	0,833	0,492	0,167	0,230	0,096	0,177	0,140	0,369	0,010	0,069	-0,027	0,186
Potenzieller Budget Impact: Durchschnitt (Log)	2,248	--	2,315	2,347	2,253	2,267	2,251	2,264	2,253	2,268	2,307	2,333	2,327	2,344
Potenzieller Budget Impact: Maximum (Log)	0,831	-0,233	1,475	1,099	0,851	0,864	0,783	0,825	0,826	1,024	0,729	0,733	0,709	0,897
Potenzieller Budget Impact: Maximum (Log) <sup>2</sup>	2,229	0,543	2,295	2,334	2,233	2,247	2,232	2,244	2,234	2,246	2,285	2,310	2,304	2,320
Potenzieller Budget Impact: Durchschnitt (Log)	2,561**	0,729												
Potenzieller Budget Impact: Maximum (Log)												1,639*	0,914	
Potenzieller Budget Impact: Maximum (Log) <sup>2</sup>												-0,049*	0,027	
Potenzieller Budget Impact: Durchschnitt (Log)														
Potenzieller Budget Impact: Maximum (Log)							0,030	1,473*						
Potenzieller Budget Impact: Maximum (Log) <sup>2</sup>							0,093	0,874						
Potenzieller Budget Impact: Durchschnitt (Log)								-0,043*						
Potenzieller Budget Impact: Maximum (Log)								0,026						

**Tabelle 20:** (Fortsetzung)

Variable	Ordinales Logit-Modell													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Potenzieller Budget Impact: NHS-Population (Log)									-0,004	1,032				0,800
Potenzieller Budget Impact: NHS-Population (Log) <sup>2</sup>									0,089	0,837				0,883
Inkrementeller Budget Impact: Durchschnitt (Log)										-0,030				-0,024
									0,024					0,025
Inkrementeller Budget Impact: NHS-Population (Log)											0,116	0,126		
											0,123	0,125		
Intercept 1	-2,347	-17,652	-3,409	2,672	-2,590	9,701	-2,018	9,656	-2,412	6,379	-3,060	9,878	0,112	0,122
Intercept 2	2,414	4413,685	2,554	6,938	2,626	7,582	2,640	7,525	2,615	7,536	2,551	7,706	0,125	0,126
	1,276	-14,967	0,288	6,381	1,031	13,359*	1,606	13,313*	1,209	10,020	0,684	13,673*	2,572	7,812
	2,407	4413,685	2,540	6,950	2,620	7,613	2,639	7,557	2,609	7,559	2,541	7,744	0,770	7,345
Veröffentlichungsjahr (Dummies)	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Indikation / Therapeutisches Gebiet (Dummies)	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Beobachtungen	255	85	251	251	254	254	254	254	254	254	235	235	232	232

ICER, Incremental Cost-Effectiveness Ratio; NHS, National Health Service; QALY, Quality-Adjusted Life Year; RCT, Randomized Controlled Trial.

Regressionsergebnisse sind statistisch signifikant bei: \*\*\* p < 0,01; \*\* p < 0,05; \* p < 0,1.



### 5.3.3.2 Ordinales Logit-Modell auf Indikationsebene

Auf Indikationsebene sind die vorliegenden Resultate der Regressionsanalyse weitgehend identisch (vgl. Tabelle 21). Während die Variable 'Vergleichstherapie' einen signifikanten Einfluss auf die positiven NICE STA Outcomes aufweist ( $p < 0,05$ ), zeigen weder die Attribute klinische Studienevidenz ( $p > 0,1$ ) noch die klinische Wirksamkeit ( $p > 0,1$ ) einen eindeutigen Effekt auf die NICE Guidance. Demgegenüber ist die Kosteneffektivität der primäre Einflussfaktor; bei einem ICER über dem Schwellenwert von 30.000 GBP/QALY erhöht sich die Wahrscheinlichkeit einer negativen Empfehlung durch NICE signifikant ( $p < 0,01$ ), außer es handelt sich um eine EoL-Therapie. Dabei zeigt sich für die EoL-Kriterien ebenfalls ein signifikanter Effekt auf positive NICE STA-Outcomes ( $p < 0,01$ ), welche bis zum Abschluss des Untersuchungszeitraums ausschließlich für onkologische Interventionen berücksichtigt wurden.

Für die in nachstehender Tabelle 21 nicht ausführlich dargestellten Indikatorvariablen lässt sich festhalten, dass die Veröffentlichungsjahre keinen eindeutigen Zusammenhang zu den STAs zu haben scheinen, wohingegen zwei der Indikationsgebiete, darunter Hautkrankheiten ( $p < 0,1$ ) und Krebserkrankungen ( $p < 0,1$ ; ohne zusätzliche Berücksichtigung der EoL-Variable), einen negativen, jedoch nicht robusten Einfluss auf die Empfehlungen durch NICE aufweisen (vgl. Anhang, Tabelle A5).

Im Gegensatz zum potenziellen BI-Szenario der NHS-Population (ex-ante Betrachtung) zeigt der BI<sup>71</sup> auf Indikationsebene in vergleichbarem Maße wie bei der Analyse auf Subgruppenebene einen Zusammenhang zu den positiven NICE STAs ( $p < 0,1$ ). Dabei zeigt sich auch auf Indikationsebene, dass der 'U-Kurven-Effekt' für das potenzielle BI-Szenario deutlich erkennbar ist, wonach die Wahrscheinlichkeit einer positiven Empfehlung durch NICE bis zu einem bestimmten BI-Schwellenwert zunimmt (vgl. hierzu Kapitel 5.3.3.1, Analyse auf Subgruppenebene). Für das inkrementelle BI-Szenario konnten wiederum keine eindeutigen Zusammenhänge ( $p > 0,1$ ) zu den Empfehlungen durch NICE festgestellt werden. Darüber hinaus zeigten sich die Resultate des Regressionsmodells auf Indikationsebene auch bei Ausschluss der Extremwerte als äußerst robust.

---

<sup>71</sup> Auf Indikationsebene weist die im Regressionsmodell vorliegende Kontrollvariable Patientenpopulation ( $p < 0,1$ ; Zusammenhang von zunehmenden Kosten und positiver Empfehlung) im Vergleich zur Kostenvariable ( $p < 0,1$ ; inverser U-Kurven-Effekt bei zunehmenden Kosten und positiver Empfehlung) einen (mindestens) vergleichbaren Effekt auf die Outcome-Variable auf (vgl. hierzu auch Anhang, Tabelle A5).

**Tabelle 21:** Resultate der Regressionsanalyse für NICE auf Indikationsebene (Eigene Darstellung)

Variable	Ordinales Logit-Modell					
	(1)	(2)	(3)	(4)	(5)	(6)
Vergleichstherapie (Dummy)	1,216**	1,323*	1,184**	1,258**	--	--
	0,591	0,709	0,602	0,608	--	--
Klinische Studienevidenz: RCT (Dummy)	-0,155	-1,088	-0,012	0,043	0,671	0,834
	0,737	0,894	0,743	0,744	0,917	0,930
ICER: 20.000 - 30.000 GBP/QALY (Dummy)	-0,526	0,162	-0,623	-0,672	-0,515	-0,683
	0,445	1,870	0,453	0,456	0,463	0,480
ICER: > 30.000 GBP/QALY (Dummy)	-2,243***	-2,917	-2,320***	-2,378***	-2,260***	-2,443***
	0,519	1,991	0,527	0,531	0,535	0,553
Klinische Wirksamkeit: Vergleichbar (Dummy)	0,262	--	0,461	0,523	0,161	0,417
	2,050	--	2,060	2,077	2,101	2,137
Klinische Wirksamkeit: Besser (Dummy)	0,878	-0,226	1,099	1,094	0,779	0,993
	2,031	0,553	2,044	2,060	2,081	2,120
End-of-Life-Kriterien (Dummy)		2,525***				
		0,746				
Potenzieller Budget Impact: Durchschnitt (Log)			-0,132	2,024*		2,326**
			0,108	1,104		1,136
Potenzieller Budget Impact: Durchschnitt (Log) <sup>2</sup>				-0,063*		-0,071**
				0,032		0,033
Inkrementeller Budget Impact: Durchschnitt (Log)					0,138	0,149
					0,126	0,127
Intercept 1	-2,415	-17,761	-4,131	13,851	-2,956	15,795
	2,301	6191,370	2,698	9,533	2,435	9,721
Intercept 2	0,597	-15,234	-1,110	16,925*	0,180	19,013*
	2,292	6191,370	2,681	9,567	2,422	9,764
Veröffentlichungsjahr (Dummies)	Ja	Ja	Ja	Ja	Ja	Ja
Indikation / Therapeutisches Gebiet (Dummies)	Ja	Ja	Ja	Ja	Ja	Ja
Beobachtungen	198	79	197	197	180	180

ICER, Incremental Cost-Effectiveness Ratio; QALY, Quality-Adjusted Life Year; RCT, Randomized Controlled Trial.

Regressionsergebnisse sind statistisch signifikant bei: \*\*\*  $p < 0,01$ ; \*\*  $p < 0,05$ ; \*  $p < 0,1$ .

### 5.3.4 Resultate der Regressionsanalyse für NICE unter Berücksichtigung des CDF

Die Analyseergebnisse auf Subgruppen- und Indikationsebene unter Heranziehung des CDF als eine zusätzliche Outcome-Variable (mit einer Empfehlung durch NICE) sind weitgehend vergleichbar mit den zuvor dargelegten Resultaten für NICE, die den CDF nicht explizit berücksichtigten.

#### 5.3.4.1 Ordinales Logit-Modell auf Subgruppenebene – inklusive CDF

Auf Subgruppenebene haben ausschließlich die Variablen Vergleichstherapie ( $p < 0,05$ ) sowie das Kosteneffektivitäts-Verhältnis ( $p < 0,01$ ; bei einem ICER  $> 30.000$  GBP/QALY) einen signifikant negativen Einfluss auf die STA-Empfehlungen. Die EoL-Kriterien zeigen demgegenüber einen eindeutig positiven Effekt auf die Bewertung von Krebstherapien ( $p < 0,01$ ) durch NICE. Ein erkennbarer Unterschied zum ordinalen Logit-Modell ohne CDF-Berücksichtigung scheint darüber hinaus zu sein, dass der klinischen Wirksamkeit eine gewichtigere Rolle zukommt und die Kosteneffektivität folglich weniger entscheidend für eine Empfehlung durch NICE ist, wobei sich die Ergebnisse als nicht robust herausstellten (vgl. Tabelle 22).

Für die im ordinalen Logit-Modell zusätzlich eingeschlossene Indikatorvariable Veröffentlichungsjahr zeigt sich ein signifikant positiver Effekt für die Jahre 2014 bis 2016 ( $p < 0,05$ ) sowie – auffällig deutlich – für das Jahr 2017 ( $p < 0,01$ ). Bei den Indikationsgebieten weisen die Haut- ( $p < 0,05$ ) und Infektionserkrankungen ( $p < 0,01$ ) sowie (zumindest teilweise auch) die Stoffwechselkrankheiten ( $p < 0,1$ ) einen negativen Zusammenhang zu den NICE-Empfehlungen auf (vgl. Anhang, Tabelle A6).

Offensichtlich sind die Bewertungen der Krebstherapien nach dem Relaunch des CDF im Juli 2016 besser bewertet als in den vorangegangenen Jahren (Januar 2011 bis Juni 2016). Da durch den CDF erstattete Technologien jedoch (nach spätestens zwei Jahren) einer erneuten Evaluation durch NICE unterzogen werden (müssen), um über eine routinemäßige Aufnahme in das NHS entscheiden zu können, unterliegt die vorliegende Erkenntnisfeststellung demzufolge Unsicherheiten hinsichtlich der zukünftigen STA-Empfehlung.

Der potenzielle BI<sup>72</sup> weist für das Durchschnitts-Szenario einen Zusammenhang zu den Empfehlungen durch NICE ( $p < 0,1$ ) auf. Vergleichbar der Analyse auf Subgruppenebene ohne explizite Berücksichtigung des CDF ist dabei erkennbar, dass NICE Technologien mit sehr hohen Kosten und damit (in der Regel) einem sehr hohen BI tendenziell negativer zu bewerten scheint ( $p < 0,1$ ; inverse U-Kurve zwischen dem potenziellen BI und den NICE STA Outcomes). Offensichtlich scheint dieser Effekt für onkologische und hämatologische Therapien jedoch (weitgehend) wegzufallen, seit (sehr) teure Krebstechnologien alternativ oder besser gesagt, nach-

---

<sup>72</sup> Die im Regressionsmodell zugrunde liegende Kontrollvariable Arzneimittelkosten pro Patient zeigt für das Durchschnitts-Szenario einen signifikanten, jedoch nicht robusten Einfluss zur Outcome-Variable ( $p < 0,1$ ; inverser U-Kurven-Effekt zwischen zunehmenden Kosten und positiver Empfehlung) auf (vgl. hierzu auch Anhang, Tabelle A6).

träglich (das heißt nach der abschließenden Empfehlung durch NICE) durch den CDF berücksichtigt und demzufolge – wenn auch zeitlich zunächst befristet – erstattet werden können. Prinzipiell bleibt der festgestellte U-kurvenförmige Effekt damit für (nicht-onkologische) Behandlungen mit einem sehr hohen BI bestehen, wenn sie durch den CDF unberührt beziehungsweise unberücksichtigt bleiben.

Im Gegensatz zur Analyse ohne explizite CDF-Berücksichtigung zeigt sich darüber hinaus auch für den potenziellen BI der NHS-Population ein positiver Effekt auf die Empfehlungen durch NICE ( $p < 0,1$ ; inverse U-Kurve zwischen dem potenziellen BI der NHS-Population und den NICE STA Outcomes). Da der Effekt ebenfalls bis zu einem bestimmten Schwellenwert (geschätztes Mittel: 350 Mio. GBP; mit einem entsprechenden Median von etwa 74,2 Mio. GBP) nachweisbar ist, ab welchem die Wahrscheinlichkeit einer negativen Bewertung durch NICE zunimmt, scheint der CDF (zumindest) auch die Empfehlungen einzelner (sehr) teurer Krebstherapien innerhalb der (ex-ante definierten) NHS-Population zu beeinflussen. Für die inkrementellen BI-Szenarien konnte kein eindeutiger Einfluss auf die STAs ( $p > 0,1$ ) festgestellt werden. Die Resultate der vorgenommenen Regression erweisen sich zudem als äußerst robust, auch wenn die obersten fünf Prozent der BI-Schätzungen aus dem ordinalen Logit-Modell ausgeschlossen wurden.

**Tabelle 22:** Resultate der Regressionsanalyse für NICE auf Subgruppenebene – inklusive CDF  
(Eigene Darstellung)

Ordinales Logit-Modell unter Berücksichtigung des Cancer Drugs Fund (CDF)														
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Vergleichstherapie (Dummy)	-1,448**	-1,829**	-1,308**	-1,254**	-1,284**	-1,218**	-1,289**	-1,228**	-1,260**	-1,141*	--	--	--	--
Klinische Studienevidenz: RCT (Dummy)	0,586	0,730	0,595	0,599	0,597	0,600	0,598	0,601	0,593	0,598	--	--	--	--
ICER: 20.000 - 30.000 GBP/QALY (Dummy)	-0,183	-1,224	-0,233	-0,204	-0,274	-0,247	-0,292	-0,262	-0,154	-0,213	-0,625	-0,687	-0,659	-0,606
ICER: > 30.000 GBP/QALY (Dummy)	0,673	0,834	0,678	0,680	0,681	0,684	0,683	0,686	0,677	0,670	0,863	0,883	0,867	0,850
Klinische Wirksamkeit: Vergleichbar (Dummy)	-0,471	0,024	-0,492	-0,478	-0,470	-0,455	-0,459	-0,450	-0,514	-0,548	-0,544	-0,508	-0,539	-0,631
End-of-Life-Kriterien (Dummy)	0,412	1,948	0,417	0,415	0,416	0,415	0,417	0,415	0,415	0,414	0,427	0,432	0,429	0,432
Potenzieller Budget Impact: Minimum (Log)	-1,913**	-2,396	-1,999**	-1,982**	-1,929**	-1,926**	-1,923**	-1,927**	-1,963**	-1,996**	-2,072**	-2,070**	-2,045**	-2,141**
Potenzieller Budget Impact: Minimum (Log) <sup>2</sup>	0,455	2,084	0,460	0,461	0,457	0,457	0,457	0,457	0,457	0,458	0,474	0,479	0,476	0,482
Potenzieller Budget Impact: Durchschnitt (Log)	-0,652	--	-0,377	-0,759	-0,652	-0,594	-0,673	-0,602	-0,514	-0,164	-0,393	-0,379	-0,361	0,166
Potenzieller Budget Impact: Maximum (Log)	2,300	--	2,351	2,389	2,309	2,324	2,306	2,319	2,317	2,345	2,391	2,425	2,415	2,460
Potenzieller Budget Impact: Maximum (Log) <sup>2</sup>	0,752	1,593**	1,037	0,618	0,786	0,802	0,765	0,803	0,929	1,238	1,039	0,998	1,050	1,543
Potenzieller Budget Impact: Durchschnitt (Log)	2,283	0,562	2,334	2,377	2,291	2,306	2,289	2,302	2,299	2,326	2,372	2,406	2,396	2,439
Potenzieller Budget Impact: Maximum (Log)	1,815**	--	--	--	--	--	--	--	--	--	--	--	--	--
Potenzieller Budget Impact: Maximum (Log) <sup>2</sup>	0,693	--	--	--	--	--	--	--	--	--	--	--	--	--
Potenzieller Budget Impact: Durchschnitt (Log)	-0,054	0,868	-0,054	0,868	-0,006	1,333	0,011	1,236	1,333	1,333	1,333	1,920**	1,920**	1,920**
Potenzieller Budget Impact: Maximum (Log)	0,090	0,893	0,090	0,893	0,095	0,901	0,095	0,892	0,901	0,901	0,901	0,951	0,951	0,951
Potenzieller Budget Impact: Maximum (Log) <sup>2</sup>	-0,028	0,027	-0,028	0,027	-0,040	0,027	-0,036	0,026	0,027	0,027	0,027	-0,056**	-0,056**	-0,056**
Potenzieller Budget Impact: Durchschnitt (Log)	0,027	--	--	--	--	--	--	--	--	--	--	0,028	0,028	0,028

**Tabelle 22:** (Fortsetzung)

Variable	Ordinales Logit-Modell unter Berücksichtigung des Cancer Drugs Fund (CDF)													
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Potenzieller Budget Impact: NHS-Population (Log)									-0,095	1,519*				1,680*
Potenzieller Budget Impact: NHS-Population (Log) <sup>2</sup>									0,093	0,859				0,931
Inkrementeller Budget Impact: Durchschnitt (Log)									-0,047*	0,025				-0,051*
											0,055	0,058		0,027
											0,124	0,125		
Inkrementeller Budget Impact: NHS-Population (Log)													0,056	0,089
													0,125	0,128
Intercept 1	-4,796*	-17,392	-5,099*	1,844	-4,815*	5,953	-4,618*	5,282	-5,844**	7,855	-4,265	11,048	-4,311	9,496
	2,483	1809,867	2,607	7,168	2,705	7,675	2,719	7,652	2,716	7,723	2,646	8,000	2,672	8,285
Intercept 2	-1,038	-14,522	-1,303	5,662	-1,047	9,760	-0,850	9,083	-2,057	11,701	-0,312	15,071*	-0,300	13,59
	2,463	1809,867	2,585	7,190	2,686	7,709	2,702	7,685	2,686	7,763	2,627	8,055	2,653	8,331
Veröffentlichungsjahr (Dummies)	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Indikation / Therapeutisches Gebiet (Dummies)	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja	Ja
Beobachtungen	255	85	251	251	254	254	254	254	254	254	235	235	232	232

ICER, Incremental Cost-Effectiveness Ratio; NHS, National Health Service; QALY, Quality-Adjusted Life Year; RCT, Randomized Controlled Trial.

Regressionsergebnisse sind statistisch signifikant bei: \*\*\* p < 0,01; \*\* p < 0,05; \* p < 0,1.

### 5.3.4.2 Ordinales Logit-Modell auf Indikationsebene – inklusive CDF

Die Resultate auf Indikationsebene unterstützen die Ergebnisse der Subgruppenanalyse (vgl. Tabelle 23). Sowohl das Vorhandensein einer Vergleichstherapie ( $p < 0,05$ ) als auch die Kosteneffektivität ( $p < 0,01$ ; bei einem ICER  $> 30.000$  GBP/QALY) weisen einen signifikant negativen Zusammenhang zu den Empfehlungen von NICE auf. Die EoL-Kriterien zeigen bei Berücksichtigung einen ebenfalls eindeutig positiven Effekt auf die NICE STAs auf und zugleich, dass bei EoL-Therapien die klinische Wirksamkeit ( $p < 0,01$ ) einen größeren Einfluss zu haben scheint als das Kosteneffektivitäts-Verhältnis.

**Tabelle 23:** Resultate der Regressionsanalyse für NICE auf Indikationsebene – inklusive CDF (Eigene Darstellung)

Variable	Ordinales Logit-Modell unter Berücksichtigung des Cancer Drugs Fund					
	(1)	(2)	(3)	(4)	(5)	(6)
Vergleichstherapie (Dummy)	-1,441**	-1,930**	-1,249**	-1,218**	--	--
	0,605	0,781	0,612	0,617	--	--
Klinische Studienevidenz: RCT (Dummy)	-0,516	-2,405*	-0,462	-0,455	-0,663	-0,543
	0,743	1,234	0,746	0,752	0,894	0,912
ICER: 20.000 - 30.000 GBP/QALY (Dummy)	-0,517	0,282	-0,605	-0,656	-0,591	-0,769
	0,456	1,911	0,466	0,468	0,475	0,495
ICER: > 30.000 GBP/QALY (Dummy)	-1,673***	-2,095	-1,768***	-1,818***	-1,893***	-2,075***
	0,518	2,053	0,527	0,531	0,544	0,567
Klinische Wirksamkeit: Vergleichbar (Dummy)	-0,557	--	-0,387	-0,328	-0,300	-0,065
	2,067	--	2,081	2,101	2,139	2,185
Klinische Wirksamkeit: Besser (Dummy)	0,723	1,263**	0,966	0,961	1,037	1,236
	2,050	0,579	2,067	2,087	2,121	2,172
End-of-Life-Kriterien (Dummy)		1,884***				
		0,727				
Potenzieller Budget Impact: Durchschnitt (Log)			-0,117	2,051*		2,707**
			0,110	1,099		1,148
Potenzieller Budget Impact: Durchschnitt (Log) <sup>2</sup>				-0,063**		-0,082**
				0,032		0,033
Inkrementeller Budget Impact: Durchschnitt (Log)					0,061	0,073
					0,127	0,129
Intercept 1	-5,178**	-18,914	-6,527**	11,408	-4,256*	17,633*
	2,375	2407,029	2,738	9,433	2,465	9,799
Intercept 2	-2,114	-16,236	-3,442	14,558	-1,004	20,997**
	2,349	2407,029	2,706	9,473	2,443	9,860
Veröffentlichungsjahr (Dummies)	Ja	Ja	Ja	Ja	Ja	Ja
Indikation / Therapeutisches Gebiet (Dummies)	Ja	Ja	Ja	Ja	Ja	Ja
Beobachtungen	198	79	197	197	180	180

ICER, Incremental Cost-Effectiveness Ratio; QALY, Quality-Adjusted Life Year; RCT, Randomized Controlled Trial.

Regressionsergebnisse sind statistisch signifikant bei: \*\*\*  $p < 0,01$ ; \*\*  $p < 0,05$ ; \*  $p < 0,1$ .

Während für die im Regressionsmodell berücksichtigte, in Tabelle 23 jedoch nicht gesondert dargestellte Dummy-Variable Indikationsgebiet eindeutige Zusammenhänge festgestellt werden konnten, zeigten die Veröffentlichungsjahre keinen signifikant robusten Effekt auf die STA-Outcomes auf. Das Indikationsgebiet der Hautkrankheiten ( $p < 0,05$ ) scheint demnach einen signifikant negativen Einfluss auf die Empfehlungen durch NICE zu haben. Ähnliches gilt für die Atemwegs-, Infektions- und Muskel-Skelett-Erkrankungen ( $p < 0,1$ ), wobei sich die Resultate nicht durchgängig als robust herausstellten (vgl. Anhang, Tabelle A5).

Vergleichbar der NHS-Population (ex-ante Betrachtung) scheint der potenzielle BI<sup>73</sup> einen signifikanten Zusammenhang zu den NICE STAs ( $p < 0,1$ ) zu haben. Der auf Subgruppenebene festgestellte Effekt, dass NICE Behandlungstherapien mit einem sehr hohen BI tendenziell negativer bewertet ( $p < 0,1$ ; inverse U-Kurve zwischen dem potenziellen BI und den NICE STA Outcomes), zeigt sich demnach auch auf Indikationsebene. Ausschließlich für Krebstherapien scheint sich der Effekt seit dem Relaunch des CDF und der damit einhergehenden Möglichkeit einer (nachträglichen) Berücksichtigung respektive Erstattung von Onkologika weitgehend aufzuheben, sodass sich diese Erkenntnis auf der Ebene der (Haupt-)Indikationen bestätigen lässt. Vergleichbar dem Regressionsmodell auf Subgruppenebene stellten sich bei Ausschluss der Extremwerte auch die hier vorliegenden Analyseresultate als robust dar.

---

<sup>73</sup> Für die im Regressionsmodell vorliegende Kontrollvariable Arzneimittelkosten pro Patient konnte ebenfalls ein Effekt auf die Outcome-Variable ( $p < 0,1$ ; inverser U-Kurven-Effekt zwischen zunehmenden Kosten und positiver Empfehlung) festgestellt werden (vgl. hierzu auch Anhang, Tabelle A5).



## 6 DISKUSSION

### 6.1 Herausstellung und kritische Reflexion der wesentlichen Erkenntnisse

#### 6.1.1 Vergleichende Untersuchung von GBA, HAS und NICE

Vergleichbar den Resultaten früherer empirischer HTA-Vergleichsstudien bestätigen die vorliegenden Ergebnisse des systematischen Abgleichs der GBA FNBs, HAS ACTs und NICE STAs, dass Outcomes respektive Empfehlungen nationaler HTA-Institutionen voneinander abweichen (Akehurst et al., 2017; Allen et al., 2017; Boucaud-Maitre et al., 2021; Cerri et al., 2015; Fischer et al., 2016; Nicod and Kanavos, 2012; Nicod et al., 2020a; Schaefer et al., 2021; Schaefer and Schlander, 2019). Vielmehr zeigt die systematische Untersuchung der HTA-Outcomes übereinstimmender Interventionspaare, dass NICE neue Gesundheitstechnologien insgesamt positiver bewertet als GBA und HAS. Teilweise bestehen jedoch erhebliche Diskrepanzen bei Betrachtung der HTA-Outcomes nach Indikations- respektive Anwendungsgebieten der Interventionen. Während die Bewertungen von onkologischen Therapien durch den GBA im Vergleich zu HAS und NICE positiver ausfallen, weisen die Evaluationen nicht-onkologischer Therapien eindeutig positivere HTA-Resultate für NICE auf (Schaefer et al., 2021).

Beim direkten, paarweisen Vergleich der drei HTA-Agenturen zeigt sich eine höhere Übereinstimmung für die Outcomes von Krebstherapien im Vergleich zu nicht-onkologischen Behandlungsinterventionen. Darüber hinaus konnte zudem festgestellt werden, dass die HTA-Ergebnisse von GBA und HAS gegenüber den Empfehlungen von NICE eine insgesamt höhere Kongruenz aufweisen (Schaefer et al., 2021; vgl. hierzu auch bei Boucaud-Maitre et al., 2021). Dies scheint primär auf die bestehenden Ähnlichkeiten zwischen den deutschen und französischen HTA-Ansätzen<sup>74</sup> zurückzuführen zu sein, wie unter anderem auf den gemeinsamen Bewertungsschwerpunkt, der explizit auf dem klinisch-therapeutischen Nutzen einer neuen Therapie oder genauer gesagt der (vergleichenden) klinischen Effektivität liegt (Allen et al., 2017; Angelis et al., 2018).

Bevor die Bewertungsfaktoren in den Vordergrund rücken, sollten jedoch zunächst die grundlegenden Unterschiede zwischen den nationalen Strukturen und den offiziellen HTA-Agenturen

---

<sup>74</sup> Neben den bestehenden Differenzen weisen die HTA-Methoden von GBA/IQWiG und HAS (sowie zumindest teilweise auch von NICE) einige grundlegende Gemeinsamkeiten an die (Qualitäts-)Anforderungen in den (klinischen) Bewertungsschwerpunkten auf, insbesondere bei: (1) (Studien-)Evidenz, (2) Krankheitslast, (3) klinisch-medizinischer Sicherheit und (4) therapeutisch-klinischem Impact (klinische Effektivität); sowie eingeschränkt für: (5) Innovationsgrad und (6) sozioökonomischen respektive Public Health Impact (Angelis et al., 2018).

hervorgehoben werden (Schaefer et al., 2021): Zum einen der bestehende institutionelle Kontext, in welchem die Einrichtungen agieren und zum anderen ihre jeweilige Rolle innerhalb der vorliegenden Entscheidungs- und Erstattungsprozesse. Während sowohl HAS in Frankreich als auch NICE in England rechtlich verbindliche Empfehlungen für die Erstattung neuer Therapien und Medikamente veröffentlichen, fließen die Beschlüsse des GBA zwar in die anschließenden Preisverhandlungen, nicht aber in die Erstattungsentscheidung ein (Akehurst et al., 2017). In England wiederum beeinflusst der Preis dabei lediglich indirekt den Erstattungsstatus einer Gesundheitstechnologie (das heißt durch die Auswirkung auf den bestmöglichen ICER/QALY), wohingegen der aus der Bewertung des SMR abgeleitete ASMR (I-IV) den Erstattungspreis respektive das Preisniveau eines Medikaments in Frankreich direkt bestimmt (Angelis et al., 2018). Das zweistufige Bewertungsverfahren in Deutschland weist im Gegensatz zu den HTA-Leitlinien von HAS und NICE keinen formellen Einspruch während des Bewertungsprozesses auf, auch wenn für die pUs bei Nichtzustandekommen eines Marktpreises (unter zusätzlicher Einbeziehung eines Schiedsgerichts die (nachgelagerte) Option einer formellen Kosten-Nutzen-Bewertung durch IQWiG besteht (Akehurst et al., 2017; Angelis et al., 2018).

Abgesehen von den genannten – und weiteren, jedoch weitaus weniger relevanten – Differenzen sowie den unterschiedlichen Gesundheitssystemen in England, Frankreich und Deutschland (NHS – ‘Beveridge-Modell’ – in England versus Sozialversicherungssystem – ‘Bismarck-Modell’ – in Frankreich und Deutschland), scheinen insbesondere die Heterogenität der analytischen Frameworks und der Bewertungsmethoden den entscheidenden Faktor für die Abweichungen in den HTA-Ergebnissen von GBA, HAS und NICE abzubilden (Schaefer et al., 2021; vgl. hierzu bei Akehurst et al., 2017; Angelis et al., 2018; Nicod and Kanavos, 2012; Nicod et al., 2020a).

Einerseits zeigt sich, dass die (vergleichende) klinische Wirksamkeit in allen drei HTA-Settings eine der Schlüsselkriterien für das jeweilige Bewertungsverfahren darstellt. Die zugrunde liegenden nationalen Leitlinien weisen demnach inhaltliche Überschneidungen auf, insbesondere wenn die evidenzbasierten Anforderungen an die klinischen Studien betrachtet werden. Sowohl GBA und HAS als auch NICE betrachten dabei die Studienevidenz in Form von RCTs (einzeln oder zusammengefasst in Form eines systematischen Reviews) als das zentrale Nachweiselement für die (vergleichende) klinische Wirksamkeit einer neuen Behandlungsintervention (Kleijnen et al., 2012; Stafinski et al., 2011a; Yuasa et al., 2020). Im Vergleich zu den GBA FNBs und HAS ACTs erscheinen NICE STAs jedoch ‘flexibler’; dies zeigt sich primär an einer

erhöhten Akzeptanz gegenüber der Einreichung von Non-RCTs oder indirekten Vergleichsstudien sowie in Ausnahmefällen auch von Interventions- oder Beobachtungsstudien (Angelis et al., 2018; Lebioda et al., 2014; Massetti et al., 2015).

Andererseits kann die Kosteneffektivität als ein wesentliches Element der NICE STAs in England betrachtet werden, während sowohl GBA in Deutschland als auch HAS in Frankreich die Kosten nicht explizit respektive nicht grundsätzlich in die Bewertung einbeziehen (Akehurst et al., 2017; Angelis et al., 2018). Die HTA-Outcomes der HAS stützen sich primär auf vergleichende klinische Wirksamkeits- und Sicherheitsdaten sowie auf den bestehenden Patientenbedarf innerhalb des französischen Krankenversicherungssystems. Demnach ist die gesundheitsökonomische Bewertung nur auf Arzneimittel mit hohem Zusatznutzen (ASMR I-III) beschränkt, was vermutlich damit zusammenhängt, dass Erstattungsentscheidungen in Frankreich direkten Einfluss auf den (effektiven) Arzneimittelzugang von Patientinnen und Patienten haben (Dubromel et al., 2018; Rémuzat et al., 2013; Toumi et al., 2017). Mit der grundsätzlichen Ablehnung eines indikationsübergreifenden Schwellenwerts (wie des QALYs) zeichnet sich für den GBA ein sehr gegensätzliches Bild ab. Während die vergleichende klinische Wirksamkeit – mit dem Fokus auf patientenrelevante Endpunkte – für den Nutznachweis einer neuen Therapie (offiziell) als der zentrale Bewertungsfaktor (sowohl in den IQWiG-Bewertungen als auch) in den GBA-Beschlüssen gilt, spielt die (allenfalls) optionale ökonomische Evaluation in Form des durch IQWiG implementierten Effizienzgrenzen-Konzepts bis heute keine Rolle (Kieslich, 2020; Schaefer and Schlander, 2019).

Eine besondere Herausforderung für die nationalen HTA-Institutionen stellt die Evaluation von Krebstherapien dar, die bei Zulassung respektive den daran anschließenden HTAs das am häufigsten vorkommende Indikationsgebiet abbilden. Der – insbesondere auch in europäischen Ländern – hohe Anteil an Krebs(todes)fällen<sup>75</sup> impliziert zudem, dass den Krebserkrankungen ein besonderes gesellschaftliches Interesse zukommt, weshalb eine (mögliche) Priorisierung gegenüber anderen lebensbedrohlichen Krankheitsbildern (zumindest) nicht per se auszuschließen ist (Culyer, 2017; Morell et al., 2017; Neumann et al., 2012; Pauwels et al., 2014).

Der Zugang zu erstattungsfähigen Krebsmedikamenten scheint dabei in Deutschland und Frankreich weniger restriktiv zu sein als in England, wie auch die zuvor dargelegten Ergebnisse

---

<sup>75</sup> Laut Global Cancer Statistics 2020 sind es weltweit 19,30 Mio. neue Krebsfälle (9,9 Mio. Todesfälle), wovon mit 4,4 Mio. Fällen etwa 22,8 % auf Europa entfallen (Sung et al., 2021). In Europa beziehungsweise der EU ist circa jeder vierte Todesfall (entspricht rund 1,4 Mio. Fällen) auf eine Krebserkrankung zurückzuführen, auch wenn die (alterstandardisierten) Mortalitätsraten weiter rückläufig sind (von 2015 bis 2020: Frauen, - 4,1 %; Männer, - 5,4 %; Carioli et al., 2020).

des paarweisen Vergleichs vermuten lassen (Schaefer et al., 2021; vgl. hierzu bei Adkins et al., 2017; Jarosławski et al., 2019; Zamora et al., 2019). Dies lässt sich auf der einen Seite durch zusätzliche Attribute beziehungsweise Einflussfaktoren erklären, wie unter anderem die seit Einführung des AMNOG geltende OD-Regelung in Deutschland. Auf der anderen Seite spiegeln sich die Abweichungen aber auch darin, in welchem Maße die zugrunde liegenden Kriterien bei der Bewertung priorisiert werden (Maynou and Cairns, 2019). Zum Beispiel scheinen die Beschlüsse neuer Krebstherapien durch den GBA primär von der Morbidität sowie dem Überlebensvorteil (progressionsfreies Überleben und, sofern verfügbar, Gesamtüberleben) gegenüber der jeweiligen Vergleichstherapie abhängig zu sein (Ruof et al., 2014). Für die HAS ACTs innovativer Krebstherapien (zur Behandlung maligner Tumorerkrankungen) zeigt sich, dass die patientenrelevanten Endpunkte progressionsfreies Überleben (oder alternativ auch das Gesamtüberleben) sowie zusätzlich das klinische Studiendesign und dessen Übertragbarkeit in die klinische Praxis stärker gewichtet werden (Chouaid et al., 2016). Zudem merken Drummond et al. (2014) an, dass die ergänzende Durchführung von Kosteneffektivitäts-Analysen – zumindest bei positivem ASMR hinsichtlich der daran anschließenden Preisverhandlungen – weiter an Bedeutung zu gewinnen scheint.

Obwohl sich NICE in seinen Empfehlungen primär auf die Kosten pro Patient stützt, indem es für die inkrementellen Kosten je QALY auf einen explizit vorliegenden Schwellenwert zurückgreift, fallen die Bewertungen neuer Krebstechnologien unter Berücksichtigung der Kriterien für EoL-Therapien (mit einem ICER  $> 30.000$  GBP/QALY) verhältnismäßig besser aus (Bovenberg et al., 2021; Pujolras and Cairns, 2015). Bereits seit seiner Einführung im Jahr 2010 wird der CDF von kontrovers geführten gesundheitspolitischen Debatten hinsichtlich seines (tatsächlichen) Nutzens für die Patientinnen und Patienten im Speziellen sowie für die Gesellschaft im Allgemeinen begleitet (Aggarwal et al., 2017; Cohen, 2017). Seit dem Relaunch im Jahr 2016 liegen jedoch erste Hinweise vor, dass der positive Empfehlungsanteil der NICE STAs für (sehr teure) Krebstechnologien insgesamt zugenommen hat – und dieser zugleich einen nachhaltigen Effekt implizieren könnte, da die pUs bei erneuter Bewertung nach spätestens zwei Jahren eine verbesserte Studienevidenz vorzulegen haben (Morrell et al., 2018; Sabry-Grant et al., 2019; Walton et al., 2019). Dies bestätigen auch die Resultate der Regressionsmodelle für NICE, die in Kapitel 6.1.3 und 6.2.2 ausführlich diskutiert werden.

Ergänzend sind zudem ODs sowie Therapien für sehr seltene Erkrankungen (sogenannte Ultra-Rare Diseases, URDs) anzuführen, die in aller Regel einen nicht unerheblichen Anteil an den evaluierten respektive den zu evaluierenden Krebstherapien darstellen. Die Schlüsselfrage, die in der einschlägigen Literatur im Zusammenhang mit HTA diskutiert wird ist, inwieweit ODs

den standardisierten HTA-Prozessen Stand halten (können) und welche Implikationen dies mit sich bringt beziehungsweise welche alternativen, flexibleren methodischen Vorgehensweisen existieren, um einen wissenschaftlich nachhaltigen Rahmen für deren Bewertung zu schaffen (Nestler-Parr et al., 2018; Nicod et al., 2019; Postma et al., 2022; Schlander et al., 2014; Simoens, 2014). Dass die systematischen Evaluationen von ODs im Vergleich zu anderen Therapien tendenziell schlechter auszufallen scheinen, wenn keine ergänzenden Faktoren berücksichtigt werden oder ein entsprechendes Bewertungsverfahren vorliegt, zeigt sich sowohl in empirischen HTA-Studien zu ODs (Kawalec et al., 2016; Malinowski et al., 2018; Nicod et al., 2017; 2020b; Tordrup et al., 2014) als auch in den vorliegenden Ergebnissen der vergleichenden Untersuchung.

Während in Deutschland ein positiver Zusatznutzen für alle Therapien (inklusive den Behandlungen von URDs) mit einem OD-Status als belegt gilt, wurden in England im Laufe der Zeit entsprechende Anpassungen respektive Ergänzungen in den TA Guidelines vorgenommen. Die methodischen Aktualisierungen durch NICE beziehen sich jedoch primär auf Krebstherapien, wie anhand der Kriterien für EoL-Therapien und der Einführung des CDFs bereits angedeutet wurde (vgl. bei Charlton, 2020). Bei der französischen HAS durchlaufen ODs dieselben Prozessschritte innerhalb des Bewertungsverfahrens wie alle anderen Medikamente, was im direkten Vergleich offensichtlich zu negativeren ACT-Outcomes zu führen scheint (Adkins et al., 2017; Nicod et al., 2017; 2020b). Jedoch bleibt anzumerken, dass der klinische Nutzen für Medikamente mit niedrigem BI (< 30 Mio. EUR), das heißt im Besonderen für Arzneimittel von URDs, als vorhanden eingestuft wird und diese wiederum entsprechend erstattet werden (Tordrup et al., 2014).

Bevor nachfolgend die Resultate aller in Deutschland und England veröffentlichten HTAs diskutiert werden, bleibt noch zu konstatieren, dass weit weniger als die Hälfte der von GBA und NICE bewerteten Therapien (zwischen Januar 2011 und Juni 2018) übereinstimmen<sup>76</sup> und demzufolge für die vergleichende Untersuchung von GBA, HAS und NICE berücksichtigt werden konnten. Diese Erkenntnis deckt sich mit den Ergebnissen von Schaefer *and* Schlander (2019), die für den Zeitraum Januar 2011 bis April 2015 (ohne Berücksichtigung der HAS ACTs) bereits eine vergleichbar geringe Übereinstimmung feststellten. Eine weitere Studie von Cerri et al. (2015), welche unter anderem einen Abgleich der Outcomes der europäischen HTA-

---

<sup>76</sup> Die Zahlen basieren auf dem Abgleich von zunächst GBA FNBs (102/298, 34 %) und NICE STAs (102/252, 40 %), die daraufhin den während des definierten Untersuchungszeitraums veröffentlichten HAS ACTs gegenübergestellt wurden.

Institutionen NICE, HAS, CVZ (College voor Zorgverzekeringen, Niederlande) und SMC (Scottish Medicines Consortium, Schottland) vornahm, identifizierte unter den extrahierten HTAs lediglich zehn Prozent übereinstimmende Wirkstoffe, die von allen vier Agenturen innerhalb eines vergleichbaren Zeitraums bewertet wurden.

Da jedoch innerhalb der EU generell alle Therapien durch die EMA zugelassen werden und sie damit theoretisch auf dem gesamten EU-Markt für Patientinnen und Patienten zugänglich gemacht werden könn(t)en, erscheint es an dieser Stelle unausweichlich, zumindest zwei der möglichen Hintergründe<sup>77</sup> aufzuzeigen, weshalb dies offensichtlich (häufig) nicht der Fall zu sein scheint. Zum einen sind zeitliche Verzögerungen insbesondere in der Einreichung der zu evaluierenden Unterlagen durch die pUs, aber auch in den jeweiligen nationalen HTA-Prozessen anzuführen (vgl. hierzu bei Cerri et al., 2015). Zum anderen scheinen die pUs innovative Therapien und Arzneimittel lediglich auf einzelnen Märkten zu positionieren, was sich primär auf die Strukturen des jeweiligen Gesundheitswesens, die länderspezifischen HTA-Systeme und die inhaltlichen Verfahrensunterschiede bei der systematischen Bewertung zurückführen lässt. Offenbar beraten pUs über die Vermarktung neuer Therapien und Arzneimittel in einzelnen Ländern, indem sie vor einer Produkteinführung die jeweiligen Gegebenheiten sowie mögliche Unsicherheiten versuchen möglichst genau einzuschätzen. Die daraus resultierenden Marktstrategien fokussieren sich darauf, in welchem Land die Aussicht auf eine positive Bewertung aufgrund der bestehenden Datenlage am realistischsten erscheint, da diese Empfehlung die (Erstattungs-)Preisfestsetzung einer Intervention weitgehend bestimmt und demzufolge (für gewöhnlich) auch eine Refinanzierung der getätigten Investitionen erheblich beeinflusst.

### **6.1.2 GBA-Beschlussfassungen und IQWiG-Dossierbewertungen**

Mit der Einführung des AMNOG zum 01. Januar 2011 wurde in Deutschland ein zweistufiges HTA-Verfahren implementiert, welches sich seitdem als ein weitgehend transparentes und funktionierendes System im Gesundheitswesen etabliert hat. Die tragende Rolle im deutschen HTA-Prozess obliegt dem GBA, welcher in der Regel das IQWiG mit der systematischen Bewertung neuer Arzneimittel beauftragt und darauf basierend Beschlüsse zum festgestellten (Zu-

---

<sup>77</sup> Die dargelegten Hintergründe stellen lediglich Annahmen dar, die der Autor aus den Resultaten dieser Arbeit sowie vergleichbarer empirischer Studien schließt und die demzufolge (zusätzlicher) belastbarer Evidenz und zugleich wissenschaftlicher Debatte bedürfen.

satz-)Nutzen (gegenüber der jeweiligen ZVT) fasst. Über den Bearbeitungszeitraum der vorliegenden Analysen lässt sich eine deutliche Zunahme der durchschnittlichen Anzahl durchgeführter FNBs feststellen, jedoch weder eine eindeutige Tendenz zu einer Ab- beziehungsweise Zunahme der positiven FNB-Resultate noch signifikante Abweichungen zwischen den einzelnen Jahren beim Vergleich der GBA-Beschlussfassungen und IQWiG-Dossierbewertungen.

Die Ergebnisse zu den veröffentlichten Beschlüssen und Bewertungen legen nahe, dass der GBA den Empfehlungen durch IQWiG weitgehend folgt. Unterschiede in den Resultaten scheinen demnach primär auf die Verfahrensanwendung respektive Berücksichtigung der offiziellen Evaluationskriterien zurückzuführen zu sein. Ein strikt evidenzbasierter HTA-Ansatz, in dessen Fokus die Evaluation der klinischen Wirksamkeit steht, stellt dabei die Grundlage für die FNB im deutschen Gesundheitssystem dar. Während die patientenrelevanten Endpunkte sowohl beim GBA als auch bei IQWiG signifikante Determinanten darstellen, scheint die zugrunde liegende Studienevidenz insbesondere Einfluss auf die Dossierbewertungen durch IQWiG zu haben. Nach der aktuell gültigen Version des Methodenpapiers (Allgemeine Methoden Version 6.0, November 2020) präferiert IQWiG systematische Reviews basierend auf RCTs oder einzelne RCTs, was sich mit den Resultaten der vorliegenden Arbeit eindeutig bestätigen lässt. Bestehorn *and* Tunder (2013) stellten bereits zwei Jahre nach Einführung des AMNOG fest, dass direkte Vergleichsstudien inklusive der Anforderungen an die ZVT von zentraler Bedeutung für eine positive Bewertung durch IQWiG sind. Demgegenüber scheint sich der GBA weniger strikt an der vorgelegten Studienevidenz zu orientieren, weshalb GBA-Beschlussfassungen insbesondere auch bei alternativen Studiendesigns, die in den pU-Dossiers verwendet wurden, positiver ausfallen als die Empfehlungen durch IQWiG (Bestehorn *and* Tunder, 2013; Fischer *and* Stargardt, 2014; Schaefer *and* Schlander, 2019).

Die jeweilige Gewichtung beziehungsweise Priorisierung der einzelnen patientenrelevanten Endpunkte scheint für die GBA-Beschlüsse nicht eindeutig, auch wenn Nachweise zur Mortalität und Morbidität wesentlich häufiger herangezogen werden als HRQoL-Daten (Bullinger *et al.*, 2015; Ruof *et al.*, 2014). Demzufolge sind die beiden Endpunkte für die Dossierbewertungen durch IQWiG von hoher Relevanz, wobei Morbiditätsdaten im Vergleich zur Beschlussfassung einer strengeren Evaluation unterliegen zu scheinen (Fischer *and* Stargardt, 2014; Herpers *and* Dintsios, 2018; Skipka *et al.*, 2016). Lebensqualitätsdaten werden in etwa drei Viertel aller durch die pUs eingereichten Dossiers angegeben, aufgrund der Auswahl des zugrunde liegenden Instruments (zum Beispiel der standardisierte, präferenzbasierte EQ-5D Lebensqualitätsfragebogen mit fünf unterschiedlichen Dimensionen) jedoch nicht durchgängig anerkannt,

da indikationsspezifische Messinstrumente aufgrund ihrer höheren Sensitivität bevorzugt werden (Blome et al., 2017; Bullinger et al., 2015). Sofern keine signifikante Nachweise auf Grundlage der patientenrelevanten Endpunkte vorliegen, begründen sich die – lediglich in seltenen Fällen vorliegenden FNB-Ergebnisse mit einem positiven Zusatznutzen – durch spezifische Daten zu möglichen Nebenwirkungen. In der Regel handelt es sich dabei um Daten zu (schweren) unerwünschten Ereignissen (zumeist einhergehend mit der Messung der Lebensqualität), die ein neuer Wirkstoff gegenüber der ZVT entweder nicht oder in sehr viel schwächerer Form aufweist (Herpers and Dintisios, 2018). Während der GBA bei einer nicht ausreichenden Datengrundlage zu patientenrelevanten Outcomes (auch) Surrogatparameter als Endpunkte zulässt, lehnt IQWiG einen Zusatznutzennachweis (primär) anhand von Surrogatendpunkten zumeist ab, da diese häufig nicht ausreichend oder nur unvollständig validiert sind (Bestehorn and Tunder, 2013; Höhle-Pasques et al., 2013).

Darüber hinaus lassen sich bestehende Unterschiede in den Resultaten von GBA und IQWiG in der Festlegung der ZVT, der Klassifikation und Einstufung von Nebenwirkungen sowie in der Definition der Teilpopulationen innerhalb einer spezifischen Indikation begründen (Broich et al., 2016; Ruof et al., 2014). Insbesondere letzteres stellt einen vieldiskutierten Themenpunkt dar, da die Definition beziehungsweise Unterteilung von Patientenpopulationen innerhalb einer zu evaluierenden Intervention ebenfalls einen Effekt auf die Outcomes zu haben scheint. Wie in nahezu allen nationalen HTA-Prozessen werden Patientenpopulationen einzelner Interventionen von GBA/IQWiG in Teilpopulationen unterteilt, um möglichst genau evaluieren zu können, ob einzelne Patientengruppen (insbesondere auch im Vergleich zu anderen Subgruppen innerhalb einer definierten Patientenpopulation) von einem neuen Wirkstoff profitieren können oder nicht (Kaiser et al., 2015; Ruof et al., 2014). Jedoch unterliegen solche Teilpopulationen immer wieder eingeschränkten Fallzahlen, wodurch zugrunde liegende Studienergebnisse ungenauer sein können (zum Beispiel aufgrund fehlender Nachweise für die Akzeptanz patientenrelevanter Endpunkte innerhalb einer bestimmten Subgruppe) und der möglicherweise vorhandene Zusatznutzen einzelner Patientengruppen damit tendenziell eher herabgestuft wird (Bestehorn and Tunder, 2013; Höhle-Pasques et al., 2013; Rychlik, 2013). Zusätzlich kommen allgemeinmethodische Herausforderungen in der Analyse von Patienten(sub)gruppen hinzu, die in den Patientenzahlen zu Abweichungen zwischen GBA-Beschlüssen und IQWiG-Bewertungen führen können (Bestehorn and Tunder, 2013; Kaiser et al., 2015).

Infolgedessen stellt sich auch die Frage, welchen Einfluss die Größe der Patientenpopulation auf die FNB-Ergebnisse von GBA und IQWiG möglicherweise hat. In der vorliegenden Analyse konnte zunächst aufgezeigt werden, dass die Angaben von GBA und IQWiG zur Größe



der Patientengruppen weitgehend vergleichbar sind. In Übereinstimmung mit Lauenroth *and* Stargardt (2017) lässt sich sowohl für die GBA-Beschlussfassungen als auch die IQWiG-Dossierbewertungen erkennen, dass Teilpopulationen oder Patientenpopulationen (auf Verfahrensebene) mit einem positiven Zusatznutzen durchschnittlich kleiner sind als Patientengruppen in FNBs ohne festgestellten Zusatznutzen (Lauenroth *and* Stargardt, 2017). Dies wiederum bestätigen im Wesentlichen auch die Resultate von Höhle-Pasques *et al.* (2013), die für die therapeutischen Gebiete mit im Durchschnitt größeren Zielpopulationen durchweg negativere FNB-Outcomes beobachteten als für Indikationsgebiete mit kleineren Patientenpopulationen.

Darüber hinaus lässt sich eindeutig feststellen, dass den FNBs auch entsprechende indikations-spezifische Unterschiede zugrunde liegen. Während beispielsweise Behandlungen von Stoffwechselerkrankungen wie Diabetes Typ 2 überwiegend negativ bewertet sind, weisen Therapien von Hauterkrankungen wie zum Beispiel der Psoriasis sehr viel häufiger einen positiven Zusatznutzen aus. Die wesentlichen Differenzen zwischen den therapeutischen Gebieten scheinen nicht ausschließlich auf einen bestimmten Grund zurückzuführen zu sein, dennoch ist allen voran die zugrunde liegende Evidenz in den pU-Dossiers explizit herauszustellen. Exemplarisch zeigt sich bei Betrachtung der Stoffwechselerkrankungen, dass Daten zur Lebensqualität häufig nur sehr lückenhaft erfasst sind. Demgegenüber weisen die pU-Dossiers der Hautbehandlungen in aller Regel ausführliche indikationsspezifische Messdaten auf, die insbesondere bei unzureichenden Mortalitäts- und / oder Morbiditätsnachweisen von erheblicher Relevanz für die FNB zu sein scheinen (Greiner *et al.*, 2019).

Vergleichbares gilt auch für Onkologika, die im Vergleich zu den nicht-onkologischen Therapien insgesamt besser bewertet sind. Für sie wie für alle anderen Erkrankungen mit ungünstigen Prognosen gilt, dass eine krankheitsspezifische Messung des Endpunktes HRQoL (mittels validierter Instrumente wie EORTC Core Quality of Life Questionnaire, QLQ-C30, oder Functional Assessment of Cancer Therapy-General, FACT-G) von gewichtiger Bedeutung ist (Lueckett *et al.*, 2011). Das Vorliegen standardisierter Instrumente sowie ausdrückliche Empfehlungen international anerkannter Organisationen begründen demnach die vergleichsweise bessere Datenbasis in pU-Dossiers onkologischer Therapien. Hinzu kommen in der Regel Daten zur Morbidität, die sich auf Nachweisen zum progressionsfreien Überleben oder, sofern verfügbar, zum Gesamtüberleben begründen (Dabisch *et al.*, 2014; Greiner *et al.*, 2019). Ein weiterer Aspekt für die positiveren Beschlussfassungen respektive Bewertungen zeigt sich auch darin, dass der Anteil an Krebstherapien mit einem OD-Status wesentlich höher ist als bei den nicht-onkologischen Therapien (Greiner *et al.*, 2019). Dies wiederum unterstreicht auch die Auffassungen

aus den zuvor diskutierten Untersuchungsergebnisse, dass Indikationen mit kleineren Patientengruppen (wie unter anderem bei Krebserkrankungen) tendenziell besser bewertet werden als solche mit größeren Populationen (wie unter anderem bei Stoffwechsel- oder Augenerkrankungen).

Im Vergleich dazu zeigen die Jahrestherapiekosten pro Patient bei der Gegenüberstellung der GBA-Beschlüsse und IQWiG-Bewertungen ein weniger stringentes Muster, auch weil die zugrunde liegenden Kosten der ZVT zusätzlich zu berücksichtigen gewesen sind. Im Gegensatz zu den vorliegenden Analyseergebnissen fanden Fischer *and* Stargardt (2014) für die GBA-Beschlüsse höhere Kosten pro Patient bei Therapien mit einem positiven Zusatznutzen im Vergleich zu Behandlungen ohne einen festgestellten Zusatznutzen. Gründe hierfür scheinen neben dem erheblich kürzeren Untersuchungszeitraum das Fehlen vergleichbarer Daten für die Bewertungen durch IQWiG und den Vergleichstherapien der Wirkstoffe ohne einen OD-Status zu sein. Dintsios et al. (2019) konnten wiederum aufzeigen, dass die Kosten der ZVT bei Zunahme einen signifikant negativen Einfluss auf die Entscheidungen des GBA im Vergleich zu den Empfehlungen durch IQWiG (mit ergänzendem Addenda) aufweisen. Für den auf die Beschlussfassung folgenden Verfahrensschritt der Preisfestlegung beziehungsweise -Verhandlung zeigt sich zudem ein eindeutiger Zusammenhang zum Ausmaß des Zusatznutzens. Demnach weisen die Therapien ohne einen festgestellten Zusatznutzen durchgängig Preise auf, die unter der jeweiligen ZVT liegen (Gandjour et al., 2020; Hammerschmidt, 2017; Lauenroth and Stargardt, 2017). Hinter dem Zusammenhang von (positiven) FNB-Resultaten und OD-Preisen steht (weiterhin) ein unverkennbares Fragezeichen, auch wenn (zumindest) anzunehmen ist, dass die 'Seltenheit' einer Erkrankung in der Preisgestaltung der einzelnen Therapie durch den OD-Status reflektiert wird (Lauenroth and Stargardt, 2017). Die Bedeutung der Therapiekosten und Patientenpopulationen sowie die daraus resultierende Verknüpfung zum (potenziellen) BI werden in Kapitel 6.2.1 nochmals gesondert aufgegriffen.

Als weitere verfahrensrelevante Faktoren respektive Prozessmerkmale, die einen unmittelbaren Einfluss auf die FNB durch GBA und IQWiG aufweisen, sind die Einbindung von Stakeholdern während der offiziellen Anhörung durch den GBA sowie (mögliche) nachträgliche Analysen durch das IQWiG – in Form ergänzender Addenda zu den durchgeführten Nutzenbewertungen – anzuführen (Dintsios et al., 2019; Fischer and Stargardt, 2014). Insbesondere die Ergänzungen durch IQWiG für die Bewertung nachträglich eingereicherter Evidenz durch die pUs, die in aller Regel aus den Beratungen resultieren, scheinen einen nicht zu unterschätzenden Effekt auf die Entscheidung durch den GBA zu haben (Dintsios et al., 2019).

In nahezu identischer Form lassen sich auch die bestehenden Unterschiede erklären, die beim Vergleich der GBA-Beschlussfassungen oder der IQWiG-Bewertungen mit den eingereichten pU-Dossiers festzustellen sind. Sowohl die Festlegung der ZVT als auch der Patientengruppen sowie die Berücksichtigung (und die damit erforderliche Akzeptanz) patientenrelevanter Endpunkte und Nebenwirkungen haben direkten Einfluss auf die FNB-Outcomes durch GBA und IQWiG (Broich et al., 2016; Ruof et al., 2014). Augenscheinlich fallen die in den pU-Dossiers beanspruchten Nutzenergebnisse eines neuen Wirkstoffs tendenziell besser aus als in den anschließenden Bewertungen respektive Beschlüssen, weshalb die Abweichungen in den Outcomes wesentlich größer sind als zwischen den GBA-Entscheidungen und IQWiG-Empfehlungen.

Letztlich bleibt nochmals hervorzuheben, dass die GBA-Beschlussfassungen gegenüber den IQWiG-Bewertungen insgesamt positiver ausfallen, auch wenn immer eine 'Fall-zu-Fall-Betrachtung' vorgenommen werden sollte, damit möglichst spezifische Aussagen bezüglich einzelner FNB-Resultate getroffen werden können. Dennoch zeigt sich in den FNB-Ergebnissen die grundlegende Tendenz, dass der GBA als eine Art 'Korrektiv' zu agieren scheint, indem er auf Grundlage der streng evidenzbasierten Bewertungen durch IQWiG 'moderatere' Beschlüsse trifft, die offensichtlich auch weitere Faktoren mitberücksichtigen (vgl. hierzu bei Dintsios et al., 2019; Fischer and Stargardt, 2014; Peinemann et al., 2019; Ruof et al., 2014; Schaefer and Schlander, 2019).

### **6.1.3 NICE STA Guidance**

Die Gründung von NICE im Jahr 1999 ist im Wesentlichen auf die Notwendigkeit zurückzuführen, innerhalb des umfassend steuerfinanzierten NHS England systematische und zugleich evidenzbasierte (Priorisierungs-)Entscheidungen zwischen unterschiedlichen Bedürfnissen der Bevölkerung zu treffen, sodass die zur Verfügung stehenden Ressourcen bei begrenztem Budget bestmöglich allokiert werden können (Littlejohns et al., 2019; Timmins, 2016). NICE TAs, die zuerst ausschließlich als MTAs, seit 2006 auch beziehungsweise vielmehr überwiegend in Form von STAs durchgeführt werden, stellen die zentralen Bewertungsprozesse auf Grundlage der aktuellsten Version des 'Methods TA Process Guide' dar (NICE, 2018). Als primäres Ziel gilt dabei, TA-Empfehlungen auf einem gerechten und zugleich nachvollziehbaren Prozess aufzubauen, indem neue sowie auch bestehende Gesundheitstechnologien innerhalb eines gesamtheitlichen, sozial-ethischen Frameworks evaluiert werden, welches sich aus zwei wesentlichen Elementen zusammensetzt (Charlton, 2020; Littlejohns et al., 2019; 2009).

Einerseits liegt ein prozessualer respektive verfahrensorientierter Teil vor, der auf Grundlage des Konzepts 'Accountability for Reasonableness' (A4R<sup>78</sup>; vgl. Daniels and Sabin, 1997) einen gerechten Ablauf des gesamten Evaluationsprozesses durch NICE gewährleisten soll. NICE 'Social Value Judgements' umfassen dabei die zugrunde liegenden Prinzipien im Zusammenhang mit der Entwicklung und Umsetzung der TA Guidance (NICE, 2008). Sie befassen sich mit den gesellschaftlichen Aspekten der Gesundheitsversorgung innerhalb des NHS und werden an anderer Stelle des Unterkapitels – unter expliziter Bezugnahme auf die zugrunde liegenden Bewertungskriterien – betrachtet.

Der TA-Prozess wird bis heute – im Sinne eines standardisierten und robusten Verfahrens – immer wieder als eine Art Rollenmodell für andere nationale HTA-Institutionen und deren methodisches Vorgehen herangezogen. Dabei zeigt sich unter Heranziehung der A4R-Bedingungen nach Daniels *and* Sabin (1997), dass auch bei NICE Gestaltungsspielräume für Prozessoptimierungen bestehen. Während der Prozess als strukturiert, nachvollziehbar und gut dokumentiert gilt, offenbaren sich an einzelnen Stellen weiterhin Transparenzdefizite ('Publicity'), wie zum Beispiel bei den unkenntlich gemachten Textstellen in den Unterlagen der pUs ('Company Evidence Submissions') sowie der mitunter nur bedingt informativen Sitzungsprotokolle ('TA Committee Meeting Minutes'). Das 'Appeal' wiederum erscheint in seiner praktischen Umsetzung zwar restriktiv, stellt dabei aber einen festen Bestandteil jedes einzelnen Verfahrens dar. Die TA Guidance umfasst alle einzelnen Evaluationschritte und bildet demnach eine vollumfängliche Grundlage der TAs, jedoch mangelt es an einem effektiven Qualitätssicherungssystem für die durchgeführten Bewertungen ('Enforcement'; Schlander, 2008).

Andererseits handelt es sich um ein inhaltliches Konzept, welches in Anlehnung an die A4R die Voraussetzung der Relevanz ('Relevance') respektive Angemessenheit ('Reasonableness') der Bewertungsempfehlungen durch NICE spezifiziert. Die sogenannten 'Scientific Value Judgements' (Rawlins et al., 2010) oder vielmehr 'Ethics of Opportunity Costs' (Rid et al., 2015, Rumbold et al., 2017) leiten aus der zur Verfügung stehenden Evidenz ab, wie die vorhandenen Ressourcen unter Berücksichtigung der Allokationseffizienz verteilt werden sollen.

Letzteres ist von besonderem Interesse, da die NICE TAs im Rahmen dieses Konstrukts von bestimmten Kriterien – allen voran der Kosteneffektivität – unmittelbar beeinflusst werden.

---

<sup>78</sup> A4R nach Daniels *and* Sabin (1997) beschreibt einen ethischen Ansatz für einen legitimen und gerechten Entscheidungs- beziehungsweise Priorisierungsprozess (im Gesundheitswesen), der vier Bedingungen voraussetzt: 1. Öffentlicher, transparenter Zugang zum Prozess ('Publicity'), 2. Kontextbezogene Relevanz und Angemessenheit der Entscheidung ('Relevance', 'Reasonableness'), 3. Möglichkeit, Entscheidungen anzufechten oder zu revidieren ('Appeal', 'Revision'), 4. Durchsetzung der drei ersten Prinzipien ('Enforcement').

Darüber hinaus sollen jedoch auch umfassendere Aspekte eine ergänzende Rolle im Entscheidungsprozess einnehmen, darunter zum Beispiel erweiternde soziale und ethische Wertkomponenten (Charlton, 2020; de Folter et al., 2018). Die vorliegenden Analysen, die den Einfluss der (offiziellen oder empirisch als relevant eingestuften) Faktoren auf die STAs (inklusive der FTAs) ausführlich untersuchten, konnten dabei retrospektiv aufzeigen, dass NICE seinen Entscheidungskriterien weitgehend zu folgen scheint.

Zunächst zeigen die Untersuchungsergebnisse eindeutig, dass sich die Empfehlungen oder Ablehnungen neuer Gesundheitstechnologien durch NICE primär durch das Kriterium der Kostenwirksamkeit erklären lassen. Zwar legen einige der im vorliegenden Studienzeitraum identifizierten STAs nahe, dass das Kosteneffektivitätsverhältnis nicht als alleiniges Bewertungsrespektive Entscheidungskriterium betrachtet werden kann, gleichwohl aber als das wichtigste (vgl. hierzu auch Charlton, 2022). Insbesondere die Wahrscheinlichkeit einer Ablehnung durch NICE erhöht sich bei einem ICER  $> 30.000$  GBP/QALY signifikant. Bei der Betrachtung weiterer Kriterien, insbesondere für die Therapien mit einem ICER über dem Schwellenwert von  $20.000 - 30.000$  GBP/QALY, wird jedoch auch die zunehmende Komplexität innerhalb des Entscheidungsprozesses von NICE deutlich erkennbar.

In den letzten beiden Dekaden ist eine Vielzahl empirischer Studien (Cerri et al., 2014; Dakin et al., 2015; 2006; Devlin and Parkin, 2004; Tappenden et al., 2007; Walton et al., 2019) mit teils unterschiedlichen Schwerpunkten zum NICE TA-Prozess veröffentlicht worden. Im Mittelpunkt dieser Arbeiten steht mit dem ICER in aller Regel die zentrale Bewertungskomponente, welche sich in den verschiedenen Analysen auch durchweg als „die“ signifikante Entscheidungsvariable für die NICE TA Guidance darstellt. Vergleichbar zur vorliegenden Untersuchung zeigt sich dabei, dass die durchschnittliche ICER ( $45.700$  GBP/QALY) bei Ablehnung einer Therapie deutlich über dem Schwellenwert<sup>79</sup> von  $20.000 - 30.000$  GBP/QALY zu liegen scheint – mit leicht ansteigender Tendenz (Claxton et al., 2015; Dakin et al., 2015; Devlin and Parkin, 2004; Schaefer and Schlander, 2019). Dies scheint vor allem darauf zurückzuführen zu sein, dass NICE bei Technologien mit einem ICER  $> 30.000$  GBP/QALY ergänzende Faktoren für die Bewertung berücksichtigt, deren konkreter Einfluss (jedoch) nicht immer eindeutig nachzuweisen ist.

---

<sup>79</sup> Die vorliegende Analyse schätzt den ICER-Grenzbereich, ab dem die Wahrscheinlichkeit von einer positiven respektive positiv eingeschränkten zu einer negativen Empfehlung durch NICE wechselt, auf  $41.000 - 51.000$  GBP/QALY. Der Wertebereich liegt in einem weitgehend vergleichbaren Rahmen wie bei Dakin et al. (2015) mit  $39.000 - 44.000$  GBP/QALY sowie Schaefer and Schlander (2019) mit  $41.000 - 49.000$  GBP/QALY.

Während Devlin *and* Parkin (2004) dabei den Effekt der Unsicherheit bezüglich des plausibelsten ICERs und die Krankheitslast anführen, stellen Dakin et al. (2006) insbesondere die Anzahl der zur Bewertung vorliegenden RCTs oder systematischen Reviews als relevantes Einflusskriterium heraus. Interessanterweise konnten Dakin et al. (2006) keinen Effekt für die Relevanz oder Qualität der RCTs feststellen, was sich im Kern mit den Ergebnissen von Dakin et al. (2015) sowie auch den in dieser Arbeit vorliegenden Analyseresultaten deckt. Tappenden et al. (2007) ergänzten die Resultate von Devlin *and* Parkin (2004) um die Faktoren Lebensqualität innerhalb bestimmter Subgruppen sowie das Vorhandensein vergleichbarer Therapien, die innerhalb des NHS bereits zur Verfügung stehen.

In aktuelleren Veröffentlichungen zeigt sich demzufolge ein sehr ähnliches Bild, wobei unter anderem Cerri et al. (2014) die zunehmende Bedeutung sozio-ökonomischer Faktoren hervorheben. Zudem konnten die Autorinnen und Autoren feststellen, dass im Laufe der Zeit tendenziell mehr restriktive und negative als (uneingeschränkt) positive Bewertungen durch NICE vorzufinden sind (Cerri et al., 2014). Dakin et al. (2015) führen in ihrer Analyse wiederum an, dass Therapien, die die EoL-Kriterien erfüllen, mit erheblich höherer Wahrscheinlichkeit positiv bewertet werden als solche, die die Kriterien nicht erfüllen. Diese Erkenntnis unterstreichen die Ergebnisse der vorliegenden Untersuchungen, die einen signifikant Effekt auf die positiven Empfehlungen durch NICE feststellen konnten, wenn die EoL-Kriterien als erfüllt gelten.

Die Implementierung der EoL-Kriterien, die auf eine Zunahme negativer Bewertungsergebnisse von Therapien für (vor allem fortgeschrittene) Krebserkrankungen und einer dadurch ausgelösten gesellschaftlichen Debatte über die (stark) kostenbasierte Bewertung von Gesundheitstechnologien zurückzuführen ist, wird in der einschlägigen Literatur auch weiterhin ausführlich untersucht (Bovenberg et al., 2021; Chalkidou, 2012; Mason and Drummond, 2009). Dabei handelt es sich um den ersten nennenswerten Eingriff in den bis dahin formal standardisierten Evaluationsprozess und damit der 'NICE Social Value Judgements', die mit den TA Guidance im Grundsatz dem Prinzip '*a QALY is a QALY is a QALY [regardless who gains or loses it]*'<sup>80</sup> folgten. Mit der ergänzenden Regelung rückte NICE erstmalig von seiner 'Equity'-Position ab, jedem QALY im Grundsatz den gleichen Wert beizumessen (vgl. hierzu auch Charlton, 2020). EoL-Therapien erfahren demzufolge eine priorisierte Behandlung, das heißt in erster Linie unheilbare (Krebs-)Erkrankungen in einem sehr späten Stadium oder Endstadium.

---

<sup>80</sup> Die vielzitierte Aussage von Weinstein (1988) umfasst die implizite Annahme des QALY-Ansatzes, dass jedes QALY den gleichen sozialen (oder gesellschaftlichen) Wert hat und das unabhängig davon, wem es zugutekommt.

Der damit zugrundeliegenden Intention, mithilfe der EoL-Kriterien den gesellschaftlichen Nutzen und zugleich die gesellschaftliche Belastung durch schwere Erkrankungen umfassender in die Entscheidungen durch NICE einzubeziehen, fehlt es laut Paulden et al. (2014) jedoch an ausreichend empirischer Evidenz sowie einer standhaften theoretischen Grundlage. Damit einhergehend stellt sich die Frage, ob die Kriterien für EoL-Therapien einem gesellschaftlichen Konsens unterliegen und demzufolge soziale Präferenzen überhaupt weitgehend abbilden. Befragungen und Experimente lassen dabei vermuten, dass sich die Kriterien für EoL-Therapien nur bedingt mit den Präferenzen der Bevölkerung decken. Vielmehr lassen die Erkenntnisse darauf schließen, dass die Einbindung von weiteren Kriterien, wie zum Beispiel der Schweregrad einer Krankheit oder der medizinische Bedarf von Patientinnen und Patienten, die gesellschaftlichen Präferenzen besser und demnach vollständiger umfassen würden (Collins and Latimer, 2013; Linley and Hughes, 2013; Shah et al., 2015).

Darüber hinaus zeigt sich in ihrer praktischen Anwendung, dass bis zum Abschluss des vorliegenden Studienzeitraums ausschließlich Krebsbehandlungen die an EoL-Therapien gestellten Anforderungen erfüllen konnten. Zugleich unterstreichen die Ergebnisse aber auch, dass die Erfüllung der Kriterien nicht (automatisch) eine Empfehlung von (Krebs-)Therapien durch NICE für die Aufnahme in das NHS impliziert. Insbesondere eine inhaltliche Anpassung in der Definition der EoL-Kriterien, die die Erfüllung des Kriteriums der Lebenszeitverlängerung ( $\geq$  drei Monate) bei erhöhter Unsicherheit deutlich erschwert, scheint eine zusätzliche Hürde für innovative (Krebs-)Therapien bei der Bewertung durch NICE darzustellen (vgl. hierzu bei Charlton, 2020).

Seit dem Relaunch des CDF im Juli 2016, der im Gegensatz zu den EoL-Regelungen einen separaten, zweckgebundenen Finanzierungsfond darstellt, können (insbesondere) Krebstherapien (welche die EoL-Kriterien nicht erfüllen) nachträglich erstattet werden (NHS, 2016). Erste empirische Studienergebnisse zeigen auf, dass sich die Empfehlungsrate von Krebstechnologien seither leicht verbessert hat (Dillon and Landells, 2018; Morrell et al., 2018; Sabry-Grant et al., 2019; Walton et al., 2019). Die Erkenntnisse decken sich im Wesentlichen mit den Resultaten der vorliegenden Analyse, die – im Vergleich zu den vor Juli 2016 veröffentlichten STAs – eine ebenfalls zunehmende Anzahl positiv bewerteter Krebstherapien aufweisen, auch wenn keine statistisch signifikanten Veränderungen über den Untersuchungszeitraum hinweg festgestellt werden konnten. Zudem bleibt festzuhalten, dass eine durch den CDF berücksichtigte Therapie nur für einen befristeten Zeitraum von maximal zwei Jahren erstattet wird, bevor eine nochmalige Bewertung durch NICE auf der Grundlage neuer Studienergebnisse über eine

Aufnahme im NHS entscheidet. Vor diesem Hintergrund bedarf es nachhaltiger Evidenz, um die langfristigen Auswirkungen der CDF-Erstattungen besser beurteilen zu können.

Sowohl die EoL-Kriterien (und deren vorrangige Anwendung für spezifische Krebstherapien) als auch die nachträgliche Erstattungsoption durch den CDF weisen dabei auf eine mögliche Bevorzugung onkologischer Interventionen hin. Insbesondere der Nutzen beziehungsweise Wert des CDF sowohl für Patientinnen und Patienten als auch für die Gesellschaft steht immer wieder im Mittelpunkt gesundheitspolitischer Debatten, wenn es um die Rechtfertigung der entstandenen Kosten – mit konkreten Schätzungen von bis zu 1,3 Mrd. GBP für den Zeitraum der Einführung im Jahr 2010 bis zum Relaunch im Jahr 2016 – geht (Aggarwal et al., 2017; Cohen, 2017). Trotz der Modifikation durch den Relaunch des CDF stellt sich darüber hinaus auch weiterhin die Frage, inwieweit sich krebsspezifische HTA-Regelungen, die im Umkehrschluss eine (mögliche) Benachteiligung nicht-onkologischer Therapien implizieren, deuten lassen (Dixon et al., 2016; Morrell et al., 2017).

Beim direkten Vergleich von Therapien seltener Krebsleiden mit ODs nicht-onkologischer Indikationsgebiete lässt sich feststellen, dass letztere wesentlich häufiger von NICE abgelehnt wurden als Onkologika während desselben Zeitraums. Dadurch entstehen wiederum Unsicherheiten bei den betroffenen pUs, die nur begrenzt abschätzen können, ob neue nicht-onkologische ODs die Anforderungen des STA-Prozesses prinzipiell erfüllen (können) und eine Einreichung aus unternehmerischer Perspektive damit als ökonomisch sinnvoll einzustufen ist (Clarke et al., 2021). Dies stellt zudem den empirisch nicht (abschließend) belegten Verdacht in Frage, dass eine (bevorzugte) Finanzierung innovativer Krebstherapien gegenüber Behandlungen, die für nicht-onkologische Erkrankungen indiziert sind, gesellschaftlich tatsächlich präferiert ist (Culyer, 2017; Linley and Hughes, 2013; Morrell et al., 2017).

Darüber hinaus besteht unter Expertinnen und Experten weitgehender Konsens darüber, dass (nicht-onkologische) ODs den standardisierten Prozessen nationaler HTA-Institutionen in aller Regel nur bedingt standhalten können und demzufolge spezifische Verfahrensweisen respektive Methoden (international) entwickelt werden sollten (Brenna et al., 2020; Facey et al., 2014; Gammie et al., 2015; Nicod et al., 2020b; Zimmermann et al., 2021). Zwar besteht in England mit dem HST-Prozess bereits seit 2013 ein eigens für URDs implementiertes Bewertungsschema – da ODs die zugrunde liegenden Kriterien jedoch in aller Regel nicht erfüllen (können), sind sie wiederum nur selten Teil dieses Verfahrens. Das zeigt sich insbesondere auch darin, dass bis Ende 2018 lediglich acht Therapien den HST-Prozess vollständig durchlaufen haben (NICE, 2020).



Im Wesentlichen basieren NICE STA Guidance auf den Reports unabhängiger ERGs, die – bestehend aus mitunter klinischen und gesundheitsökonomischen Expertinnen und Experten – die von den pUs eingereichten Unterlagen systematisch evaluieren. Demnach unterliegt die Bewertung und Beschlussfassung formal ebenfalls zwei voneinander getrennten Verfahrensschritten, wobei die ERG Reports keine ausdrücklichen Empfehlungen ausweisen (vgl. hierzu auch Kaltenthaler et al., 2011). Dabei können die ERG Reports auch ergänzende Analysen enthalten, die nicht ausschließlich auf den pU-Daten beruhen müssen und damit zusätzlich in den Entscheidungsprozess von NICE einfließen (Carroll et al., 2017; Kaltenthaler et al., 2017). Darüber hinaus spielt beim ‘Appraisal’ sowohl die Involvierung von Patientinnen und Patienten als auch von unterschiedlichen Stakeholdern eine nicht minder bedeutende Rolle, wie unter anderem die Möglichkeit eines ‘Appeals’ vor der abschließenden Veröffentlichung der STA Guidance darlegt.

## **6.2 Die (potenzielle) Rolle des Budget Impact bei GBA/IQWiG und NICE**

### **6.2.1 Die ökonomische Komponente in der frühen (Zusatz-)Nutzenbewertung**

Während die jeweiligen HTA-Outcomes von GBA und IQWiG eindeutige Rückschlüsse zu den offiziellen Bewertungskriterien – allen voran der vergleichenden klinischen Wirksamkeit in Form patientenrelevanter Endpunkte – ermöglichen, ist die Untersuchung der potenziellen Rolle des BI als wesentlich komplexer einzustufen. Zum einen stellt der BI weder bei den GBA-Beschlüssen noch den IQWiG-Bewertungen einen offiziellen Evaluationsfaktor dar, weshalb die zugrunde liegenden (Kontroll-)Variablen (Kostendaten und Patientenzahlen) zu berücksichtigen sind (vgl. G-BA, 2020; IQWiG, 2020). Zum anderen liegen den FNBs von GBA und IQWiG zumeist unvollständige Informationen zu den möglichen Kostenfolgen zugrunde, insbesondere bei Patienten(teil)populationen neuer Wirkstoffe, die keinen positiven Zusatznutzen aufweisen. Vielmehr basieren die in den IQWiG-Dossierbewertungen angegebenen BI-Schätzungen auf den Daten beziehungsweise Berechnungen der pUs, die in den Dossiers explizite Angaben zu den gesamten Jahrestherapiekosten einer neuen Therapie ausweisen müssen, damit die zukünftigen Kosten der als relevant einzustufenden GKV-Zielpopulation möglichst genau prognostiziert werden können (IQWiG, 2020).

Als ein potenzieller Einflussfaktor der FNB-Resultate von GBA und IQWiG unterliegt der BI folglich den zur Verfügung stehenden Informationen zu Patienten(teil)populationen und jährli-

chen Behandlungskosten pro Patient, deren Effekt in den durchgeführten Analysen entsprechend kontrolliert wurde. Die Anzahl der möglichen zu behandelnden Patientinnen und Patienten innerhalb der GKV, die für die Bewertung und Beschlussfassung herangezogen werden, scheint dabei nur bedingt Einfluss auf den BI zu haben. Als wesentlich einflussreichere Variable stellte sich die Kostenkomponente heraus, die sowohl zu den GBA-Beschlussfassungen als auch den IQWiG-Bewertungen einen Zusammenhang aufweist. Während Wirkstoffe mit positiven Bewertungsempfehlungen durch IQWiG durchschnittlich höhere Jahrestherapiekosten pro Patient aufweisen (positiv linearer Zusammenhang), zeigen Therapien mit positiven GBA-Beschlüssen im Durchschnitt niedrigere Behandlungskosten pro Patient auf (negativ linearer Zusammenhang).

Auch Fischer *and* Stargardt (2014) sowie Dintsios et al. (2019) berichten über einen möglichen Einfluss der Behandlungskosten auf die FNB-Resultate (vgl. hierzu auch Kapitel 2.4). Während Fischer *and* Stargardt (2014) für Wirkstoffe mit höheren Therapiekosten, die zwischen Januar 2011 und Juni 2013 veröffentlicht wurden, eine höhere Wahrscheinlichkeit für einen positiven Zusatznutzen andeuten, stellen Dintsios et al. (2019) heraus, dass eine Zunahme der Behandlungskosten (je 1.000 EUR) der zugehörigen ZVT einen negativen Effekt auf den jeweiligen GBA-Beschluss hat. Zumindest auf Verfahrensebene weist die vorliegende Untersuchung vergleichbare Ergebnisse auf, wobei die Resultate von Dintsios et al. (2019), welche auf einen vergleichbaren Untersuchungszeitraum (01/2011–12/2017) zurückgehen, sich vorrangig auf Beschlussfassungen stützen, denen eine Bewertung inklusive nachträglicher Addenda zugrunde liegen. Die Erkenntnisse von Fischer *and* Stargardt (2014) entsprechen hingegen den vorliegenden Ergebnissen zu den Dossierbewertungen durch IQWiG, welche in der Studie jedoch nicht gesondert analysiert wurden. Keine der beiden Untersuchungen (Dintsios et al., 2019; Fischer *and* Stargardt, 2014) konnte dabei einen eindeutigen Zusammenhang zwischen den GBA-Beschlussresultaten und dem BI feststellen.

Demgegenüber weisen die vorliegenden Regressionsergebnisse einen Zusammenhang zwischen einem zunehmenden BI und einer positiven Bewertungsempfehlung durch IQWiG auf. Jedoch scheint sich die Wahrscheinlichkeit einer Bewertung ohne positiven Zusatznutzen ab einem bestimmten (nicht eindeutig quantifizierbaren) Schwellenwert zu erhöhen. Dies könnte darauf hindeuten, dass die Kostenfolgen eine (zumindest implizite) Rolle im zweistufigen Verfahrensverlauf einnehmen; während IQWiG nämlich den Prinzipien der EbM 'strikt' zu folgen scheint, fasst der GBA – basierend auf den Nutzenempfehlungen – tendenziell 'moderatere' Beschlüsse, deren zugrunde liegenden Faktoren nicht eindeutig nachzuweisen sind (vgl. hierzu auch bei Ruof et al., 2014).

Im Umkehrschluss könnte der potenzielle BI demzufolge, wenn auch aller Wahrscheinlichkeit nach nur als einer von mehreren Determinanten (wie beispielsweise die mündliche Anhörung der Stakeholder), Einfluss auf die Entscheidungen durch den GBA haben, da Therapien mit einem durchschnittlich höheren BI seltener positive Beschlüsse ausweisen als die vorausgehenden Bewertungen durch IQWiG. Denn im Gegensatz zu den streng evidenzbasierten Bewertungen durch IQWiG dürfte der GBA, der das höchste Beschlussgremium der gemeinsamen Selbstverwaltung im deutschen Gesundheitswesen darstellt und sich unter anderem aus dem GKV-SV (als die zentrale Interessenvertretung der gesetzlichen Kranken- und Pflegekassen) zusammensetzt, gesellschaftliche beziehungsweise soziale Präferenzen in seine versorgungspraktisch relevanten Beschlussfassungen einbeziehen. Dies wiederum würde dann implizieren, dass den Entscheidungen die Zahlungsbereitschaft der Bürgerinnen und Bürger für einzelne Interventionen – das heißt die potenziellen Kostenfolgen geteilt durch die Anzahl der Versicherten in der GKV – zugrunde liegen.

Offensichtlich unterliegt die Interpretation des Einflusses ökonomischer Komponenten damit erheblichen Unsicherheiten und damit gewissermaßen auch (nicht belegbaren) Annahmen. Zwar legen die gesetzlichen Regelungen innerhalb des AMNOG-Prozesses nahe, dass die Berücksichtigung von Kosten und damit auch den Kostenfolgen als ein (praktischer) Bestandteil der Bewertungen vorgesehen sind. Eine formale Kosten-Nutzen-Bewertung<sup>81</sup> durch IQWiG kann – nach expliziter Beauftragung durch den GBA – jedoch allenfalls optional durchgeführt werden, weshalb die ökonomische Komponente noch immer Gegenstand kontroverser Debatten rund um das Thema HTA in Deutschland ist (Gerber-Grote et al., 2014; Höhle-Pasques et al., 2013).

Insbesondere in der praktischen Umsetzung zeigt sich, dass die Anwendung der sogenannten Effizienzgrenzen-Analyse durch IQWiG, die entsprechend um eine Analyse der Kostenfolgen zu ergänzen ist, unter AMNOG bisher (noch) keine Rolle gespielt hat. Gerber-Grote et al. (2014) sehen genau hierin eine verpasste Chance und ein zugleich zunehmendes Problem, wenn eine langfristige Finanzierung medizinischer Leistungen durch die FNB gewährleistet werden soll. Denn der aktuelle AMNOG-Report 2020 zeigt indes eindrücklich, dass die Preise neuer Arzneimittel über die Jahre hinweg kontinuierlich angestiegen sind (Greiner et al., 2020).

---

<sup>81</sup> Bei einer Kosten-Nutzen-Bewertung durch IQWiG werden unterschiedliche Kostenarten berücksichtigt und zugleich verschiedene Kostenperspektiven eingenommen, wobei insbesondere die potenziellen Ausgaben aus Sicht des Kostenträgers, das heißt der GKV, von Interesse sind (IQWiG, 2020).

Vielmehr wird daher die obligatorische Bewertung von Nutzen und Kosten als ein ergänzendes Element der FNB gefordert, um eine adäquate, faire und zugleich nachhaltige Preisbildung sicherzustellen (Kaier and Fetzer, 2015; Windeler and Lange, 2015). Dabei kann die BIA ein mögliches Teilelement einer ganzheitlichen (ökonomischen) Bewertung darstellen. Cassel *and* Ulrich (2015) empfehlen darüber hinaus eine institutionelle Trennung der Nutzenbewertung und der anschließenden Preisverhandlung. Für die Preisverhandlungen wird bereits seit längerem von verschiedensten Stakeholdern eine deutlich höhere Transparenz gefordert, insbesondere wenn – bei vorliegendem positiven Zusatznutzenbeschluss durch den GBA – eine Schiedsstelle über eine gescheiterte Erstattungspreisverhandlung zwischen GKV-SV und pU entscheiden muss (Cassel and Ulrich, 2015).

Das von IQWiG verwendete Konzept der Effizienzgrenze stellt in der von CEAs respektive CUAs dominierten internationalen HTA-Praxis eine ebenfalls vieldiskutierte Ausnahme bei der ökonomischen Bewertung neuer Therapien dar, da sie unter Berücksichtigung des festgestellten Zusatznutzens ausschließlich einen indikationsbezogenen Kosten-Nutzen-Vergleich ermöglicht (Mathes et al., 2013; Riedel et al., 2013). Dabei deutet die (zumindest praktische) Ablehnung eines indikationsübergreifenden Nutzenmaßes – wie zum Beispiel des QALYs – daraufhin, dass sowohl gesundheitspolitische als auch gesellschaftliche (und damit ethische) Beweggründe zugrunde liegen müssen (Caro et al., 2010; Kieslich, 2012). Zudem fehlt es allen voran aus Komplexitätsgründen an einem einzelnen indikationsübergreifenden Outcome-Maß, das alternativ verwendet werden könnte (Beresniak and Dupont, 2016). Gandjour (2013; 2020) sieht in der Effizienzgrenzen-Analyse vielmehr eine plausibel definierte Wertefunktion, die prinzipiell über alle therapeutischen Gebiete hinweg eingesetzt werden kann. Sandmann et al. (2018) weisen darüber hinaus auf die lediglich marginalen Unterschiede zwischen der Effizienzgrenzen-Analyse und der CEA hin und führen die – zumindest in der Theorie bestehende – Möglichkeit einer Zusammenführung von essentiellen Kernelementen der beiden Ansätze an, auch wenn die grundsätzliche Frage nach der praktischen Umsetzbarkeit (bisher) unbeantwortet bleibt (Sandmann et al., 2018; vgl. hierzu auch bei Gandjour, 2020).

In den zurückliegenden zehn Jahren hat sich die FNB durch den GBA als ein 'lernendes System' im deutschen Gesundheitswesen weitgehend etabliert. Nachweislich basieren seit Einführung der FNB nach § 35a SGB V sowohl die Bewertungen als auch Beschlussfassungen auf den Kriterien der EbM, die den Beleg der klinischen Wirksamkeit einer neuen Therapie gegenüber der definierten ZVT auf Grundlage qualitativ möglichst hochwertiger Studienresultate (RCTs als 'Goldstandard') fordert. Jedoch zeigt sich sukzessive, dass die Herausforderungen an eine möglichst ganzheitliche Evaluation von neuen Arzneimitteln stetig zunehmen. Allen

voran sind hier die steigenden Preise innovativer Therapien sowie demgegenüber die bis heute nur bedingt generierten Einsparungen durch die FNB nach AMNOG zu nennen (Greiner et al., 2020; Höer and Berkemeier, 2020). Ein weiteres und darüber hinaus durchaus kontroverses Thema stellen die Evaluationen von ODs dar, die aufgrund der bestehenden OD-Regelung offensichtlich priorisiert behandelt werden (Bouslouk, 2016; Schulz et al., 2018), obwohl die der Bewertung zugrunde liegende Studienevidenz im Vergleich zu Nicht-ODs häufig als schlechter einzustufen ist (Greiner et al., 2020).

Folgerichtig scheinen weitere Adaptionen innerhalb des bestehenden AMNOG-Verfahrens unabweichlich, womit wiederum die Grundsatzfrage in den gesundheitspolitischen Fokus rückt, inwieweit dem Vergleich von Kosten und Nutzen – und demzufolge auch den Kostenfolgen für das deutsche Gesundheitssystem – eine (obligatorische) Rolle bei der FNB zukommen sollte(n).

### **6.2.2 STAs und die Kosteneffektivität als (primärer) Bewertungsfaktor**

Bei Betrachtung der untersuchten NICE Guidance zeigt sich deutlich, dass die ICER/QALY den primären Einflussfaktor der STA Outcomes darstellt. Diese Erkenntnis bestätigt die vorhandenen Resultate aus verschiedenen empirischen Arbeiten (Cerri et al., 2014; Dakin et al., 2015; 2006; Devlin and Parkin, 2004; Tappenden et al., 2007; Walton et al., 2019), auch wenn – wie zuvor bereits ausführlich dargelegt – die Kosteneffektivität einer Technologie nicht als alleiniges Entscheidungskriterium für eine Aufnahme in den NHS England betrachtet werden kann (vgl. hierzu auch Charlton, 2022). Da die NICE Guidance demnach verbindliche Empfehlungen hinsichtlich der Erstattungsentscheidung neuer Therapien darstellen, beeinflussen diese (zumindest indirekt) das Budget, welches innerhalb des NHS England zur Verfügung steht.

Die damit implizierte Allokation der verfügbaren Ressourcen führt wiederum auf einen in der einschlägigen Literatur vielfach diskutierten Themenschwerpunkt zurück, inwiefern die Berücksichtigung der Kosten je QALY als ein (immer noch) zentraler Faktor ausreicht, um nachhaltige Entscheidungen (vorrangig) in einem Beveridge-System<sup>82</sup> treffen zu können (vgl. hierzu bei Devlin and Lorgelly, 2017; Drummond et al., 2009; Garau et al., 2011; Loomes and McKenzie, 1989). Einerseits stellt sich die Frage nach der Verwendung und folglich der (adäquaten)

---

<sup>82</sup> Unter Rekurs auf die Theorie der neoklassischen Wohlfahrtsökonomie kann ein variabler ICER-Schwellenwert unter der Voraussetzung eines restriktiven (oder zumindest festgelegten) Budgets, wie es im NHS England der Fall ist, durchaus geeignet sein. Im Vergleich zu Sozialversicherungssystemen wie in Deutschland erscheint die Verwendung von QALYs damit weitaus praktikabler, da vielmehr eine Maximierung der medizinischen Versorgung mittels der zur Verfügung stehenden Ressourcen zu erreichen ist (vgl. hierzu bei Cleemput et al., 2011).

Höhe des ICER-Schwellenwerts, der unter Einbeziehung ergänzender Kriterien empfehlungsrelevant ist (Appleby et al., 2009; Cleemput et al., 2011; Paulden et al., 2017). Andererseits, inwieweit die Kosteneffektivität als vorrangiges (ökonomisches) Kriterium ausreicht, um innerhalb eines begrenzten Budgets Kostenübernahmeentscheidungen für einzelne Therapien treffen zu können (Claxton et al., 2015; Drummond and Sorenson, 2009; Eichler et al., 2004).

Gafni *and* Birch (2006) zweifeln vielmehr an, ob ein ICER-Schwellenwert überhaupt eine effiziente Ressourcenallokation sicherstellen kann. Darüber hinaus argumentieren die beiden Autoren, dass ICER-Schwellenwerte (zumindest indirekt) zu einem Anstieg der Gesundheitsausgaben beitragen, insbesondere wenn Informationen zu den Opportunitätskosten von Behandlungsinterventionen fehlen beziehungsweise bei der gesundheitspolitischen Entscheidungsfindung unberücksichtigt bleiben (Gafni and Birch, 2006). Daran anknüpfend folgten kontrovers geführte Debatten um die bereits seit vielen Jahren steigenden Kosten im (englischen) Gesundheitswesen und die damit einhergehende Frage, wie eine nachhaltige Finanzierung (‘Affordability’) von innovativen Gesundheitstechnologien innerhalb des NHS England sichergestellt werden kann (Aggarwal and Sullivan, 2014; Ewbank et al., 2018).

Die Kostenfolgen stellen bis dato kein (offizielles) Evaluationskriterium bei NICE dar, das in den STA-Empfehlungen (explizit) Berücksichtigung findet. Jedoch haben bereits mehrere Studien den BI als einen (möglichen) Einflussfaktor der NICE STAs untersucht. Während Devlin *and* Parkin (2004) den BI im Hinblick auf die Empfehlungen durch NICE als nicht aussagekräftig beschreiben, stellen Dakin et al. (2006) einen möglichen Einfluss des potenziellen BI in den Entscheidungen durch NICE heraus. Die Autorinnen und Autoren stellten fest, dass der BI für restriktive Empfehlungen deutlich höher ist als für uneingeschränkt positive Bewertungen, wobei der potenzielle BI negativer Empfehlungen durch NICE nicht weiterführend analysiert wurde (Dakin et al., 2006). Jedoch bleibt anzumerken, dass Dakin et al. (2006) in ihrem multinomialen Regressionsmodell keinen statistisch signifikanten Effekt für die Kostenfolgen feststellen konnten (vgl. hierzu auch Kapitel 2.4).

Erweiternd sind die Resultate einer multivariaten Regressionsanalyse von Mauskopf et al. (2013) hervorzuheben, die insbesondere den Effekt des BI auf die durch NICE veröffentlichten TAs für den Zeitraum Januar 2001 bis März 2011 analysierten. Sowohl für den maximalen als auch den bereinigten BI (für die als Grundlage die sogenannte ‘UK marketing-approved population’ diente) zeigt sich, dass dieser für TAs mit negativen Empfehlungen durchschnittlich höher ist als bei positiven Bewertungen, wobei die uneingeschränkt positiven Empfehlungen niedriger sind als die restriktiven Bewertungen (Mauskopf et al., 2013). Eine vergleichbare

Tendenz der BI-Schätzwerte zeigt sich auch in der Studie von Cerri et al. (2014), die die deskriptiven Resultate zum BI aber nicht weiter in ihrer statistischen Analyse zu den zwischen 2004 und 2009 veröffentlichten TAs berücksichtigten.

Die statistisch signifikanten Zusammenhänge zwischen den potenziellen BI-Szenarien und den Empfehlungen durch NICE aus der Arbeit von Mauskopf et al. (2013) konnten durch die vorliegenden Analyseergebnisse, die primär STAs (inklusive einzelner FTAs) aus dem daran anschließenden Zeitraum (Januar 2011 bis Juni 2018) berücksichtigten, weitgehend bestätigt werden. Darüber hinaus zeigt sich in der vorliegenden Arbeit, dass die Bewertungen durch NICE von Technologien bis zu einem bestimmten BI-Schätzbereich in der Regel positiv ausfallen und demzufolge die Wahrscheinlichkeit einer Empfehlung für Interventionen mit einem oberhalb dieses 'Schwellenwerts' liegenden BI abnimmt. Dieser Effekt bestätigt sich prinzipiell auch bei gesonderter Berücksichtigung des CDF (ab Juli 2016), wobei insbesondere die Bewertungen von Krebstherapien mit vergleichsweise hohen Kosten beziehungsweise vielmehr mit verhältnismäßig hohen Kostenfolgen besser auszufallen scheinen als vor dem CDF-Relaunch.

Spätestens mit der Einführung des BI-Tests im April 2017 wird eine zunehmende Bedeutung des potenziellen BI im Evaluationsprozess neuer Technologien durch NICE deutlich erkennbar. Denn dadurch werden die Erstattungspreise neu bewerteter Therapien, die das NHS England innerhalb eines der ersten drei Jahre insgesamt mehr als 20 Mio. GBP kosten, nochmals neu verhandelt, bis eine einvernehmliche Preisvereinbarung (das sogenannte 'Commercial Agreement') zwischen dem NHS und dem pU getroffen wird (NICE, 2017). Andernfalls kann es zu einem zumindest verspäteten Zugang für die Patientinnen und Patienten innerhalb des NHS kommen, und dies unabhängig davon, ob die Therapie bereits als kosteneffektiv – gegebenenfalls von sozialen und / oder ethischen Werten unterstützt – eingestuft wurde. Kritische Stimmen sehen genau darin eine zusätzliche 'Hürde' für den Marktzugang innovativer Gesundheitstechnologien, die zugleich die eigentliche Kompetenz beziehungsweise das vorgesehene Kerngebiet der NICE TAs überschreitet (Charlton et al., 2017; Edwards et al., 2019).

Die praktische Umsetzung stellt sich – bis Juni 2018 – dementsprechend dar, dass alle der potenziell betroffenen Technologien aufgrund von Preis(nach)verhandlungen innerhalb des NHS zugänglich gemacht wurden. Inwiefern der BI-Test zu einem verspäteten Zugang für Patientinnen und Patienten geführt haben könnte, lässt sich anhand der vorliegenden Daten jedoch nicht eindeutig feststellen. Die erste Technologie, die nach nochmaliger Preisvereinbarung zwischen dem NHS und dem pU weiterhin über den jährlichen Gesamtkosten von maximal 20 Mio. GBP lag, war das Cochlea-Implantat im März 2019 (TA566; Hörprothese für gehörlose und ertaubte

Personen). Aufgrund einer nachträglichen Anpassung der NICE Guidance hinsichtlich der erstattungsfähigen NHS-Zielpopulation wurde das Implantat jedoch für eine routinemäßige Erstattung aufgenommen (Timmins, 2019).

Von unterschiedlichen Stakeholdern geäußerte Bedenken, dass der neue BI-Grenzwert zu niedrig angesetzt sei und dies negative Auswirkungen auf die Erstattung innovativer Technologien zur Folge haben könnte, sind (bisher) nicht auszumachen (vgl. hierzu auch bei Ogden, 2017). Dennoch bleibt nicht auszuschließen, dass der BI-Test nachhaltigen Einfluss auf die Erstattungsentscheidungen haben könnte, sei es einerseits in Form finanzieller Einsparungen aus Sicht des NHS oder andererseits durch verspätet zugängliche Behandlungsoptionen, die insbesondere die Patientinnen und Patienten direkt betreffen würden.

Die vorliegenden Analyseergebnisse zeigen – unterstützt von den durch Mauskopf et al. (2013) veröffentlichten Resultaten, dass der BI – und das nicht erst seit Einführung des BI-Tests im April 2017 – einen Effekt auf die STA-Empfehlungen durch NICE zu haben scheint. NICE, das in seinen TAs die Kosteneffektivitäts-Daten unter Hinzuziehung sozialer und ethischer Komponenten priorisiert, wird weiterhin Modifikationen, Erweiterungen und Neuerungen in seinen Methoden und den zugrunde liegenden 'Frameworks' vorzunehmen haben, um einen robusten sowie transparenten und damit zugleich gerechten Entscheidungsprozess sicherstellen zu können (Charlton, 2020; 2022; Kieslich, 2020; Rid et al., 2015; Rumbold et al., 2017).

Dass dabei auch die Kostenfolgen eine zumindest ergänzende Rolle spielen werden, erscheint aufgrund der zurückliegenden Entwicklungen folgerichtig – denn bereits heute ist nicht mehr von der Hand zu weisen, dass der potenzielle BI innerhalb des aktuellen Entscheidungsprozesses an Bedeutung gewonnen hat. Schlussendlich stellt der Zugang zu innovativen und zugleich (sowohl klinisch als auch ökonomisch) effektiven Therapien sowie deren Bezahlbarkeit die zentrale Herausforderung für den NHS England dar, die es aktuell – gerade aber auch zukünftig – sicherzustellen gilt.



### 6.3 Nationale HTA-Institutionen und -Verfahren – quo vadis?

In nationalen Gesundheitssystemen werden regelmäßig Ressourcenallokationsentscheidungen getroffen, die eine Abwägung von potenziellen Gesundheitsgewinnen für unterschiedliche Patientenpopulationen implizieren (Stafinski et al., 2011b). Als ein ganzheitlicher Ansatz verfolgt HTA dabei das Ziel, solche Entscheidungen mittels systematischer und zugleich evidenzbasierter Evaluationen zu unterstützen, wobei seine praktische Umsetzung Diskrepanzen sowohl auf gesundheitspolitischer als auch auf institutionell-methodischer Ebene offenlegt (Akehurst et al., 2017; Angelis et al., 2018; Kleijnen et al., 2012).

Einerseits weisen die Resultate der vergleichenden Untersuchung von GBA, HAS und NICE darauf hin, dass unterschiedliche Methoden zu signifikanten Unterschieden in den HTA-Outcomes führen. Andererseits bestätigen die Ergebnisse der Regressionsmodelle sowohl für GBA/IQWiG als auch für NICE, dass die HTA-Agenturen ihren offiziellen Bewertungskriterien weitgehend zu folgen scheinen. Darüber hinaus zeigen die bestehenden Unterschiede in den Bewertungsmethoden gemeinsam mit den institutionsspezifischen Einflussfaktoren auf, weshalb die Outcomes respektive Empfehlungen der HTA-Agenturen entweder stärker oder – wie insbesondere bei den identifizierten Krebstherapien deutlich wurde – weniger stark voneinander abweichen.

Während bei der Bewertung der klinischen Effektivität weitgehend Konsens darüber besteht, sich an den Grundsätzen der Kriterien der EbM zu orientieren, zeigt sich insbesondere bei der Integration der ökonomischen Komponente ein durchaus heterogenes Bild innerhalb der bestehenden HTA-Landschaft (Nicod et al., 2019; 2020a; Stafinski et al., 2011b). Zwar hat sich das QALY, wie unter anderem in Australien, England oder den Niederlanden, als ein De-facto-Standard für die Bewertung der Kosteneffektivität durchgesetzt; in einzelnen Ländern, wie beispielsweise Deutschland und Spanien, wird die Verwendung des QALYs und demnach die Berechnung der Kostenwirksamkeit mittels indikationsübergreifenden Nutzenmaßes jedoch weitgehend abgelehnt (vgl. hierzu bei Angelis et al., 2018; Nicod et al., 2020a).

Dies scheint wiederum primär eine Folge der zunehmenden Kritik am QALY(-Ansatz) zu sein, das als gesundheitsökonomisches Konzept eine extra-welfaristische Position einnimmt. Zum einen sind die grundsätzlichen methodischen Probleme anzuführen, die der individuellen Nutzen- respektive QALY-Maximierung, das heißt der Maximierung von Lebensdauer und Lebensqualität, zugrunde liegen (vgl. bei Beresniak et al., 2015). Zum anderen ist die fehlende adäquate Erfassung gesellschaftlicher oder sozialer Präferenzen zu nennen, die gerade auch aus ethischen Gesichtspunkten kritisiert wird (vgl. bei Gu et al., 2015).

Aktuelle Weiterentwicklungen des QALY-Konzepts verdeutlichen zudem, dass seine 'begrenzt' (oder im Sinne der praktischen Umsetzung vielmehr 'eng') definierte Anwendung zu erweitern ist, indem nicht eingeschlossene Komponenten, wie zum Beispiel die Auswirkung auf Autonomie respektive Selbstbestimmtheit, einbezogen und ergänzende Einflussfaktoren, wie beispielsweise soziale Präferenzen, in einer möglichst deliberativen Weise berücksichtigt werden (vgl. bei Brazier and Tsuchiya, 2015; Garrison et al., 2017). Exemplarisch lässt sich unter anderem die Anpassung von ICER-Schwellenwerten an den Schweregrad einer Erkrankung anführen, die sich (zumindest weitgehend) in den NICE STAs von EoL-Therapien widerspiegelt (Collins and Latimer, 2013).

Die Berücksichtigung von BIAs als eine erweiternde ökonomische Komponente – allen voran ergänzend zur Durchführung von CEAs respektive CUAs – erscheint dabei nicht minder kontrovers, wengleich ihre zunehmende Bedeutung im Kontext von HTA nicht von der Hand zu weisen ist (vgl. hierzu bei Chugh et al., 2021). Gerade bei der praktischen Umsetzung des QALY-Konzepts zeigt sich dabei sehr häufig, dass Therapien für seltene und sehr seltene Krankheiten nicht die konventionellen Kosteneffektivitäts-Benchmarks erfüllen können, obwohl sie in aller Regel finanziell tragbar sind und zugleich nur bedingt zum andauernden Kostenwachstum innovativer Arzneimittel beitragen (Schey et al., 2011; Schlander et al., 2018). Bilinski et al. (2017) unterstreichen zudem das zugrunde liegende Potenzial, welches aus der Integration von BIAs und (irgendeiner Form von) CEAs bei der Festlegung von Prioritäten für globale Gesundheitsprogramme resultieren könnte (vgl. hierzu auch Lomas, 2019). Dies gilt umso mehr vor dem Hintergrund, dass einzeln neue Therapien, die als kosteneffektiv eingestuft werden, innerhalb bestehender Budgetrestriktionen kaum noch finanzierbar sind (Bilinski et al., 2017).

Andere Expertinnen und Experten weisen ebenfalls auf eine zunehmende Anzahl von BIAs in der gesundheitsökonomischen Fachliteratur hin, wengleich ihrer Ansicht nach ein Großteil der durchgeführten Analysen nicht dem Anspruch allerhöchster wissenschaftlicher Qualität genügen (Orlewska and Gulácsi, 2009; van der Vooren et al., 2014). Darüber hinaus deuten van de Vooren et al. (2014) an, dass insbesondere von den pUs in Auftrag gegebene Studien die jeweiligen BIAs dementsprechend konzipieren, damit möglichst kurzfristige Einsparungen durch den Einsatz innovativer und zugleich hochpreisiger Medikamente aufgezeigt werden können.

Demgegenüber stellt die von IQWiG verwendete Effizienzgrenzen-Analyse, welche bei Anwendung ebenfalls um eine BIA ergänzt werden soll, ein Methodenkonzept für die Umsetzung

von Kosten-Nutzen-Bewertungen dar, mit dem die (implizite) Zahlungsbereitschaft von medizinischen Leistungen innerhalb eines therapeutischen Gebiets ermittelt werden kann (Caro et al., 2010). Neben der implizierten Schwierigkeit, dass bei einem fehlenden Referenzwert ein Preis als Maßstab aus einem anderen therapeutischen Gebiet herangezogen werden muss, mangelt es bisher aber auch an aussagekräftiger Evidenz zur Relevanz in der bestehenden HTA-Praxis, wie bereits in Kapitel 6.2.1 aufgezeigt werden konnte.

Alternative methodische HTA-Ansätze<sup>83</sup>, die in Fachkreisen (sowie darüber hinaus) diskutiert werden, versuchen für gewöhnlich weitergefasste Faktoren bei der systematischen Evaluation neuer Gesundheitstechnologien zu berücksichtigen. Dabei ist allen voran die 'Multi-Criteria Decision Analysis' (MCDA) zu nennen, die es zum Ziel hat, eine zugrundeliegende Entscheidungsproblematik zu analysieren, um dann möglichst alle potenziell relevanten Faktoren, die den Entscheidungsprozess beeinflussen, systematisch identifizieren und daraufhin evaluieren zu können (Marsh et al., 2016; Thokala et al., 2016). Im theoretischen sowie praktischen Kontext von HTA gewinnt die MCDA sukzessive an Bedeutung, auch wenn ihre Integration in bestehende Strukturen und 'HTA-Frameworks' – oftmals aufgrund der vorhandenen Komplexitätsgründe, weitere Bemühungen erfordert (Angelis and Kanavos, 2017; Baltussen et al., 2019). Ähnliches gilt sowohl für die sogenannte (relative) soziale Zahlungsbereitschaft, die einen primären Fokus auf soziale Transferzahlungen und zugleich eine größere Rolle für die budgetären Auswirkungen von Gesundheitsprogrammen implizieren (vgl. hierzu bei Richardson et al., 2014), sowie für die in formellen HTAs noch (weitgehend) unberücksichtigten 'Fairness-based Evaluation Frameworks' (vgl. hierzu bei Nord, 2015).

Neben den aufgezeigten, primär methodischen Gründen verstärken gesundheitspolitische Regulierungen und zugrunde liegende Erstattungspreisprozesse der einzelnen (europäischen<sup>84</sup>) Länder die bestehenden Unterschiede offizieller HTA-Empfehlungen, die zugleich auch den Rahmen für weitere Differenzen innerhalb des institutionellen Kontexts bilden (Akehrst et al., 2017; Angelis et al., 2018; Kleijnen et al., 2012). Im Gegensatz zur arzneimittelrechtlichen Zulassung, die auf europäischer Ebene seit 31. März 2004 harmonisiert und damit weitgehend zentralisiert werden konnte (vgl. Verordnung (EG) Nr. 726/2004 des Europäischen Parlaments

---

<sup>83</sup> Die angeführten HTA-Ansätze stellen lediglich einen kurzen, exemplarischen Auszug der Bewertungsverfahren dar, die als (potenziell) relevant einzustufen sind. Eine ausführliche Übersicht und Diskussion zu den aktuell vorhandenen HTA-Methoden findet sich unter anderem bei Caro et al. (2019).

<sup>84</sup> England hat die EU formal zum 31. Januar 2020 verlassen ('Brexit'). Dies schließt jedoch nicht aus, dass NICE auch weiterhin eine wichtige Rolle in internationalen Initiativen, Kollaborationen und Netzwerken – auch auf europäischer Ebene – einnehmen kann. Aktuelle politische Entwicklungen auf EU-Ebene werden demzufolge aber keine (direkten) Auswirkungen auf NICE und dessen Arbeitsweise haben.

und des Rates, Amtsblatt Nr. L 136), ist die Bewertung medizinischer Technologien im Sinne eines nachgelagerten HTA ausschließlich nationalen Regelungen und Verfahren unterworfen. Internationale Kollaborationen von nationalen HTA-Institutionen und politischen sowie wissenschaftlichen Einrichtungen existieren bereits seit vielen Jahren. Eine mögliche politische Regulierung auf europäischer Ebene befindet sich aktuell jedoch noch in einem frühen Entwicklungsstadium. Die angekündigte EU-weite Zusammenarbeit im Rahmen von HTA, die im September 2021 mit den Entwürfen von Leitliniendokumenten und Rechtsabhandlungen startete, plant den stufenweisen Aufbau – basierend auf den abgeschlossenen und laufenden EUnetHTA-Projektinitiativen<sup>85</sup> (vgl. hierzu bei Ruether et al., 2022), einem europäischen HTA-Netzwerk und der unverbindlichen Unterstützung der EU-Mitgliedsstaaten – eines vollumfänglichen ‘Joint Clinical Assessment Scope’ für den Zeitraum von 2024 bis 2029 (EU-Kommission, 2021).<sup>86</sup>

Während auf der Nutzenseite offensichtlich die Reduktion von Redundanz sowie Effizienzsteigerungen und Qualitätsverbesserungen in den Bewertungen liegen, gerade auch für Länder mit stark limitierten Ressourcen, weist die Gegenseite insbesondere auf die adäquate Berücksichtigung der unterschiedlichen Prozesse, Methoden und Instrumente hin, die in einigen Ländern den anschließenden Erstattungsprozess (direkt oder indirekt) beeinflussen (Kanavos et al., 2019; Vela Bonanno et al., 2019). Torbica et al. (2018; 2020) unterstreichen zudem die bestehenden Unterschiede sowohl innerhalb des institutionellen Kontexts als auch von kulturellen sowie sozialen respektive gesellschaftlichen Werten in einzelnen Ländern. Während beispielsweise in den methodischen Leitlinien von GBA/IQWiG und NICE – trotz der elementar unterschiedlichen HTA-Ansätze – eine Vielzahl von Überschneidungen vorzufinden sind, weichen gesundheitspolitische und regulatorische Faktoren teils erheblich voneinander ab (Ivandic, 2014; Kieslich, 2020; Wright et al., 2017). Husereau et al. (2016) sehen in den aktuellen Diskussionen und Entwicklungen jedoch zugleich auch die Chance für einen grundlegenden Para-

---

<sup>85</sup> Die grundlegende Basis stellen hierbei EUnetHTA Joint Action (JA) 1 (2010-2012), JA 2 (2012-2015) und JA 3 (2016-2021) sowie aktuell EUnetHTA 21 (2021-2023) dar. Zudem gibt es gemeinsame Initiativen von EUnetHTA und EMA, wie beispielsweise den ‘Early Dialogue’, der für Unternehmen als eine europaweite, frühe Beratung zur (klinischen) Studienplanung im Hinblick auf die Zulassung und anschließende Nutzenbewertung implementiert wurde (EUnetHTA, 2021).

<sup>86</sup> Insbesondere der aktiven Einbindung aller potenziell relevanten Stakeholder kommt dabei eine zentrale Rolle zu, deren Berücksichtigung – zum Beispiel in Form von Erkenntnissen aus einer Multi-Stakeholder-Umfrage im Fachbereich der Hämatologie/Onkologie zu ‘Joint European HTA’ (vgl. hierzu Julian et al., 2022) – eine notwendige Voraussetzung für eine nachhaltige und damit auch möglichst erfolgreiche Entwicklung sowie Implementierung darstellt.

digmenwechsel in der (europäischen) HTA-Landschaft, für den sie neben adaptiven methodischen Verfahren vor allem einen verstärkten wissenschaftlichen Dialog und eine frühe sowie umfassende Involvierung aller relevanten Stakeholder fordern.

Die langfristige Zusammenarbeit von HTA-Agenturen untereinander und zusätzlich mit anderen relevanten Institutionen fördert prinzipiell einen stärker integrierten Entscheidungsprozess und damit eine nachhaltige und zugleich systematische Anpassung des (europäischen) HTA-Umfelds. Die europäischen HTA-Institutionen weisen dabei sowohl aus regulatorischer als auch aus methodisch-technischer Perspektive nachweislich Übereinstimmungen auf, insbesondere bei der Bewertung klinischer Aspekte (wie beispielsweise klinisch-relevanter Endpunkte) sowie einzelner Prozess- und Konsultationskomponenten, die weiter aneinander angepasst werden können (vgl. bei Allen et al., 2017; Chassagnol et al., 2020; Kleijnen et al., 2016; Oortwijn et al., 2017; Tafuri et al., 2016). Ergänzend sollten nationale und internationale HTA-Netzwerke als wertvolle Informationsquellen genutzt werden, um mittels Daten- und Wissensaustausch zwischen den einzelnen Regulierungsinstitutionen eine vollständigere und zugleich verbesserte Evidenzgrundlage für die Durchführung von HTAs zu schaffen (Panteli et al., 2016).

Darüber hinaus sollte die systematische Evaluation von Gesundheitstechnologien durch offizielle HTA-Agenturen von evidenzbasierten Beratungsprozessen unterstützt werden, um deren Legitimität bei Allokationsentscheidungen von begrenzt zur Verfügung stehenden finanziellen Ressourcen zu verbessern (Baltussen et al., 2017; Oortwijn et al., 2020). Unter HTA-Expertinnen und -Experten stößt der vielseitige Ansatz vorwiegend auf Zuspruch, wobei die Implementierung und praktische Umsetzung eines solchen Prozesses immer abhängig ist von den bestehenden nationalen Verfahren und Methoden (Culyer, 2020; Gopinathan et al., 2021; Schlander, 2021). Schließlich steht im Gesundheitswesen eine systematische und transparente Allokation der verfügbaren Ressourcen im Vordergrund, mit dem Ziel eines bestmöglichen Zugangs für Patientinnen und Patienten zu wirksamen medizinischen Leistungen, der Sicherstellung eines vergleichsweise effektiven Kosten-Nutzen-Verhältnisses therapeutischer Behandlungen ('Value for Money') und damit einhergehend einer nachhaltigen Finanzierbarkeit von innovativen sowie auch bestehenden Interventionen.<sup>87</sup>

---

<sup>87</sup> Im Kontext der vorliegenden Arbeit bleibt abschließend nochmals anzumerken, dass HTA-Agenturen in vielen (europäischen) Gesundheitssystemen zwar eine bedeutende Rolle innerhalb der nationalen Entscheidungsprozesse zukommt, sie zugleich aber nur einen von mehreren wichtigen Akteuren darstellen. Ein aktuelles – unter Ägide der WHO – initiiertes Projekt greift dieses Thema in seinem Kern auf und stellt die einzelnen relevanten Akteure in den Vordergrund; im Fokus steht dabei insbesondere deren Vernetzung und interdisziplinäre Zusammenarbeit, sodass ein weiter gefasstes und besseres Verständnis von gesundheitspolitischen Entscheidungsfindungen im Allgemeinen ermöglicht wird (vgl. Schönemann et al., 2022).

Aufbauend auf den zuvor dargelegten Ergebnisdarstellungen und Diskussionsschwerpunkten werden nachfolgend einige schlussfolgernde Aspekte<sup>88</sup> abgeleitet, die der Autor im Kontext von HTA gegenwärtig sowie perspektivisch als gesundheitspolitisch und -ökonomisch relevant erachtet:

- Die identifizierten Abweichungen in den Empfehlungen offizieller HTA-Institutionen begründen sich primär durch die Unterschiede in den nationalen HTA-Verfahren und -Methoden sowie den vorgegebenen politisch-rechtlichen Rahmen, der den institutionellen Kontext bestimmt. Zudem haben die gesundheitspolitischen Regulierungen und strukturellen Gegebenheiten, die dem jeweiligen Gesundheitssystem zugrunde liegen, einen nicht unerheblichen Einfluss auf die abschließenden Outcomes nationaler HTAs.
- Bei der Bewertung neuer Gesundheitstechnologien werden allen voran indikationsspezifische Unterschiede offensichtlich, die – wenn auch nicht ausschließlich – auf den zentralen Bewertungsschwerpunkten (klinische Wirksamkeit nach den Prinzipien der EbM bei GBA/IQWiG versus Kosteneffektivität basierend auf dem ICER/QALY bei NICE) beruhen. Eine besondere Herausforderung stellen dabei Krebstherapien sowie ODs für seltene und sehr seltene Erkrankungen dar, die den standardisierten Evaluationsprozessen in der Regel nur bedingt standhalten können, weshalb ergänzende Bewertungsfaktoren oder eigens dafür implementierte HTA-Verfahren Anwendung finden (sollten).
- Das vorrangige Ziel von systematischen HTAs offizieller Einrichtungen liegt grundsätzlich darin, sowohl klinisch als auch ökonomisch effektive Therapien flächendeckend zugänglich zu machen, indem deren Bezahlbarkeit sowie nachhaltige Finanzierung innerhalb des Gesundheitssystems sichergestellt wird, damit die zur Verfügung stehenden Ressourcen transparent und zugleich möglichst gerecht allokiert werden.
- Neben den Faktoren zur Evaluation der klinischen und ökonomischen Wirksamkeit sollten elementare Gesichtspunkte wie die Verteilungsgerechtigkeit, gesellschaftliche Präferenzen und soziale Bedürfnisse (verstärkt) in die Bewertung innovativer Therapien einbezogen

---

<sup>88</sup> Die Schlussfolgerungen beziehungsweise Implikationen beziehen sich primär auf (europäische) Länder mit weitgehend etablierten HTA-Institutionen und -Verfahren. Ergänzende und zugleich weiterführende Literatur zum Thema HTA in Europa findet sich beispielsweise bei Kalo et al. (2016) und Beletsi et al. (2018). Eine weltweite Übersicht bietet unter anderem der 'WHO Global Survey on HTA' (WHO, 2015).

werden, damit die Entscheidungsfindung einen (möglichst) ganzheitlichen Kontext umfasst und zugleich eine qualitativ hochwertige Versorgung aller Patientinnen und Patienten ermöglicht.

- Die Vereinheitlichung oder zumindest weitgehende Harmonisierung von HTA-Verfahren – insbesondere auf europäischer Ebene – schreitet (zumindest in der Theorie) voran und setzt dabei einen lernfähigen, kooperativen und ganzheitlichen Ansatz voraus, der alle (entscheidungs-)relevanten Institutionen und Stakeholder einbeziehen sollte. Dabei gilt es jedoch immer zu berücksichtigen, dass (mögliche) nachgelagerte Erstattungsentscheidungs- sowie Preisfindungsprozesse den nationalen Regulierungen des jeweils vorhandenen Gesundheitssystems unterliegen und darüber hinaus Diskrepanzen sowohl auf sozial-kultureller Ebene als auch hinsichtlich bestehender Wertevorstellungen existieren, die von mitentscheidender Bedeutung für die Akzeptanz und damit auch für die Erfolgsaussichten einer nachhaltigen Umsetzung in die Praxis sind.

Die aus den vorliegenden Resultaten und Diskussionsabschnitten der Dissertation abgeleiteten Schlussfolgerungen legen den Status quo sowie potenzielle und bereits eingeleitete Entwicklungsschritte dar, die zugleich die bestehenden und zukünftigen Herausforderungen herausstellen. Einerseits liegen aufgrund der Unterschiede in den Gesundheitssystemen und Versorgungsstrukturen sowie den (teilweise) heterogenen HTA-Verfahren nur bedingt vergleichbare HTA-Outcomes respektive Empfehlungen durch nationale HTA-Agenturen vor, wie exemplarisch für GBA/IQWiG und NICE – und ergänzend für die HAS – aufgezeigt werden konnte. Andererseits handelt es sich bei den dargelegten Punkten um inhaltlich sowie thematisch relevante Ausführungen beziehungsweise abschließende Überlegungen, die in ihrer Darlegung weder vollständig noch „kohärent“ (letzteres im Sinne wissenschaftlicher Unanfechtbarkeit) sein können. Vielmehr sollen sie einen fundierten Beitrag für einen weiterführenden Diskurs im Kontext gesundheitspolitischer und -ökonomischer Fragestellungen rund um das Thema HTA beitragen, der in den zurückliegenden Jahren angestoßen wurde.

## 6.4 Limitationen und weiterer Forschungsbedarf

Die Untersuchungen der vorliegenden Arbeit zu den offiziellen HTAs von GBA/IQWiG in Deutschland und NICE in England basieren auf einem umfassenden Erhebungszeitraum, der die relevanten Daten von Januar 2011 bis einschließlich Juni 2018 einschließt. Dabei wurden jedoch ausschließlich zum Zeitpunkt der Datenerfassung gültige HTAs einbezogen, sodass abgelaufene oder zurückgezogene Bewertungen nicht berücksichtigt werden konnten. NICE TAs, die bereits vor Januar 2011 – und damit vor der Einführung des AMNOG in Deutschland – durchgeführt und öffentlich zugänglich gemacht wurden, sind ebenfalls kein Bestandteil der durchgeführten Analysen. Darüber hinaus wurden für NICE ausschließlich STAs und FTAs herangezogen, weshalb sowohl MTAs als auch HSTs nicht Bestandteil der Analysen sind.

### Vergleichende Untersuchung der HTA-Outcomes von GBA, HAS und NICE

Für die vergleichende Untersuchung der HTA-Outcomes wurden zusätzlich die von der HAS in Frankreich veröffentlichten ACTs herangezogen, die auf Grundlage der zuvor durchgeführten systematischen Gegenüberstellung der GBA FNBs und NICE STAs identifiziert werden konnten. Als wesentliche Limitation ist dabei herauszustellen, dass der Fokus ausschließlich auf den übereinstimmenden Interventionspaaren liegt, da die Definitionen der Patienten(sub)gruppen von GBA, HAS und NICE voneinander abweichen können. Um die identifizierten HTA-Paare folglich eindeutig vergleichen zu können, sind entsprechende Annahmen zu treffen gewesen. Für die Untersuchung bedeutet das konkret, dass bei Gesundheitstechnologien mit mehreren Patienten(sub)gruppen die HTA-Outcomes für die Teilpopulation mit der größten Patientenzahl herangezogen wurden, um die Ergebnisse auf der Hauptindikationsebene vergleichbar zu machen. Demnach wurde durchgehend die Hauptindikation der Therapien erfasst, damit die HTA-Paare auf einer vergleichbaren (Indikations-)Ebene gegenübergestellt werden konnten, da sich Patientensubgruppen wie auch Vergleichstherapien unterscheiden können.

Die Empfehlungen respektive Entscheidungen der HTA-Agenturen wurden dabei in 'positiv' (vorhanden) und 'negativ' (nicht vorhanden) unterteilt, weshalb mögliche Einschränkungen oder darüber hinaus gehende Spezifikationen – beispielsweise von einer bestimmten Teilpopulation innerhalb der vorliegenden Indikation – nicht weiter untersucht werden konnten. Schließlich wurden die (potenziellen) Einflusskriterien nicht explizit analysiert, sondern vorrangig solche Faktoren, die hinsichtlich der HTA-Outcomes onkologischer Interventionen – als das Indikationsgebiet mit der größten Anzahl veröffentlichter HTAs – von besonderem Interesse sind, wie unter anderem der OD-Status in Deutschland oder die EoL-Kriterien in England.



### Multivariate Regressionsanalysen bei GBA/IQWiG und NICE

Demgegenüber beziehen die multivariaten Regressionsanalysen sowohl für GBA und IQWiG als auch für NICE alle identifizierten und aktuellen HTAs innerhalb des vorliegenden Untersuchungszeitraums ein. Dabei wurden ausschließlich offizielle sowie in anderen vergleichenden oder länderspezifischen Studien als relevant herausgestellte Bewertungskriterien berücksichtigt, die demzufolge Evaluationsbestandteil der veröffentlichten HTAs von GBA und IQWiG sowie NICE sein sollten (vgl. hierzu Kapitel 6.1 und 6.2).

Als zentrale Limitation der Analysen sind die auf den von GBA/IQWiG und NICE angegebenen (potenziellen) Patientenzahlen und Kostenangaben basierenden Berechnungen der unterschiedlichen BI-Szenarien anzuführen, die lediglich Schätzungen darstellen können und den Einschränkungen beziehungsweise Unsicherheiten der beiden zugrunde liegenden Einflussvariablen unterliegen. Dabei basieren die jeweils relevanten Zielpopulationen neuer Therapien auf den abgeschlossenen Bewertungen (sogenannte ex-post Betrachtung), wonach Angaben zu einer gesamten theoretisch geeigneten respektive behandelbaren Patientenpopulation innerhalb der jeweiligen Versorgungsstruktur für gewöhnlich nur rückwirkend identifiziert oder rechnerisch bestimmt werden können (sogenannte ex-ante Betrachtung). Die vom Autor gewählte Vorgehensweise begründet sich darin, dass einerseits der GBA keine und das IQWiG nur teilweise eigens ermittelte Angaben zu den Kostenfolgen neuer Therapien für die GKV ausweisen. Die potenziellen Gesamtkostenauswirkungen positiv bewerteter Teilpopulationen der Therapien finden sich lediglich in den pU-Dossiers, die zu einer solchen Angabe verpflichtet sind. Andererseits veröffentlicht NICE Angaben zum potenziellen BI nur für Technologien, die eine (eingeschränkt oder uneingeschränkt) positive Empfehlung für eine routinemäßige Anwendung innerhalb des NHS erhalten und damit lediglich für die Patienten(sub)gruppen, die durch die Empfehlung explizit eingeschlossen sind. Demnach liegen für durch NICE negativ bewertete STAs keine Schätzungen zu den möglichen Kostenfolgen vor, welche daher nur den durch die pUs eingereichten Unterlagen zu entnehmen (gewesen) sind, sofern die relevanten Angaben aus Vertraulichkeitsgründen vor der Veröffentlichung nicht unkenntlich gemacht wurden.

Um die Regressionen für GBA und IQWiG vergleichbar modellieren zu können, wurden Anpassungen der IQWiG-Bewertungen an die GBA-Beschlussfassungen vorgenommen, indem bestehende Unterschiede in den Definitionen von Patiententeilpopulationen zusammengefasst wurden. Damit wird zugleich offensichtlich, dass eine direkte Vergleichbarkeit der GBA-Beschlüsse und IQWiG-Nutzenbewertungen, sowie auch zu den Angaben der pUs auf Teilpopulationsebene, nicht immer eindeutig möglich gewesen ist. Für die unterstützenden Analysen auf

Verfahrensebene wurde die jeweils am positivsten evaluierte Teilpopulation (mit den entsprechend zugehörigen Evaluationskriterien) herangezogen, wobei die Outcome-Variable um die Kodierung 'eingeschränkt positiv' ergänzt wurde (das heißt nur für einzelne Teilpopulationen innerhalb einer Indikation liegen positive Outcomes vor). Weiterhin bleibt zu konstatieren, dass durch IQWiG veröffentlichte Addenda (sofern vorhanden aufgrund einer nachträglichen Beauftragung durch den GBA) nicht weiter berücksichtigt wurden. Darüber hinaus wurden Therapien mit einem OD-Status aus den Regressionsmodellen für GBA und IQWiG ausgeschlossen, wenn aufgrund des definitionsgemäß vorhandenen Zusatznutzens kein nachweislicher Rückschluss von den Outcomes auf einzelne Evaluationsfaktoren möglich gewesen ist beziehungsweise relevante Daten unvollständig gewesen sind.

Vergleichbar den früheren empirischer Studien (Cerri et al., 2014; Dakin et al., 2006; Devlin and Parkin, 2004; Mauskopf et al., 2013) zu NICE TAs und den (potenziellen) Evaluationsfaktoren berücksichtigt die vorliegende Arbeit drei Outcome-Variablen (empfohlen, restriktiv, nicht empfohlen), um möglichst aussagekräftige Erkenntnisse aus den Empfehlungen für einzelne Patienten(sub)gruppen ziehen zu können. Im Gegensatz zu zwei Outcome-Variablen bedeutet dies (vgl. hierzu beispielsweise bei Dakin et al., 2015), dass sich die in den restriktiven Empfehlungen berücksichtigte Evidenz zwischen (potenziell vorhandenen) Patientengruppen unterscheidet und zugleich der Anteil der Patientinnen und Patienten, die die Therapie (tatsächlich) in Anspruch nehmen (werden), variieren kann. Folglich unterlag die Kategorisierung mittels der nach Raftery (2006) definierten Kriterien einer durch den Autor nachgelagerten Beurteilung (ex-post), die auf den abschließenden Empfehlungen von NICE basiert (vgl. hierzu bei O'Neill and Devlin, 2010). Die unterstützenden Analysen auf Indikationsebene entsprechen der Vorgehensweise für GBA/IQWiG in Deutschland, wobei die dreistufige Outcome-Variable dabei unverändert bleibt. Bei fehlenden Informationen in den NICE STA Guidance oder Costing Templates / Resource Impact Statements, wie unter anderem Angaben zum ICER/QALY oder der Patientenpopulation, wurde auf den zugehörigen ERG Report zurückgegriffen. Kostenangaben zur neuen Technologie, die ausschließlich für Therapiezyklen angegeben wurden (wie zum Beispiel bei bestimmten Krebsbehandlungen), mussten entsprechend auf Jahresbasis hochgerechnet werden. Bei fehlenden Informationen zur (relevanten) Vergleichstherapie, insbesondere bei fehlenden Angaben zu deren Kosten, wurde in Anlehnung an Mauskopf et al. (2013) vorgegangen (vgl. hierzu Kapitel 4.3.2). Abschließend bleibt anzumerken, dass die zusätzliche CDF-Variable im alternativen Modell – welches ansonsten identisch ist zur Modellierung ohne deren gesonderte Berücksichtigung – nur für einen wesentlich kürzeren Zeitraum eingeschlossen werden konnte, da der Relaunch innerhalb des NHS erst seit Juli 2016 besteht.

### Erkenntnisgewinn der Arbeit als Grundlage für weitere Forschungsbemühungen

Die vorliegende Dissertation liefert aktuelle Erkenntnisse zur (potenziellen) Rolle des BI in den HTA-Prozessen von GBA/IQWiG in Deutschland und NICE in England. Während die empirische Analyse für die von GBA und IQWiG veröffentlichten FNBs die erste ihrer Art darstellt, die unterschiedliche BI-Szenarien untersucht, ergänzen (und bestätigen) die Analysen der NICE STAs die Studienergebnisse von Mauskopf et al. (2013). Dabei weist der (potenzielle) BI einen Effekt sowohl auf die Outcomes von GBA/IQWiG als auch von NICE auf, obgleich er in beiden HTA-Prozessen keinen (offiziellen) Bewertungsfaktor darstellt. Um den Effekt der verschiedenen BI-Szenarien hinsichtlich der Outcomes von GBA/IQWiG und NICE auf Robustheit sowie (statistische) Signifikanz zu testen, wurden die BI-Berechnungen durch die Hinzunahme der jeweils offiziellen oder in der einschlägigen Literatur als empirisch relevant eingestuften Evaluationsfaktoren kontrolliert. Darüber hinaus werden die Analyseergebnisse durch die Resultate der vergleichenden Untersuchung der HTA-Outcomes übereinstimmender Interventionspaare von GBA und NICE – sowie ergänzend der französischen HAS – untermauert (vgl. hierzu bei Schaefer et al., 2021; Schaefer and Schlander, 2019).

Nichtsdestotrotz sind weitere Bemühungen erforderlich, um ein noch besseres Verständnis der BIAs und ihrer (bestehenden) Rolle in den von GBA/IQWiG und NICE veröffentlichten HTAs zu bekommen und einen Vergleich mit anderen HTA-Institutionen und deren Vorgehensweisen vornehmen zu können, die den BI als ein offizielles Bewertungskriterium in den Entscheidungsprozessen berücksichtigen. Zudem besteht zusätzlicher Forschungsbedarf hinsichtlich der indikationsspezifischen Unterschiede in den Empfehlungen nationaler HTA-Institutionen. Dabei sind die onkologischen und hämatologischen Therapien von besonderem Interesse, um die bestehenden Diskrepanzen zwischen den HTA-Methoden der einzelnen Agenturen sowie dem zugrunde liegenden institutionellen Kontext und den gesundheitspolitischen Regulierungen besser erklären und zugleich gesamtheitlich einordnen zu können. Krebstherapien bilden dabei einen zentralen Ansatzpunkt für die ausführliche Untersuchung von Interventionen seltener und sehr seltener Erkrankungen, bei der jedoch gerade auch nicht-onkologische Therapien explizit hinzugezogen und entsprechend analysiert werden sollten.

## 7 ZUSAMMENFASSUNG

Die Implementierung von HTA zur systematischen Bewertung medizinischer Interventionen ist auf das Ziel einer möglichst effektiven und effizienten Nutzung innovativer Therapien zurückzuführen, um die gesundheitspolitische und versorgungspraktische Entscheidungsfindung im Gesundheitssektor unterstützen zu können. In seiner Umsetzung stützt sich HTA vorrangig auf zwei Säulen: Einerseits auf die Grundsätze der EbM zur Bewertung der klinischen Wirksamkeit; andererseits auf eine Bewertung des Kosten-Nutzen-Verhältnisses, welches im Beispiel Deutschlands implizit und im Beispiel Englands explizit – unter Rekurs auf die Logik der Kosteneffektivität – erfolgt. Unter der Prämisse, dass sich die Analyse der Kosteneffektivität auf die Kosten pro Patient beziehungsweise Maßnahme fokussiert, lässt der gegenwärtige Forschungsstand vermuten, dass BIAs eine größere Bedeutung für die gesamtheitliche Entscheidungsfindung im Gesundheitswesen zukommen sollte. Sie werden als ergänzende Komponente einer umfassenden ökonomischen Bewertung diskutiert, um den möglichen Einfluss einer neuen Therapie auf das zur Verfügung stehende Budget oder die Gesamtkosten (besser) abschätzen zu können.

Vor diesem Hintergrund befasst sich die Dissertation aufbauend auf dem systematischen Vergleich offizieller HTA-Outcomes mit den jeweiligen Einflussfaktoren – unter besonderer Berücksichtigung des BI – und nimmt dabei Bezug auf GBA/IQWiG und NICE, welche die unterschiedlichen methodischen HTA-Ansätze exemplifizieren. Während sich GBA und IQWiG in Deutschland seit Einführung des AMNOG im Jahr 2011 (maßgeblich) auf die Prinzipien der EbM stützen, begründen sich die Empfehlungen durch NICE in England (aus empirischer Sicht) vorrangig auf dem Nutzenmaß der QALYs. Der BI spielt in beiden HTA-Systemen eine (bisher eher) nachgeordnete Rolle und stellt demnach keinen formellen Bewertungsfaktor dar.

Für den Untersuchungszeitraum von Januar 2011 bis einschließlich Juni 2018 sind alle aktuellen und öffentlich zugänglichen HTAs der beiden Agenturen identifiziert worden. Daraufhin wurden zuerst alle übereinstimmenden Interventionspaare von GBA und NICE – sowie ergänzend von der HAS in Frankreich – bestimmt und systematisch gegenübergestellt. Auf Grundlage der extrahierten Daten aller relevanten HTAs wurden sowohl für GBA/IQWiG als auch für NICE deskriptive Untersuchungen vorgenommen und daran anschließend multivariate Regressionsanalysen zu den einzelnen HTA-Outcomes sowie den (offiziellen) Evaluationsfaktoren durchgeführt.

Die Untersuchung der Outcomes der übereinstimmenden HTA-Paare zeigt, dass NICE innovative Therapien insgesamt positiver zu bewerten scheint als GBA und HAS. Jedoch existieren deutliche Unterschiede bei Betrachtung der HTA-Outcomes nach Indikationsgebieten; beispielsweise fallen Bewertungen onkologischer und hämatologischer Therapien beim GBA tendenziell besser aus als bei HAS und NICE. Die Empfehlungen von Krebsmedikamenten weichen zudem weniger voneinander ab als die der nicht-onkologischen Therapien. Darüber hinaus weisen die Outcomes von GBA/HAS eine weitaus höhere Übereinstimmung auf als GBA/NICE oder HAS/NICE. Die Ergebnisse bestätigen damit, dass unterschiedliche methodische Vorgehensweisen mit abweichenden HTA-Outcomes verknüpft sind, welche – insbesondere bei Krebstherapien – von weiteren Determinanten sowie möglichen Diskrepanzen innerhalb des jeweiligen institutionellen Kontexts beeinflusst werden.

Sowohl die GBA-Beschlussfassungen als auch die IQWiG-Bewertungen orientieren sich an den Prinzipien der EbM respektive den evidenzbasierten Evaluationskriterien. Die Analysereultate zeigen eindrücklich, dass die vergleichende klinische Wirksamkeit in Form patientenrelevanter Endpunkte als das zentrale Evaluationskriterium einen signifikanten Einfluss sowohl auf die GBA-Beschlüsse als auch die IQWiG-Empfehlungen hat. Das klinische Studiendesign in Form relevanter RCTs weist ausschließlich einen Zusammenhang mit den Bewertungen durch IQWiG auf. Die Outcomes von GBA und IQWiG stimmen dabei weitgehend überein; es zeigt sich jedoch, dass der GBA auf Grundlage der Nutzenbewertungen 'moderatere' Beschlüsse zu fassen scheint, wohingegen IQWiG den (eigenen) Methoden unter Berücksichtigung der zur Verfügung stehenden Evidenz strikt folgt. Interessanterweise korrelieren zudem die potenziellen BI-Szenarien mit den IQWiG-Bewertungen (linearer Zusammenhang positiver Bewertungsempfehlungen und zunehmender BI-Schätzungen). Dies wiederum unterstreicht jedoch vielmehr die (im Grunde evidente) Tatsache, dass den GBA-Beschlussfassungen weitere Entscheidungsfaktoren zugrunde liegen (müssen), wozu – im Umkehrschluss – dann auch die Kostenfolgen einer neuen Therapie gehören könnten.

Die Analyseergebnisse für NICE zeigen, dass STAs primär vom ICER/QALY beeinflusst werden. Dabei wird deutlich, dass mit zunehmendem ICER die Wahrscheinlichkeit für eine Ablehnung durch NICE signifikant steigt. Bei (teuren) Krebstherapien scheint die Berücksichtigung der EoL-Kriterien einen Effekt auf die STAs zu haben. Der potenzielle BI weist einen Einfluss auf die Empfehlungen durch NICE auf, wobei sich die Wahrscheinlichkeit einer negativen Bewertung ab einem bestimmten Wert respektive Wertebereich deutlich erhöht (U-förmiger Zusammenhang positiver STA-Empfehlungen und zunehmender BI-Schätzungen). Dieser Effekt scheint sich für die Krebstherapien weitgehend aufzuheben, seit sie (nachträglich) durch den

CDF erstattet werden können. Dabei sind die Resultate unter Heranziehung des CDF prinzipiell vergleichbar, lediglich STAs onkologischer Interventionen fallen seit dem CDF-Relaunch (im Juli 2016) besser aus. Zudem scheint der klinischen Wirksamkeit eine höhere Priorität zuzukommen, sodass sich infolgedessen die Kosteneffektivität als weniger entscheidend für eine Empfehlung durch NICE darstellt.

Abschließend bleibt festzuhalten, dass erstens die vergleichenden Untersuchungsergebnisse bestehende Unterschiede in den Empfehlungen nationaler HTA-Institutionen bestätigen und zweitens, dass sowohl GBA/IQWiG als auch NICE den (eigenen) offiziellen Kriterien innerhalb der jeweiligen Evaluationsprozesse in robuster und zugleich konsistenter Weise folgen. Während sich die GBA-Beschlüsse und IQWiG-Bewertungen rigoros an den Prinzipien der EbM orientieren, werden die Empfehlungen durch NICE primär von den Kosten je gewonnenem QALY bestimmt. Offensichtlich scheinen die potenziellen Kostenfolgen (auch) nach Kontrolle der sich als relevant herausgestellten Evaluationsfaktoren Einfluss auf die HTA-Ergebnisse in Deutschland und England zu haben, wenngleich ihre (praktische) Rolle innerhalb der Entscheidungsverfahren von GBA/IQWiG und NICE nicht eindeutig geklärt werden konnte.

## 8 LITERATURVERZEICHNIS

Adkins, E.M., Nicholson, L., Floyd, D., Ratcliffe, M., and Chevrou-Severac, H. (2017). Oncology drugs for orphan indications: how are HTA processes evolving for this specific drug category? *Clinicoecon Outcomes Res* 9, 327-342.

Aggarwal, A., and Sullivan, R. (2014). Affordability of cancer care in the United Kingdom – Is it time to introduce user charges? *Journal of Cancer Policy* 2, 31-39.

Aggarwal, A., Fojo, T., Chamberlain, C., Davis, C., and Sullivan, R. (2017). Do patient access schemes for high-cost cancer drugs deliver value to society?-lessons from the NHS Cancer Drugs Fund. *Ann Oncol* 28, 1738-1750.

Akehurst, R.L., Abadie, E., Renaudin, N., and Sarkozy, F. (2017). Variation in Health Technology Assessment and Reimbursement Processes in Europe. *Value Health* 20, 67-76.

Allen, N., Walker, S.R., Liberti, L., and Salek, S. (2017). Health Technology Assessment (HTA) Case Studies: Factors Influencing Divergent HTA Reimbursement Recommendations in Australia, Canada, England, and Scotland. *Value Health* 20, 320-328.

Angelis, A., and Kanavos, P. (2017). Multiple Criteria Decision Analysis (MCDA) for evaluating new medicines in Health Technology Assessment and beyond: The Advance Value Framework. *Soc Sci Med* 188, 137-156.

Angelis, A., Lange, A., and Kanavos, P. (2018). Using health technology assessment to assess the value of new medicines: results of a systematic review and expert consultation across eight European countries. *Eur J Health Econ* 19, 123-152.

Appleby, J., Devlin, N., Parkin, D., Buxton, M., and Chalkidou, K. (2009). Searching for cost effectiveness thresholds in the NHS. *Health Policy* 91, 239-245.

Baltussen, R., Jansen, M.P.M., Bijlmakers, L., Grutters, J., Kluytmans, A., Reuzel, R.P., Tummers, M., and der Wilt, G.J.V. (2017). Value Assessment Frameworks for HTA Agencies: The Organization of Evidence-Informed Deliberative Processes. *Value Health* 20, 256-260.

Baltussen, R., Marsh, K., Thokala, P., Diaby, V., Castro, H., Cleemput, I., Garau, M., Iskrov, G., Olyaeemanesh, A., Mirelman, A., Mobinizadeh, M., Morton, A., Tringali, M., van Til, J., Valentim, J., Wagner, M., Youngkong, S., Zah, V., Toll, A., Jansen, M., Bijlmakers, L., Oortwijn, W., and Broekhuizen, H. (2019). Multicriteria Decision Analysis to Support Health Technology Assessment Agencies: Benefits, Limitations, and the Way Forward. *Value Health* 22, 1283-1288.

Banta, D. (2003). The development of health technology assessment. *Health Policy* 63, 121-132.

Banta, D., and Jonsson, E. (2009). History of HTA: Introduction. *Int J Technol Assess Health Care* 25 *Suppl 1*, 1-6.

Battista, R.N., and Hodge, M.J. (1999). The evolving paradigm of health technology assessment: reflections for the millennium. *Can Med Assoc J* 160, 1464-1467.

Beletsi, A., Koutrafouris, V., Karampli, E., and Pavi, E. (2018). Comparing Use of Health Technology Assessment in Pharmaceutical Policy among Earlier and More Recent Adopters in the European Union. *Value Health Reg Issues* 16, 81-91.

Beresniak, A., and Dupont, D. (2016). Is there an alternative to quality-adjusted life years for supporting healthcare decision making? *Expert Rev Pharmacoecon Outcomes Res* 16, 351-357.

Beresniak, A., Medina-Lara, A., Auray, J.P., De Wever, A., Praet, J.C., Tarricone, R., Torbica, A., Dupont, D., Lamure, M., and Duru, G. (2015). Validation of the underlying assumptions of the quality-adjusted life-years outcome: results from the ECHOUTCOME European project. *Pharmacoeconomics* 33, 61-69.

Bestehorn, M., and Tunder, R. (2013). AMNOG – Erste Erfahrungen und mögliche Auswirkungen auf die Klinische Forschung. *Pharmacoeconomics German Research Articles* 11, 3-11.

BfArM (2020): Bundesinstitut für Arzneimittel und Medizinprodukte (BfArM) Dienstsitz Köln. Online: <https://www.dimdi.de/dynamic/de/das-dimdi/>, Stand: 01.08.2021.



Bilinski, A., Neumann, P., Cohen, J., Thorat, T., McDaniel, K., and Salomon, J.A. (2017). When cost-effective interventions are unaffordable: Integrating cost-effectiveness and budget impact in priority setting for global health programs. *PLoS Med* 14, e1002397.

Blome, C., Augustin, M., Metin, H., and Lohrberg, D. (2017). Four years of early benefit assessment of new drugs in Germany: a qualitative study on methodological requirements for quality of life data. *Eur J Health Econ* 18, 181-193.

Boucaud-Maitre, D., Berdaï, D., and Salvo, F. (2021). Added Therapeutic Value of Medicinal Products for French and German Health Technology Assessment Organizations: A Systematic Comparison. *Value Health* 24, 346-352.

Bouslouk, M. (2016). G-BA benefit assessment of new orphan drugs in Germany: the first five years. *Expert Opinion on Orphan Drugs* 4, 453-455.

Bovenberg, J., Penton, H., and Buyukkaramikli, N. (2021). 10 Years of End-of-Life Criteria in the United Kingdom. *Value Health* 24, 691-698.

Brazier, J., and Tsuchiya, A. (2015). Improving Cross-Sector Comparisons: Going Beyond the Health-Related QALY. *Appl Health Econ Health Policy* 13, 557-565.

Brenna, E., Polistena, B., and Spandonaro, F. (2020). The implementation of health technology assessment principles in public decisions concerning orphan drugs. *Eur J Clin Pharmacol*, 76, 755-764.

Broich, K., Löbker, W., Schulte, A., Beinlich, P., and Müller, T. (2016). Anforderungen an Zulassung und Zusatznutzenbewertung von Arzneimitteln: Regulatorische Aspekte und Erfahrungen. *Nervenarzt* 87, 376-385.

Bullemt, A., Taylor, M., McMordie, S.T., Waters, E., and Hatswell, A.J. (2019). NICE, in Confidence: An Assessment of Redaction to Obscure Confidential Information in Single Technology Appraisals by the National Institute for Health and Care Excellence. *Pharmacoeconomics* 37, 1383-1390.

Bullinger, M., Blome, C., Sommer, R., Lohrberg, D., and Augustin, M. (2015). Gesundheitsbezogene Lebensqualität – ein zentraler patientenrelevanter Endpunkt in der Nutzenbewertung medizinischer Maßnahmen. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 58, 283-290.

Busse, R., Blümel, M., Knieps, F., and Bärnighausen, T. (2017). Statutory health insurance in Germany: a health system shaped by 135 years of solidarity, self-governance, and competition. *Lancet* 390, 882-897.

Carioli, G., Bertuccio, P., Boffetta, P., Levi, F., La Vecchia, C., Negri, E., and Malvezzi, M. (2020). European cancer mortality predictions for the year 2020 with a focus on prostate cancer. *Ann Oncol* 31, 650-658.

Caro, J.J., Brazier, J.E., Karnon, J., Kolominsky-Rabas, P., McGuire, A.J., Nord, E., and Schlander, M. (2019). Determining Value in Health Technology Assessment: Stay the Course or Tack Away? *Pharmacoeconomics* 37, 293-299.

Caro, J.J., Nord, E., Siebert, U., McGuire, A., McGregor, M., Henry, D., de Pouvourville, G., Atella, V., and Kolominsky-Rabas, P. (2010). The efficiency frontier approach to economic evaluation of health-care interventions. *Health Econ* 19, 1117-1127.

Carroll, C., Kaltenthaler, E., Hill-McManus, D., Scope, A., Holmes, M., Rice, S., Rose, M., Tappenden, P., and Woolacott, N. (2017). The Type and Impact of Evidence Review Group Exploratory Analyses in the NICE Single Technology Appraisal Process. *Value Health* 20, 785-791.

Cassel, D., and Ulrich, V. (2015). AMNOG auf dem ökonomischen Prüfstand: Funktionsweise, Ergebnisse und Reformbedarf der Preisregulierung für neue Arzneimittel in Deutschland (Baden-Baden: Nomos).

Cerri, K.H., Knapp, M., and Fernandez, J. (2015). Untangling the Complexity of Funding Recommendations: A Comparative Analysis of Health Technology Assessment Outcomes in Four European Countries. *Pharm Med* 29, 341-359.

Cerri, K.H., Knapp, M., and Fernandez, J.L. (2014). Decision making by NICE: examining the influences of evidence, process and context. *Health Econ Policy Law* 9, 119-141.

Chalkidou, K. (2012). Evidence and values: paying for end-of-life drugs in the British NHS. *Health Econ Policy Law* 7, 393-409.

Charlton, V. (2020). NICE and Fair? Health Technology Assessment Policy Under the UK's National Institute for Health and Care Excellence, 1999-2018. *Health Care Anal* 28, 193-227.

Charlton, V. (2022). The normative grounds for NICE decision-making: a narrative cross-disciplinary review of empirical studies. *Health Econ Policy Law* 1, 10.1017/S1744133122000032.

Charlton, V., Littlejohns, P., Kieslich, K., Mitchell, P., Rumbold, B., Weale, A., Wilson, J., and Rid, A. (2017). Cost effective but unaffordable: an emerging challenge for health systems. *BMJ* 356, j1402.

Chassagnol, F., Marcelli, G., Wagle, J., Giuliani, G., Traub, D., Schaub, V., and Ruof, J. (2020). Review of Relative effectiveness assessments (REAs) of pharmaceuticals at the European network for health technology assessment (EUnetHTA): A first step towards a consolidated European perspective on comparative effectiveness & safety? *Health Policy* 124, 943-951.

Chevreur, K., Berg Brigham, K., Durand-Zaleski, I., and Hernández-Quevedo, C. (2015). France: Health system review (Copenhagen: European Observatory on Health Systems and Policies).

Chouaid, C., Borget, I., Braun, E., Bazil, M.L., Schaetz, D., Rémuzat, C., and Toumi, M. (2016). French Health Technology Assessment of Antineoplastic Drugs Indicated in the Treatment of Solid Tumours: Perspective for Future Trends. *Target Oncol* 11, 515-534.

Chugh, Y., De Francesco, M., and Prinja, S. (2021). Systematic Literature Review of Guidelines on Budget Impact Analysis for Health Technology Assessment. *Appl Health Econ Health Policy* 19, 825-838.

Clarke, S., Ellis, M., and Brownrigg, J. (2021). The impact of rarity in NICE's health technology appraisals. *Orphanet J Rare Dis* 16, 218.

Claxton, K., Martin, S., Soares, M., Rice, N., Spackman, E., Hinde, S., Devlin, N., Smith, P.C., and Sculpher, M. (2015). Methods for the estimation of the National Institute for Health and Care Excellence cost-effectiveness threshold. *Health Technol Assess* 19, 1-503, v-vi.

Cleemput, I., Neyt, M., Thiry, N., De Laet, C., and Leys, M. (2011). Using threshold values for cost per quality-adjusted life-year gained in healthcare decisions. *Int J Technol Assess Health Care* 27, 71-76.

Cochrane (2021): The Cochrane Collaboration: About us. Online: <https://www.cochrane.org/>, Stand: 01.07.2021.

Cohen, D. (2017). Most drugs paid for by £1.27bn Cancer Drugs Fund had no "meaningful benefit". *BMJ* 357, j2097.

Collins, M., and Latimer, N. (2013). NICE's end of life decision making scheme: impact on population health. *BMJ* 346, f1363.

Concato, J., Shah, N., and Horwitz, R.I. (2000). Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med* 342, 1887-1892.

Culyer, A.J. (2017). Ethics, priorities and cancer. *J Cancer Policy* 11, 6-11.

Culyer, A.J. (2020). Use of Evidence-Informed Deliberative Processes - Learning by Doing Comment on "Use of Evidence-informed Deliberative Processes by Health Technology Assessment Agencies Around the Globe". *Int J Health Policy Manag* 9, 263-265.

Dabisch, I., Dethling, J., Dintsios, C.M., Drechsler, M., Kalanovic, D., Kaskel, P., Langer, F., Ruof, J., Ruppert, T., and Wirth, D. (2014). Patient relevant endpoints in oncology: current issues in the context of early benefit assessment in Germany. *Health Econ Rev* 4, 2.

Dakin, H., Devlin, N., Feng, Y., Rice, N., O'Neill, P., and Parkin, D. (2015). The Influence of

Cost-Effectiveness and Other Factors on Nice Decisions. *Health Econ* 24, 1256-1271.

Dakin, H.A., Devlin, N.J., and Odeyemi, I.A. (2006). "Yes", "No" or "Yes, but"? Multinomial modelling of NICE decision-making. *Health Policy* 77, 352-367.

Daniels, N., and Sabin, J. (1997). Limits to health care: fair procedures, democratic deliberation, and the legitimacy problem for insurers. *Philos Public Aff* 26, 303-350.

de Folter, J., Trusheim, M., Jonsson, P., and Garner, S. (2018). Decision-Components of NICE's Technology Appraisals Assessment Framework. *Int J Technol Assess Health Care* 34, 163-171.

Detournay, B., Borel, T., Trancart, M., Emery, C., Coudray-Omnès, C., and LEEM working group on health technology assessment of drugs (2021). Clinical added value of drugs: An empirical survey of French transparency committee opinions. *Therapie* 76, 639-645.

Devlin, N., and Parkin, D. (2004). Does NICE have a cost-effectiveness threshold and what other factors influence its decisions? A binary choice analysis. *Health Econ* 13, 437-452.

Devlin, N.J., and Lorgelly, P.K. (2017). QALYs as a measure of value in cancer. *J Cancer Policy* 11, 19-25.

Dillon, A., and Landells, L.J. (2018). NICE, the NHS, and Cancer Drugs. *J Am Med Assoc* 319, 767-768.

Dintsios, C.M., Worm, F., Ruof, J., and Herpers, M. (2019). Different interpretation of additional evidence for HTA by the commissioned HTA body and the commissioning decision maker in Germany: whenever IQWiG and Federal Joint Committee disagree. *Health Econ Rev* 9, 35.

Dixon, P., Chamberlain, C., and Hollingworth, W. (2016). Did It Matter That the Cancer Drugs Fund Was Not NICE? A Retrospective Review. *Value Health* 19, 879-884.

Drummond, M., and Banta, D. (2009). Health technology assessment in the United Kingdom.

Int J Technol Assess Health Care 25 *Suppl 1*, 178-181.

Drummond, M., Brixner, D., Gold, M., Kind, P., McGuire, A., and Nord, E. (2009). Toward a consensus on the QALY. *Value Health 12 Suppl 1*, S31-35.

Drummond, M., de Pouvourville, G., Jones, E., Haig, J., Saba, G., and Cawston, H. (2014). A comparative analysis of two contrasting European approaches for rewarding the value added by drugs for cancer: England versus France. *Pharmacoeconomics 32*, 509-520.

Drummond, M., and Sorenson, C. (2009). Nasty or nice? A perspective on the use of health technology assessment in the United Kingdom. *Value Health 12 Suppl 2*, S8-13.

Drummond, M.F., Sculpher, M.J., Claxton, K., Stoddart, G.L., and Torrance, G.W. (2015). *Methods for the Economic Evaluation of Health Care Programmes*, 4th edn (Oxford: Oxford University Press).

Drummond, M.F., Schwartz, J.S., Jönsson, B., Luce, B.R., Neumann, P.J., Siebert, U., and Sullivan, S.D. (2008). Key principles for the improved conduct of health technology assessments for resource allocation decisions. *Int J Technol Assess Health Care 24*, 244-258; discussion 362-368.

Dubromel, A., Geffroy, L., Aulagner, G., and Dussart, C. (2018). Assessment and diffusion of medical innovations in France: an overview. *J Mark Access Health Policy 6*, 1458575.

Eddy, D.M. (2005). Evidence-based medicine: a unified approach. *Health Aff (Millwood) 24*, 9-17.

Edwards, N., Appleby, J., and Timmins, N. (2019). NICE at 20. *BMJ 364*, 11343.

Eichler, H.G., Kong, S.X., Gerth, W.C., Mavros, P., and Jonsson, B. (2004). Use of cost-effectiveness analysis in health-care resource allocation decision-making: how are cost-effectiveness thresholds expected to emerge? *Value Health 7*, 518-528.

EU-Kommission (2018): Europäische Kommission: Gesundheitswesen: Bewertung der Gesundheitstechnologie. Online:

[https://ec.europa.eu/health/technology\\_assessment/overview\\_de](https://ec.europa.eu/health/technology_assessment/overview_de), Stand: 31.01.2020.

EU-Kommission (2021): European Commission: Public Health: Health technology assessment: 11th meeting of the HTA Network: HTA Regulation. Online: [https://ec.europa.eu/health/technology\\_assessment/events/ev\\_20210707\\_en](https://ec.europa.eu/health/technology_assessment/events/ev_20210707_en), Stand: 07.07.2021.

EUnetHTA (2021): European Network for Health Technology Assessment: About EUnetHTA: History of EUnetHTA. Online: <https://www.eunetha.eu/>, Stand: 01.12.2021.

Ewbank, L., Omojomolo, D., Sullivan, K., and McKenna, H. (2018). The rising cost of medicines to the NHS: What's the story? (London: The King's Fund).

Facey, K., Granados, A., Guyatt, G., Kent, A., Shah, N., van der Wilt, G.J., and Wong-Rieger, D. (2014). Generating health technology assessment evidence for rare diseases. *Int J Technol Assess Health Care* 30, 416-422.

Fischer, K.E., Heisser, T., and Stargardt, T. (2016). Health benefit assessment of pharmaceuticals: An international comparison of decisions from Germany, England, Scotland and Australia. *Health Policy* 120, 1115-1122.

Fischer, K.E., and Stargardt, T. (2014). Early benefit assessment of pharmaceuticals in Germany: manufacturers' expectations versus the Federal Joint Committee's decisions. *Med Decis Making* 34, 1030-1047.

Foroutan, N., Tarride, J.E., Xie, F., and Levine, M. (2018). A methodological review of national and transnational pharmaceutical budget impact analysis guidelines for new drug submissions. *Clinicoecon Outcomes Res* 10, 821-854.

Gafni, A., and Birch, S. (2006). Incremental cost-effectiveness ratios (ICERs): the silence of the lambda. *Soc Sci Med* 62, 2091-2100.

Gammie, T., Lu, C.Y., and Babar, Z.U. (2015). Access to Orphan Drugs: A Comprehensive

Review of Legislations, Regulations and Policies in 35 Countries. *PLoS One* 10, e0140002.

Gandjour, A. (2013). A theoretical and empirical investigation into the willingness-to-pay function for new innovative drugs by Germany's health technology assessment agency (IQWiG). *Health Serv Manage Res* 26, 103-109.

Gandjour, A. (2020). Willingness to pay for new medicines: a step towards narrowing the gap between NICE and IQWiG. *BMC Health Serv Res* 20, 343.

Gandjour, A., Schübler, S., Hammerschmidt, T., and Dintsios, C.M. (2020). Predictors of negotiated prices for new drugs in Germany. *Eur J Health Econ* 21, 1049-1057.

Garattini, L., and van de Vooren, K. (2011). Budget impact analysis in economic evaluation: a proposal for a clearer definition. *Eur J Health Econ* 12, 499-502.

Garau, M., Shah, K.K., Mason, A.R., Wang, Q., Towse, A., and Drummond, M.F. (2011). Using QALYs in cancer: a review of the methodological limitations. *Pharmacoeconomics* 29, 673-685.

Garrison, L.P., Jr., Kamal-Bahl, S., and Towse, A. (2017). Toward a Broader Concept of Value: Identifying and Defining Elements for an Expanded Cost-Effectiveness Analysis. *Value Health* 20, 213-216.

G-BA (2019): Gemeinsamer Bundesausschuss: Bewertungsverfahren: Nutzenbewertung nach § 35a SGB V. Online: <https://www.g-ba.de/bewertungsverfahren/nutzenbewertung>, Stand: 31.01.2019.

G-BA (2020): Verfahrensordnung des Gemeinsamen Bundesausschusses. Online: [https://www.g-ba.de/downloads/62-492-2323/VerfO\\_2020-07-16\\_iK-2020-12-16.pdf](https://www.g-ba.de/downloads/62-492-2323/VerfO_2020-07-16_iK-2020-12-16.pdf), Stand: 16.12.2020.

G-BA (2021): Gemeinsamer Bundesausschuss: Über den G-BA. Online: <https://www.g-ba.de/>, Stand: 01.10.2021.



Gerber-Grote, A., Sandmann, F.G., Zhou, M., Ten Thoren, C., Schwalm, A., Weigel, C., Balg, C., Mensch, A., Mostardt, S., Seidl, A., and Lhachimi, S.K. (2014). Decision making in Germany: is health economic evaluation as a supporting tool a sleeping beauty? *Z Evid Fortbild Qual Gesundhwes* 108, 390-396.

Gerhardus, A., Dorendorf, E., Perleth, M., and Zechmeister-Koss, I. (2014). Die Effekte von HTA-Berichten im Gesundheitssystem. In *Health Technology Assessment: Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung*, M. Perleth, R. Busse, A. Gerhardus, B. Gibis, D. Lühmann, and A. Zentner, eds. (Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft), pp. 355-368.

Ghabri, S., Autin, E., Poullié, A.I., and Josselin, J.M. (2018). The French National Authority for Health (HAS) Guidelines for Conducting Budget Impact Analyses (BIA). *Pharmacoeconomics* 36, 407-417.

Gopinathan, U., Ottersen, T., Cyr, P.R., and Chalkidou, K. (2021). Evidence-Informed Deliberative Processes for HTA Around the Globe: Exploring the Next Frontiers of HTA and Best Practices Comment on "Use of Evidence-informed Deliberative Processes by Health Technology Assessment Agencies Around the Globe". *Int J Health Policy Manag* 10, 232-236.

Greiner, W., Witte, J., and Gensorowsky, D. (2019). AMNOG-Report 2019. Nutzenbewertung von Arzneimitteln in Deutschland. Schwerpunkt: Höchstpreise für ein Mehr an Lebensqualität? In *Beiträge zur Gesundheitsökonomie und Versorgungsforschung (Band 29)*, A. Storm, ed. (Heidelberg: medhochzwei Verlag).

Greiner, W., Witte, J., Gensorowsky, D., and Pauge, S. (2020). AMNOG-Report 2020. 10 Jahre AMNOG – Rückblick und Ausblick. In *Beiträge zur Gesundheitsökonomie und Versorgungsforschung (Band 32)*, A. Storm, ed. (Heidelberg: medhochzwei Verlag).

Grosios, K., Gahan, P.B., and Burbidge, J. (2010). Overview of healthcare in the UK. *EPMA J* 1, 529-534.

Gu, Y., Lancsar, E., Ghijben, P., Butler, J.R., and Donaldson, C. (2015). Attributes and weights in health care priority setting: A systematic review of what counts and to what extent. *Soc Sci*

Med 146, 41-52.

Hammerschmidt, T. (2017). Analyse der AMNOG-Erstattungsbeträge im europäischen Preisumfeld. *Gesundheitsökonomie & Qualitätsmanagement* 22, 43-53.

HAS (2015): Haute Autorité de Santé: Transparency Committee: Assessment Process. Online: [https://www.has-sante.fr/jcms/c\\_1729421/en/transparency-committee#toc\\_1\\_4\\_1](https://www.has-sante.fr/jcms/c_1729421/en/transparency-committee#toc_1_4_1), Stand: 01.07.2021.

HAS (2018): Haute Autorité de Santé: Missions de la HAS. Online: [https://www.has-sante.fr/jcms/c\\_1002212/fr/missions-de-la-has](https://www.has-sante.fr/jcms/c_1002212/fr/missions-de-la-has) Stand: 01.07.2021.

HAS (2019): Haute Autorité de Santé: Professionels: toutes nos publications par themes. Online: [www.has-sante.fr/jcms](http://www.has-sante.fr/jcms), Stand: 31.01.2019.

HAS (2020): Transparency Committee doctrine - Principles of medicinal product assessments and appraisal for reimbursement purposes. Online: [https://www.has-sante.fr/upload/docs/application/pdf/2019-07/doctrine\\_de\\_la\\_commission\\_de\\_la\\_transparence\\_-\\_version\\_anglaise.pdf](https://www.has-sante.fr/upload/docs/application/pdf/2019-07/doctrine_de_la_commission_de_la_transparence_-_version_anglaise.pdf), Stand: 01.10.2021.

HAS (2021): Haute Autorité de Santé: Commission d'évaluation économique et de santé publique. Online: [https://www.has-sante.fr/jcms/c\\_419565/commission-evaluation-des-strategies-de-sante](https://www.has-sante.fr/jcms/c_419565/commission-evaluation-des-strategies-de-sante), Stand: 01.10.2021.

Henshall, C., and Schuller, T. (2013). Health technology assessment, value-based decision making, and innovation. *Int J Technol Assess Health Care* 29, 353-359.

Herpers, M., and Dintsios, C.M. (2019). Methodological problems in the method used by IQWiG within early benefit assessment of new pharmaceuticals in Germany. *Eur J Health Econ* 20, 45-57.

Höer, A., and Berkemeier, F. (2020). AMNOG-Reporting. In *Arzneimittel-Atlas 2020*, B. Häussler, and A. Höer, eds. (Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft), pp. 145-176.

Höhle-Pasques, S., Hankowitz, J., and Oberender, P. (2013). Drei Jahre frühe Nutzenbewertung nach §35a SGB V – kritische Würdigung und Lösungsvorschlag. *Pharmacoeconomics German Res Articles 11*, 13-24.

Hofmann, S., Branner, J., Misra, A., and Lintener, H. (2021). A Review of Current Approaches to Defining and Valuing Innovation in Health Technology Assessment. *Value Health 24*, 1773-1783.

HTAi (2021): Health Technology Assessment international: About HTAi. Online: <https://htai.org/>, Stand: 01.07.2021.

Husereau, D., Henshall, C., Sampietro-Colom, L., and Thomas, S. (2016). CHANGING HEALTH TECHNOLOGY ASSESSMENT PARADIGMS? *Int J Technol Assess Health Care 32*, 191-199.

INAHTA (2020): International Network of Agencies for Health Technology Assessment: About INAHTA. Online: <https://www.inahta.org/>, Stand: 01.12.2020.

IQWiG (2020): Allgemeine Methoden (Version 6.0). Online: [https://www.iqwig.de/methoden/allgemeine-methoden\\_version-6-0.pdf](https://www.iqwig.de/methoden/allgemeine-methoden_version-6-0.pdf), Stand: 01.07.2021.

IQWiG (2021): Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen: Über uns: Aufgaben und Ziele des IQWiG. Online: <https://www.iqwig.de/>, Stand: 01.10.2021.

Ivandic, V. (2014). Requirements for benefit assessment in Germany and England - overview and comparison. *Health Econ Rev 4*, 12.

Jarosławski, S., Hanna, E., Dabbous, M., Chachoua, L., and Toumi, M. (2019). Heterogeneous Recommendations for Oncology Products Among Different HTA Systems: A Comparative Assessment. *Recent Results Cancer Res 213*, 39-55.

Julian, E., Gianfrate, F., Sola-Morales, O., Mol, P., Bergmann, J.-F., Salmonson, T., Hebborn, A., Grande, M., Ruof, J. (2022). How can a joint European health technology assessment provide an ‘additional benefit’ over the current standard of national assessments? *Health Econ Rev*

12, 30.

Kaier, K., and Fetzer, S. (2015). Das Arzneimittelmarktneuordnungsgesetz (AMNOG) aus ökonomischer Sicht. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 58, 291-297.

Kaiser, T., Vervölgyi, V., and Wieseler, B. (2015). Nutzenbewertung von Arzneimitteln. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 58, 232-239.

Kaló, Z., Gheorghe, A., Huic, M., Csanádi, M., and Kristensen, F.B. (2016). HTA Implementation Roadmap in Central and Eastern European Countries. *Health Econ* 25 *Suppl 1*, 179-192.

Kaltenthaler, E., Carroll, C., Hill-McManus, D., Scope, A., Holmes, M., Rice, S., Rose, M., Tappenden, P., and Woolacott, N. (2017). Issues Related to the Frequency of Exploratory Analyses by Evidence Review Groups in the NICE Single Technology Appraisal Process. *Pharmacoecoon Open* 1, 99-108.

Kaltenthaler, E., Papaioannou, D., Boland, A., and Dickson, R. (2011). The National Institute for Health and Clinical Excellence Single Technology Appraisal process: lessons from the first 4 years. *Value Health* 14, 1158-1165.

Kanavos, P., Angelis, A., and Drummond, M. (2019). An EU-wide approach to HTA: An irrelevant development or an opportunity not to be missed? *Eur J Health Econ* 20, 329-332.

Kawalec, P., Sagan, A., and Pilc, A. (2016). The correlation between HTA recommendations and reimbursement status of orphan drugs in Europe. *Orphanet J Rare Dis* 11, 122.

Kieslich, K. (2012). Social values and health priority setting in Germany. *J Health Organ Manag* 26, 374-383.

Kieslich, K. (2020). Paradigms in operation: explaining pharmaceutical benefit assessment outcomes in England and Germany. *Health Econ Policy Law* 15, 370-385.

Klarman, H.E., Francis, J.O'S., and Rosenthal, G.D. (1968). Cost Effectiveness Analysis Applied to the Treatment of Chronic Renal Disease. In *Medical Care* (Philadelphia: Lippincott Williams & Wilkins).

Kleijnen, S., George, E., Goulden, S., d'Andon, A., Vitre, P., Osinska, B., Rdzany, R., Thirstrup, S., Corbacho, B., Nagy, B.Z., Leufkens, H.G., de Boer, A., and Goettsch, W.G. (2012). Relative effectiveness assessment of pharmaceuticals: similarities and differences in 29 jurisdictions. *Value Health* 15, 954-960.

Kleijnen, S., Lipska, I., Leonardo Alves, T., Meijboom, K., Elsada, A., Vervölgyi, V., d'Andon, A., Timoney, A., Leufkens, H.G., De Boer, A., and Goettsch, W.G. (2016). Relative effectiveness assessments of oncology medicines for pricing and reimbursement decisions in European countries. *Ann Oncol* 27, 1768-1775.

Lauenroth, V.D., and Stargardt, T. (2017). Pharmaceutical Pricing in Germany: How Is Value Determined within the Scope of AMNOG? *Value Health* 20, 927-935.

Lebioda, A., Gasche, D., Dippel, F.W., Theobald, K., and Plantor, S. (2014). Relevance of indirect comparisons in the German early benefit assessment and in comparison to HTA processes in England, France and Scotland. *Health Econ Rev* 4, 31.

Linley, W.G., and Hughes, D.A. (2013). Societal views on NICE, cancer drugs fund and value-based pricing criteria for prioritising medicines: a cross-sectional survey of 4118 adults in Great Britain. *Health Econ* 22, 948-964.

Littlejohns, P., Chalkidou, K., Culyer, A.J., Weale, A., Rid, A., Kieslich, K., Coultas, C., Max, C., Manthorpe, J., Rumbold, B., Charlton, V., Roberts, H., Faden, R., Wilson, J., Krubiner, C., Mitchell, P., Wester, G., Whitty, J.A., and Knight, S. (2019). National Institute for Health and Care Excellence, social values and healthcare priority setting. *J R Soc Med* 112, 173-179.

Littlejohns, P., Garner, S., Doyle, N., Macbeth, F., Barnett, D., and Longson, C. (2009). 10 years of NICE: still growing and still controversial. *Lancet Oncol* 10, 417-424.

Lomas, J.R.S. (2019). Incorporating Affordability Concerns Within Cost-Effectiveness Analysis for Health Technology Assessment. *Value Health* 22, 898-905.

Loomes, G., and McKenzie, L. (1989). The use of QALYs in health care decision making. *Soc Sci Med* 28, 299-308.

Luce, B.R., Drummond, M., Jönsson, B., Neumann, P. J., Schwartz, J. S., Siebert, U., and Sullivan, S.D. (2010). EBM, HTA, and CER: clearing the confusion. *The Milbank Quarterly* 88, 256-276.

Luckett, T., King, M.T., Butow, P.N., Oguchi, M., Rankin, N., Price, M.A., Hackl, N.A., and Heading, G. (2011). Choosing between the EORTC QLQ-C30 and FACT-G for measuring health-related quality of life in cancer clinical research: issues, evidence and recommendations. *Ann Oncol* 22, 2179-2190.

Lühmann, D., Rüter, A., Schwarzer, R., Gawlik, C., Schiffner, R., and Perleth, M. (2014). Grundlagen und Prinzipien von Health Technology Assessment (HTA). In *Health Technology Assessment: Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung*, M. Perleth, R. Busse, A. Gerhardus, B. Gibis, D. Lühmann, and A. Zentner, eds. (Berlin: Medizinische Wissenschaftliche Verlagsgesellschaft), pp. 1-57.

Malinowski, K.P., Kawalec, P., Trabka, W., Sowada, C., and Pilc, A. (2018). Reimbursement of Orphan Drugs in Europe in Relation to the Type of Authorization by the European Medicines Agency and the Decision Making Based on Health Technology Assessment. *Front Pharmacol* 9, 1263.

Marsh, K., Ijzerman, M., Thokala, P., Baltussen, R., Boysen, M., Kaló, Z., Lönngren, T., Musson, F., Peacock, S., Watkins, J., and Devlin, N. (2016). Multiple Criteria Decision Analysis for Health Care Decision Making--Emerging Good Practices: Report 2 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health* 19, 125-137.

Marshall, D.A., Douglas, P.R., Drummond, M.F., Torrance, G.W., Macleod, S., Manti, O., Cheruvu, L., and Corvari, R. (2008). Guidelines for conducting pharmaceutical budget impact analyses for submission to public drug plans in Canada. *Pharmacoeconomics* 26, 477-495.

Mason, A.R., and Drummond, M.F. (2009). Public funding of new cancer drugs: Is NICE getting nastier? *Eur J Cancer* 45, 1188-1192.

Massetti, M., Aballea, S., Videau, Y., Remuzat, C., Roiz, J., and Toumi, M. (2015). A comparison of HAS & NICE guidelines for the economic evaluation of health technologies in the context of their respective national health care systems and cultural environments. *J Mark Access Health Policy* 3, 10.3402/jmahp.v3.24966.

Mathes, T., Jacobs, E., Morfeld, J.C., and Pieper, D. (2013). Methods of international health technology assessment agencies for economic evaluations--a comparative analysis. *BMC Health Serv Res* 13, 371.

Mauskopf, J. (1998). Prevalence-based economic evaluation. *Value Health* 1, 251-259.

Mauskopf, J., Earnshaw, S.R., Brogan, A., Wolowacz, S., and Brodtkorb, T.-H. (2017). *Budget-Impact Analysis of Health Care Interventions: A Practical Guide* (Berlin, Heidelberg: Adis).

Mauskopf, J., Chirila, C., Birt, J., Boye, K.S., and Bowman, L. (2013). Drug reimbursement recommendations by the National Institute for Health and Clinical Excellence: have they impacted the National Health Service budget? *Health Policy* 110, 49-59.

Mauskopf, J.A., Sullivan, S.D., Annemans, L., Caro, J., Mullins, C.D., Nuijten, M., Orlewska, E., Watkins, J., and Trueman, P. (2007). Principles of good practice for budget impact analysis: report of the ISPOR Task Force on good research practices--budget impact analysis. *Value Health* 10, 336-347.

Maynou, L., and Cairns, J. (2019). What is driving HTA decision-making? Evidence from cancer drug reimbursement decisions from 6 European countries. *Health Policy* 123, 130-139.

Morrell, L., Wordsworth, S., Rees, S., and Barker, R. (2017). Does the Public Prefer Health Gain for Cancer Patients? A Systematic Review of Public Views on Cancer and its Characteristics. *Pharmacoeconomics* 35, 793-804.

Morrell, L., Wordsworth, S., Schuh, A., Middleton, M.R., Rees, S., and Barker, R.W. (2018).

Will the reformed Cancer Drugs Fund address the most common types of uncertainty? An analysis of NICE cancer drug appraisals. *BMC Health Serv Res* 18, 393.

Nestler-Parr, S., Korchagina, D., Toumi, M., Pashos, C.L., Blanchette, C., Molsen, E., Morel, T., Simoens, S., Kaló, Z., Gatermann, R., and Redekop, W. (2018). Challenges in Research and Health Technology Assessment of Rare Disease Technologies: Report of the ISPOR Rare Disease Special Interest Group. *Value Health* 21, 493-500.

Neumann, P.J., Russell, L.B., Siegel, J.E., Prosser, L.A., Krahn, M., Mandelblatt, J.S., Daniels, N., and Gold, M.R. (2016). Using Cost-Effectiveness Analysis in Health and Medicine. In *Cost-Effectiveness in Health and Medicine*, P.J. Neumann, T.G. Ganiats, L.B. Russell, G.D. Sanders, and J.E. Siegel, eds. (Oxford: Oxford University Press).

Neumann, P.J., Bliss, S.K., and Chambers, J.D. (2012). Therapies for advanced cancers pose a special challenge for health technology assessment organizations in many countries. *Health Aff (Millwood)* 31, 700-708.

Neyt, M., Cleemput, I., Sande, S.V., and Thiry, N. (2015). Belgian guidelines for budget impact analyses. *Acta Clin Belg* 70, 175-180.

NHS England (2016): Appraisal and Funding of Cancer Drugs from July 2016 (including the new Cancer Drugs Fund) - A new deal for patients, taxpayers and industry. Online: <https://www.england.nhs.uk/wp-content/uploads/2013/04/cdf-sop.pdf>, Stand: 01.12.2020.

NICE (2008): National Institute for Health and Care Excellence: NICE Process and Methods Guides: Social Value Judgements: Principles for the Development of NICE Guidance. Online: [https://www.ncbi.nlm.nih.gov/books/NBK395865/pdf/Bookshelf\\_NBK395865.pdf](https://www.ncbi.nlm.nih.gov/books/NBK395865/pdf/Bookshelf_NBK395865.pdf), Stand: 01.12.2020.

NICE (2009): National Institute for Health and Clinical Excellence: Appraising life-extending, end of life treatments. Online: <https://www.nice.org.uk/guidance/gid-tag387/documents/appraising-life-extending-end-of-life-treatments-paper2>, Stand: 01.12.2020.



NICE (2013): National Institute for Health and Care Excellence: Guide to the methods of technology appraisal 2013. Online: <https://www.nice.org.uk/process/pmg9>, Stand: 01.12.2020.

NICE (2017): National Institute for Health and Care Excellence: Technology appraisal guidance: Budget impact test. Online: <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-technology-appraisal-guidance/budget-impact-test>, Stand: 01.10.2021.

NICE (2018): National Institute for Health and Care Excellence: Guide to the processes of technology appraisal. Online: <https://www.nice.org.uk/process/pmg19>, Stand: 30.05.2018.

NICE (2019): National Institute for Health and Care Excellence: Guidance and advice list: Published. Online: <http://www.nice.org.uk/guidance/published?type=cg>, Stand: 31.01.2019.

NICE (2020): National Institute for Health and Care Excellence: Highly specialised technologies guidance. Online: <https://www.nice.org.uk/about/what-we-do/our-programmes/nice-guidance/nice-highly-specialised-technologies-guidance>, Stand: 31.01.2021.

NICE (2021): National Institute for Health and Care Excellence: About. Online: <https://www.nice.org.uk/about>, Stand: 01.10.2021.

Nicod, E., Annemans, L., Bucsics, A., Lee, A., Upadhyaya, S., and Facey, K. (2019). HTA programme response to the challenges of dealing with orphan medicinal products: Process evaluation in selected European countries. *Health Policy* 123, 140-151.

Nicod, E., Berg Brigham, K., Durand-Zaleski, I., and Kanavos, P. (2017). Dealing with Uncertainty and Accounting for Social Value Judgments in Assessments of Orphan Drugs: Evidence from Four European Countries. *Value Health* 20, 919-926.

Nicod, E., and Kanavos, P. (2012). Commonalities and differences in HTA outcomes: a comparative analysis of five countries and implications for coverage decisions. *Health Policy* 108, 167-177.

- Nicod, E., Maynou, L., Visintin, E., and Cairns, J. (2020a). Why do health technology assessment drug reimbursement recommendations differ between countries? A parallel convergent mixed methods study. *Health Econ Policy Law* 15, 386-402.
- Nicod, E., Whittall, A., Drummond, M., and Facey, K. (2020b). Are supplemental appraisal/reimbursement processes needed for rare disease treatments? An international comparison of country approaches. *Orphanet J Rare Dis* 15, 189.
- Nord, E. (2015). Cost-value analysis of health interventions: introduction and update on methods and preference data. *Pharmacoeconomics* 33, 89-95.
- Ogden, J. (2017). How will NICE's budget impact test affect new drug availability? *Prescriber (August)*, 9-12.
- Oortwijn, W., Determann, D., Schiffers, K., Tan, S.S., and van der Tuin, J. (2017). Towards Integrated Health Technology Assessment for Improving Decision Making in Selected Countries. *Value Health* 20, 1121-1130.
- Oortwijn, W., Jansen, M., and Baltussen, R. (2020). Use of Evidence-Informed Deliberative Processes by Health Technology Assessment Agencies Around the Globe. *Int J Health Policy Manag* 9, 27-33.
- Orlewska, E., and Gulácsi, L. (2009). Budget-impact analyses: a critical review of published studies. *Pharmacoeconomics* 27, 807-827.
- Orlewska, E., and Mierzejewski, P. (2004). Proposal of Polish guidelines for conducting financial analysis and their comparison to existing guidance on budget impact in other countries. *Value Health* 7, 1-10.
- O'Rourke, B., Oortwijn, W., and Schuller, T. (2020). The new definition of health technology assessment: A milestone in international collaboration. *Int J Technol Assess Health Care* 36, 187-190.
- Orvain, J., Xerri, B., and Matillon, Y. (2004). Overview of health technology assessment in

France. *Int J Technol Assess Health Care* 20, 25-34.

Panteli, D., Nolting, A., Eckhardt, H., Kulig, M., and Busse, R. (2016). Published and unpublished evidence in coverage decision-making for pharmaceuticals in Europe: existing approaches and way forward. *Health Res Policy Syst* 14, 6.

Paulden, M., O'Mahony, J., and McCabe, C. (2017). Determinants of Change in the Cost-effectiveness Threshold. *Med Decis Making* 37, 264-276.

Paulden, M., O'Mahony, J.F., Culyer, A.J., and McCabe, C. (2014). Some inconsistencies in NICE's consideration of social values. *Pharmacoeconomics* 32, 1043-1053.

Pauwels, K., Huys, I., Casteels, M., De Nys, K., and Simoens, S. (2014). Market access of cancer drugs in European countries: improving resource allocation. *Target Oncol* 9, 95-110.

PBAC (2016): Guidelines for preparing a submission to the Pharmaceutical Benefits Advisory Committee (Version 5.0). Online: <https://pbac.pbs.gov.au/content/information/files/pbac-guidelines-version-5.pdf>, Stand: 01.10.2021.

Peinemann, F., and Labeit, A. (2019). Varying results of early benefit assessment of newly approved pharmaceutical drugs in Germany from 2011 to 2017: A study based on federal joint committee data. *J Evid Based Med* 12, 9-15.

Perleth, M., Zentner, A., Hoffmann, C., and Gibis, B. (2014). Priorisierung von HTA-Themen. In *Health Technology Assessment: Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung*, M. Perleth, R. Busse, A. Gerhardus, B. Gibis, D. Lühmann, and A. Zentner, eds. (Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft), pp. 114-134.

Perleth, M., Gibis, B., and Göhlen, B. (2009). A short history of health technology assessment in Germany. *Int J Technol Assess Health Care* 25 (*Suppl 1*), 112-119.

Postma, M. J., Noone, D., Rozenbaum, M. H., Carter, J. A., Botteman, M. F., Fenwick, E., and Garrison, L. P. (2022). Assessing the value of orphan drugs using conventional cost-effectiveness analysis: Is it fit for purpose? *Orphanet J Rare Dis* 17, 157.

Pujolras, L.M., and Cairns, J. (2015). Why do some countries approve a cancer drug and others don't? *J Cancer Policy* 4, 21-25.

Raftery, J. (2006). Review of NICE's recommendations, 1999-2005. *BMJ* 332, 1266-1268.

Rawlins, M., Barnett, D., and Stevens, A. (2010). Pharmacoeconomics: NICE's approach to decision-making. *Br J Clin Pharmacol* 70, 346-349.

Rémuzat, C., Toumi, M., and Falissard, B. (2013). New drug regulations in France: what are the impacts on market access? Part 1 - Overview of new drug regulations in France. *J Mark Access Health Policy* 1, 10.3402/jmahp.v1i0.20891.

Richardson, J., Iezzi, A., Sinha, K., Khan, M.A., and McKie, J. (2014). An instrument for measuring the social willingness to pay for health state improvement. *Health Econ* 23, 792-805.

Richardson, J., and Schlander, M. (2018). Health technology assessment (HTA) and economic evaluation: efficiency or fairness first. *J Mark Access Health Policy* 7, 1557981.

Rid, A., Littlejohns, P., Wilson, J., Rumbold, B., Kieslich, K., and Weale, A. (2015). The importance of being NICE. *J R Soc Med* 108, 385-389.

Riedel, R., Repschläger, U., Griebenow, R., Breitkopf, S., Schmidt, S., and Guhl, A. (2013). International standards for health economic evaluation with a focus on the German approach. *J Clin Pharm Ther* 38, 277-285.

Ruether, A., Imaz-Iglesia, I., Bêlorgey, C., Lo Scalzo, A., Garrett, Z., and Guardian, M. (2022). European collaboration on health technology assessment: looking backward and forward. *Int J Technol Assess Health Care* 38, e34.

Rumbold, B., Weale, A., Rid, A., Wilson, J., and Littlejohns, P. (2017). Public Reasoning and Health-Care Priority Setting: The Case of NICE. *Kennedy Inst Ethics J* 27, 107-134.

Ruof, J., Schwartz, F.W., Schulenburg, J.M., and Dintsios, C.M. (2014). Early benefit assessment (EBA) in Germany: analysing decisions 18 months after introducing the new AMNOG

legislation. *Eur J Health Econ* 15, 577-589.

Rychlik, R.P.T. (2013). Target patient population: how does the regulatory and reimbursement view differ? Insights from the German IQWiG. *J Public Health* 21, 109-114.

Sabry-Grant, C., Malottki, K., and Diamantopoulos, A. (2019). The Cancer Drugs Fund in Practice and Under the New Framework. *Pharmacoeconomics* 37, 953-962.

Sackett, D.L., Rosenberg, W.M., Gray, J.A., Haynes, R.B., and Richardson, W.S. (1996). Evidence based medicine: what it is and what it isn't. *BMJ* 312, 71-72.

Sandmann, F.G., Mostardt, S., Lhachimi, S.K., and Gerber-Grote, A. (2018). The efficiency-frontier approach for health economic evaluation versus cost-effectiveness thresholds and internal reference pricing: combining the best of both worlds? *Expert Rev Pharmacoecon Outcomes Res* 18, 475-486.

Schäfer, R. (2015). Health Technology Assessment (HTA): Eine vergleichende Untersuchung der Implementierung bei NICE in England und IQWiG in Deutschland. Masterthesis (unveröffentlicht). Hochschule für Wirtschaft und Gesellschaft Ludwigshafen am Rhein.

Schaefer, R., Hernandez, D., and Schlander, M. (2020). PDG70 The Role of the Budget Impact: A Multivariate Analysis of Health Technology Assessment (HTA) Outcomes by GBA/IQWiG in Germany and NICE in England. *Value in Health* 23, S530.

Schaefer, R., Hernandez, D., Selberg, L., and Schlander, M. (2021). Health technology assessment (HTA) in England, France and Germany: What do matched drug pairs tell us about recommendations by national HTA agencies? *J Comp Eff Res* 10, 1187-1195.

Schaefer, R., and Schlander, M. (2019). Is the National Institute for Health and Care Excellence (NICE) in England more 'innovation-friendly' than the Federal Joint Committee (G-BA) in Germany? *Expert Rev Pharmacoecon Outcomes Res* 19, 453-462.

Schey, C., Milanova, T., and Hutchings, A. (2011). Estimating the budget impact of orphan medicines in Europe: 2010 - 2020. *Orphanet J Rare Dis* 6, 62.

Schlender, M. (2007). *Health Technology Assessments by the National Institute for Health and Clinical Excellence: A Qualitative Study* (New York: Springer).

Schlender, M. (2021). HTA Agencies Need Evidence-Informed Deliberative Processes Comment on "Use of Evidence-Informed Deliberative Processes by Health Technology Assessment Agencies Around the Globe". *Int J Health Policy Manag* 10, 158-161.

Schlender, M. (2008). The use of cost-effectiveness by the National Institute for Health and Clinical Excellence (NICE): no(t yet an) exemplar of a deliberative process. *J Med Ethics* 34, 534-539.

Schlender, M., Dintsiou, C.M., and Gandjour, A. (2018). Budgetary Impact and Cost Drivers of Drugs for Rare and Ultrarare Diseases. *Value Health* 21, 525-531.

Schlender, M., Garattini, S., Holm, S., Kolominsky-Rabas, P., Nord, E., Persson, U., Postma, M., Richardson, J., Simoons, S., de Sola Morales, O., Tolley, K., and Toumi, M. (2014). Incremental cost per quality-adjusted life year gained? The need for alternative methods to evaluate medical interventions for ultra-rare disorders. *J Comp Eff Res* 3, 399-422.

Schünemann, H. J., Rezapour, M., Piggott, T., Laidmäe, E., Köhler, K., Pöld, M., Ens, B., Irs, A., Akl, E. A., Cuello, C. A., Falavigna, M., Gibbens, M., Neamtii, L., Parmelli, E., Jameleddine, M., Pyke, L., Verstijnen, I., Alonso-Coello, P., Tugwell, P., Zhang, Y., Saz-Parkinson, Z., Kuchenmüller, T., and Moja, L. (2022). The ecosystem of health decision making: from fragmentation to synergy. *Lancet Public Health* 7, e378-e390.

Schulz, S., Passon, A., Kulig, M., Perleth, M., and Matthias, K. (2018). Studien bei seltenen Erkrankungen: eine deskriptive Analyse abgeschlossener Orphan Drug Bewertungen im Gemeinsamen Bundesausschuss. *Gesundheitswesen* 80, e54-e61.

Shah, K.K., Tsuchiya, A., and Wailoo, A.J. (2015). Valuing health at the end of life: a stated preference discrete choice experiment. *Soc Sci Med* 124, 48-56.

Sheskin, D.J. (2011). *Handbook of Parametric and Nonparametric Statistical Procedures*, 5th edn (London: Chapman & Hall/CRC).

Simoens, S. (2014). Health technologies for rare diseases: does conventional HTA still apply? *Expert Rev Pharmacoecon Outcomes Res* 14, 315-317.

Skipka, G., Wieseler, B., Kaiser, T., Thomas, S., Bender, R., Windeler, J., and Lange, S. (2016). Methodological approach to determine minor, considerable, and major treatment effects in the early benefit assessment of new drugs. *Biom J* 58, 43-58.

Solidarites-sante (2019): Ministère des Solidarités de la Santé: Base de données publique des médicaments: Rechercher par substance active. Online: <https://m.base-donnees-publique.medicaments.gouv.fr/>, Stand: 31.03.2019.

Sorenson, C., and Chalkidou, K. (2012). Reflections on the evolution of health technology assessment in Europe. *Health Econ Policy Law* 7, 25-45.

Stafinski, T., Menon, D., Davis, C., and McCabe, C. (2011a). Role of centralized review processes for making reimbursement decisions on new health technologies in Europe. *Clinicoecon Outcomes Res* 3, 117-186.

Stafinski, T., Menon, D., Philippon, D.J., and McCabe, C. (2011b). Health technology funding decision-making processes around the world: the same, yet different. *Pharmacoeconomics* 29, 475-495.

Sullivan, S.D., Mauskopf, J.A., Augustovski, F., Caro, J., Lee, K.M., Minchin, M., Orlewska, E., Penna, P., Rodriguez Barrios, J.M., and Shau, W.Y. (2014). Budget impact analysis-principles of good practice: report of the ISPOR 2012 Budget Impact Analysis Good Practice II Task Force. *Value Health* 17, 5-14.

Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., and Bray, F. (2021). Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 71, 209-249.

Tafari, G., Pagnini, M., Moseley, J., Massari, M., Petavy, F., Behring, A., Catalan, A., Gajraj, E., Hedberg, N., Obach, M., Osipenko, L., Russo, P., Van De Castele, M., Zebedin, E.M., Rasi, G., and Vamvakas, S. (2016). How aligned are the perspectives of EU regulators and HTA

bodies? A comparative analysis of regulatory-HTA parallel scientific advice. *Br J Clin Pharmacol* 82, 965-973.

Tappenden, P., Brazier, J., Ratcliffe, J., and Chilcott, J. (2007). A stated preference binary choice experiment to explore NICE decision making. *Pharmacoeconomics* 25, 685-693.

Thokala, P., Devlin, N., Marsh, K., Baltussen, R., Boysen, M., Kalo, Z., Longrenn, T., Mussen, F., Peacock, S., Watkins, J., and Ijzerman, M. (2016). Multiple Criteria Decision Analysis for Health Care Decision Making--An Introduction: Report 1 of the ISPOR MCDA Emerging Good Practices Task Force. *Value Health* 19, 1-13.

Timmins, N. (2016). The first sixteen years: from Relenza to the Cancer Drugs Fund. In *A Terrible Beauty: A Short History of NICE: The National Institute for Health and Care Excellence*, N. Timmins, M. Rawlins, and J. Appleby, eds. (Bangkok: Amarin Printing & Publishing), pp. 8-135.

Timmins, N. (2019): NICE at 20: a cause for celebration. Online: <https://www.kingsfund.org.uk/blog/2019/03/nice-at-20>, Stand: 01.04.2021.

Torbica, A., Fornaro, G., Tarricone, R., and Drummond, M.F. (2020). Do Social Values and Institutional Context Shape the Use of Economic Evaluation in Reimbursement Decisions? An Empirical Analysis. *Value Health* 23, 17-24.

Torbica, A., Tarricone, R., and Drummond, M. (2018). Does the approach to economic evaluation in health care depend on culture, values, and institutional context? *Eur J Health Econ* 19, 769-774.

Tordrup, D., Tzouma, V., and Kanavos, P. (2014). Orphan drug considerations in health technology assessment in eight European countries. *Rare Diseases and Orphan Drugs Journal* 1, 83-97.

Toumi, M., Motrunich, A., Millier, A., Remuzat, C., Chouaid, C., Falissard, B., and Aballea, S. (2017). Analysis of health economics assessment reports for pharmaceuticals in France - understanding the underlying philosophy of CEESP assessment. *J Mark Access Health Policy* 5,



1344088.

Trueman, P., Drummond, M., and Hutton, J. (2001). Developing guidance for budget impact analysis. *Pharmacoeconomics* *19*, 609-621.

van de Vooren, K., Duranti, S., Curto, A., and Garattini, L. (2014). A critical systematic review of budget impact analyses on drugs in the EU countries. *Appl Health Econ Health Policy* *12*, 33-40.

Vella Bonanno, P., Bucsics, A., Simoens, S., Martin, A.P., Oortwijn, W., Gulbinovič, J., Rothe, C., Timoney, A., Ferrario, A., Gad, M., Salem, A., Hoxha, I., Sauermann, R., Kamusheva, M., Dimitrova, M., Petrova, G., Laius, O., Selke, G., Kourafalos, V., Yfantopoulos, J., Magnusson, E., Joppi, R., Jakupi, A., Bochenek, T., Wladysiuk, M., Furtado, C., Marković-Peković, V., Mardare, I., Meshkov, D., Fürst, J., Tomek, D., Cortadellas, M.O., Zara, C., Haycox, A., Campbell, S., and Godman, B. (2019). Proposal for a regulation on health technology assessment in Europe - opinions of policy makers, payers and academics from the field of HTA. *Expert Rev Pharmacoecon Outcomes Res* *19*, 251-261.

Walton, M.J., O'Connor, J., Carroll, C., Claxton, L., and Hodgson, R. (2019). A Review of Issues Affecting the Efficiency of Decision Making in the NICE Single Technology Appraisal Process. *Pharmacoecon Open* *3*, 403-410.

Weill, C., and Banta, D. (2009). Development of health technology assessment in France. *Int J Technol Assess Health Care* *25 Suppl 1*, 108-111.

Weinstein, M.C. (1988). A QALY is a QALY--or is it? *J Health Econ* *7*, 289-290.

Weinstein, M.C., and Stason, W.B. (1977). Foundations of cost-effectiveness analysis for health and medical practices. *N Engl J Med* *296*, 716-721.

Weiß, C. (2019). *Basiswissen Medizinische Statistik*, 7th edn (Berlin, Heidelberg: Springer).

WHO (2015). *2015 Global Survey on Health Technology Assessment by National Authorities. Main findings*; World Health Organization (Geneva: WHO Document Production Services).

Windeler, J., and Lange, S. (2015). Nutzenbewertung medizinischer Leistungen im deutschen Gesundheitswesen – rechtlicher Rahmen, historische und internationale Perspektive. *Bundesgesundheitsblatt – Gesundheitsforschung – Gesundheitsschutz* 58, 220-226.

Wright, J.S.F., Barron, A.J.G., Shah, S.M.B., and Klingler, C. (2017). Convergence, Divergence and Hybridity: A Regulatory Governance Perspective on Health Technology Assessment in England and Germany. *Global Policy* 8, 69-75.

Yuasa, A., Yonemoto, N., Demiya, S., Foellscher, C., and Ikeda, S. (2021). Investigation of Factors Considered by Health Technology Assessment Agencies in Eight Countries. *Pharmacoecon Open* 5, 57-69.

Zamora, B., Maignen, F., O'Neill, P., Mestre-Ferrandiz, J., and Garau, M. (2019). Comparing access to orphan medicinal products in Europe. *Orphanet J Rare Dis* 14, 95.

Zentner, A., Perleth, M., Garrido, M.V., and Janßen, I. (2014). Vorgehensweise und Methodenspektrum in der HTA-Berichterstellung. In *Health Technology Assessment: Konzepte, Methoden, Praxis für Wissenschaft und Entscheidungsfindung*, M. Perleth, R. Busse, A. Gerhardus, B. Gibis, D. Lühmann, and A. Zentner, eds. (Berlin: Medizinisch Wissenschaftliche Verlagsgesellschaft), pp. 135-161.

Zimmermann, B.M., Eichinger, J., and Baumgartner, M.R. (2021). A systematic review of moral reasons on orphan drug reimbursement. *Orphanet J Rare Dis* 16, 292.

## 9 VERÖFFENTLICHUNGEN

Schaefer, R., Hernandez, D., Selberg, L., and Schlander, M. (2021). Health technology assessment (HTA) in England, France and Germany: What do matched drug pairs tell us about recommendations by national HTA agencies? *J Comp Eff Res* *10*, 1187-1195.

Schaefer, R., and Schlander, M. (2019). Is the National Institute for Health and Care Excellence (NICE) in England more 'innovation-friendly' than the Federal Joint Committee (G-BA) in Germany? *Expert Rev Pharmacoecon Outcomes Res* *19*, 453-462.

## 10 ANHANG

Tabellarischer Anhang zu den Regressionsanalysen für GBA und IQWiG [A1–A3] sowie für NICE [A4–A6]:

Tabelle A1: Ergänzende Resultate der Regressionsanalyse für GBA auf Teilpopulationsebene .....	180
Tabelle A2: Ergänzende Resultate der Regressionsanalyse für IQWiG auf Teilpopulationsebene .....	183
Tabelle A3: Ergänzende Resultate der Regressionsanalyse für GBA und IQWiG auf Verfahrensebene .....	186
Tabelle A4: Ergänzende Resultate der Regressionsanalyse für NICE auf Subgruppenebene .....	189
Tabelle A5: Ergänzende Resultate der Regressionsanalyse für NICE auf Indikationsebene – exklusive und inklusive CDF .....	192
Tabelle A6: Ergänzende Resultate der Regressionsanalyse für NICE auf Subgruppenebene – inklusive CDF .....	195

**Tabelle A1:** Ergänzende Resultate der Regressionsanalyse für GBA auf Teilpopulationsebene  
(Eigene Darstellung)

Variable	GBA / Logit-Modell								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>[Weitere Indikatorvariablen, vgl. hierzu Tabelle 12]</i>									
Veröffentlichungsjahr 2011 (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Veröffentlichungsjahr 2012 (Dummy)	1.633	1.675	1.694	1.671	1.697	1.664	1.691	1.678	1.692
Veröffentlichungsjahr 2013 (Dummy)	2.981	2.932	2.935	2.942	2.956	2.950	2.969	2.724	2.939
Veröffentlichungsjahr 2014 (Dummy)	-0.278	-0.207	-0.156	-0.224	-0.165	-0.241	-0.182	-0.286	-0.232
Veröffentlichungsjahr 2015 (Dummy)	3.019	2.974	2.977	2.985	2.998	2.993	3.011	2.736	2.981
Veröffentlichungsjahr 2016 (Dummy)	0.701	0.781	0.842	0.761	0.839	0.743	0.822	0.713	0.754
Veröffentlichungsjahr 2017 (Dummy)	2.990	2.955	2.960	2.966	2.982	2.974	2.996	2.727	2.960
Veröffentlichungsjahr 2018 (Januar-Juni) (Dummy)	-0.227	-0.071	-0.019	-0.088	-0.017	-0.104	-0.029	-0.022	-0.054
Therapeutisches Gebiet: Augen (Dummy)	2.926	2.880	2.884	2.891	2.907	2.900	2.921	2.648	2.888
Therapeutisches Gebiet: Haut (Dummy)	-1.702	-1.715	-1.685	-1.726	-1.673	-1.738	-1.676	-1.446	-1.709
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	2.923	2.873	2.876	2.884	2.898	2.892	2.911	2.651	2.881
Therapeutisches Gebiet: ...	-1.689	-1.624	-1.565	-1.637	-1.553	-1.652	-1.561	-1.414	-1.699
Therapeutisches Gebiet: ...	2.949	2.900	2.904	2.910	2.926	2.919	2.940	2.685	2.909
Therapeutisches Gebiet: ...	-0.875	-0.851	-0.815	-0.864	-0.802	-0.878	-0.807	-0.768	-0.913
Therapeutisches Gebiet: ...	2.978	2.930	2.933	2.941	2.955	2.949	2.969	2.710	2.938
Therapeutisches Gebiet: ...	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Therapeutisches Gebiet: ...	--	--	--	--	--	--	--	--	--
Therapeutisches Gebiet: ...	4.310**	4.318**	4.390**	4.277**	4.406**	4.241**	4.387**	4.086**	4.344**
Therapeutisches Gebiet: ...	1.892	1.895	1.898	1.906	1.912	1.909	1.917	1.914	1.871
Therapeutisches Gebiet: ...	2.377	2.450	2.540	2.421	2.527	2.400	2.501	1.581	2.458
Therapeutisches Gebiet: ...	1.709	1.727	1.739	1.727	1.740	1.726	1.739	1.842	1.713

Tabelle A1: (Fortsetzung)

Variable	GBA / Logit-Modell								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Therapeutisches Gebiet: Infektionen (Dummy)	1.545	1.546	1.638	1.523	1.621	1.510	1.596	1.250	1.556
	1.509	1.523	1.537	1.519	1.531	1.516	1.526	1.523	1.511
Therapeutisches Gebiet: Atmungssystem (Dummy)	2.279	2.363	2.463	2.339	2.452	2.322	2.425	2.118	2.358
	1.631	1.643	1.658	1.643	1.657	1.642	1.654	1.636	1.635
Therapeutisches Gebiet: Blut und blutbildende Organe (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	--	--	--	--	--	--	--	--	--
Therapeutisches Gebiet: Muskel-Skelett-System (Dummy)	0.545	0.589	0.679	0.550	0.659	0.522	0.629	0.305	0.628
	1.867	1.893	1.906	1.893	1.907	1.892	1.906	1.902	1.871
Therapeutisches Gebiet: Nervensystem (Dummy)	0.642	0.674	0.700	0.655	0.673	0.642	0.651	0.529	0.671
	1.659	1.672	1.675	1.670	1.673	1.668	1.671	1.660	1.667
Therapeutisches Gebiet: Urogenital (Dummy)	3.148	3.248	3.434	3.218	3.415	3.199	3.371	2.988	3.259
	2.474	2.478	2.593	2.463	2.587	2.453	2.563	2.451	2.476
Therapeutisches Gebiet: Verdauungssystem (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	--	--	--	--	--	--	--	--	--
Therapeutisches Gebiet: Onkologie (Dummy)	2.682*	2.775*	2.855*	2.757*	2.839*	2.743*	2.815*	2.417	2.874*
	1.526	1.534	1.545	1.535	1.543	1.535	1.542	1.554	1.535
Therapeutisches Gebiet: Psychische Gesundheit (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	--	--	--	--	--	--	--	--	--
Therapeutisches Gebiet: Stoffwechselsystem (Dummy)	0.545	0.553	0.622	0.530	0.597	0.515	0.569	-0.084	0.401
	1.552	1.570	1.578	1.569	1.575	1.568	1.572	1.605	1.566
Therapeutisches Gebiet: Sonstige (Dummy)	1.500	1.487	1.603	1.516	1.657	1.536	1.665	1.093	1.503
	1.875	1.892	1.895	1.889	1.890	1.885	1.885	1.920	1.885

Tabelle A1: (Fortsetzung)

Variable	GfBA / Logit-Modell								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
<i>Potenzieller Budget Impact: Minimum (Log)</i>		-0.001	0.483						
		0.095	0.933						
<i>Potenzieller Budget Impact: Minimum (Log)<sup>2</sup></i>		-0.014							
		0.027							
<i>Potenzieller Budget Impact: Durchschnitt (Log)</i>				0.012	0.664				0.860
				0.099	1.042				1.075
<i>Potenzieller Budget Impact: Durchschnitt (Log)<sup>2</sup></i>					-0.018				-0.024
					0.029				0.030
<i>Potenzieller Budget Impact: Maximum (Log)</i>						0.021	0.687		
						0.098	1.075		
<i>Potenzieller Budget Impact: Maximum (Log)<sup>2</sup></i>							-0.019		
							0.030		
<i>Inkrementeller Budget Impact: Durchschnitt (Log)</i>								0.297	0.300
								0.228	0.230
<i>Konstante</i>	-5.445	-5.443	-9.64	-5.627	-11.442	-5.767	-11.781	-6.034**	-13.812
	3.457	3.677	8.845	3.709	9.981	3.717	10.388	3.014	10.254
Beobachtungen	493	486	486	486	486	486	486	474	474

Regressionsergebnisse sind statistisch signifikant bei: \*\*\* p < 0.01; \*\* p < 0.05; \* p < 0.1.

Anmerkung: Tabelle A1 stellt eine Ergänzung zu Tabelle 12 dar und bildet die Regressionsergebnisse der Dummy-Variablen Veröffentlichungsjahr und Therapeutisches Gebiet ab.

**Tabelle A2:** Ergänzende Resultate der Regressionsanalyse für IQWiG auf Teilpopulationsebene  
(Eigene Darstellung)

Variable	IQWiG / Logit-Modell										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Veröffentlichungsjahr 2011 (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Veröffentlichungsjahr 2012 (Dummy)	-1.476	-0.957	-1.036	-0.610	-0.621	-0.550	-0.554	-0.560	-0.538	-0.716	-0.919
Veröffentlichungsjahr 2013 (Dummy)	5.090	1.207	1.224	1.179	1.181	1.173	1.173	1.162	1.169	1.155	1.207
Veröffentlichungsjahr 2014 (Dummy)	-1.630	-1.515	-1.563	-1.301	-1.313	-1.267	-1.272	-0.702	-0.680	-1.007	-1.183
Veröffentlichungsjahr 2015 (Dummy)	5.122	1.440	1.444	1.416	1.415	1.411	1.410	1.385	1.383	1.377	1.440
Veröffentlichungsjahr 2016 (Dummy)	-2.629	-2.702**	-2.722**	-2.437**	-2.415**	-2.395**	-2.384**	-1.707	-1.684	-2.104*	-2.461**
Veröffentlichungsjahr 2017 (Dummy)	5.086	1.189	1.191	1.137	1.138	1.128	1.130	1.073	1.080	1.088	1.148
Veröffentlichungsjahr 2018 (Januar-Juni) (Dummy)	-2.523	-2.589**	-2.612**	-2.309**	-2.291**	-2.259**	-2.251**	-1.637	-1.624	-1.950*	-2.117**
Therapeutisches Gebiet: Augen (Dummy)	5.068	1.146	1.151	1.094	1.096	1.085	1.087	1.056	1.045	1.042	1.078
Therapeutisches Gebiet: Haut (Dummy)	-3.327	-3.215***	-3.263***	-2.960***	-2.958***	-2.924***	-2.922***	-2.292**	-2.275**	-2.676**	-2.996***
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	5.069	1.146	1.154	1.098	1.098	1.091	1.091	1.045	1.040	1.050	1.110
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	-2.778	-2.458**	-2.430**	-2.232*	-2.175*	-2.204*	-2.174*	-1.744	-1.744	-2.110*	-2.504**
Therapeutisches Gebiet: Augen (Dummy)	5.097	1.224	1.229	1.179	1.194	1.171	1.186	1.117	1.106	1.108	1.168
Therapeutisches Gebiet: Haut (Dummy)	-0.618	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	5.083	--	--	--	--	--	--	--	--	--	--
Therapeutisches Gebiet: Augen (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Therapeutisches Gebiet: Haut (Dummy)	4.474*	3.791	3.764	3.561	3.562	3.505	3.511	4.511*	4.540*	4.044	3.886
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	2.483	2.428	2.425	2.496	2.501	2.526	2.529	2.551	2.553	2.496	2.518
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	1.783	1.325	1.364	1.276	1.293	1.267	1.274	1.728	1.760	1.538	1.316
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	1.717	1.758	1.757	1.776	1.776	1.789	1.789	1.739	1.735	1.751	1.763

[Weitere Indikatorvariablen, vgl. hierzu Tabelle 13]



Tabelle A2: (Fortsetzung)

Variable	IQWiG / Logit-Modell										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Therapeutisches Gebiet: Infektionen (Dummy)	1.242	0.532	0.610	0.496	0.534	0.520	0.536	1.185	1.203	1.110	1.070
	1.581	1.630	1.638	1.650	1.654	1.661	1.664	1.614	1.612	1.605	1.607
Therapeutisches Gebiet: Atmungssystem (Dummy)	1.738	1.286	1.382	1.285	1.321	1.312	1.326	1.766	1.785	1.597	1.588
	1.685	1.724	1.737	1.742	1.746	1.753	1.756	1.712	1.716	1.711	1.713
Therapeutisches Gebiet: Blut und blutbildende Organe (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	--	--	--	--	--	--	--	--	--	--	--
Therapeutisches Gebiet: Muskel-Skelett-System (Dummy)	-0.202	-1.342	-1.322	-1.325	-1.323	-1.310	-1.310	-0.203	-0.185	-0.600	-0.736
	1.890	1.988	1.986	1.992	1.991	1.999	1.998	1.940	1.926	1.940	1.956
Therapeutisches Gebiet: Nervensystem (Dummy)	0.894	0.392	0.393	0.406	0.386	0.427	0.414	0.889	0.903	0.766	0.911
	1.755	1.785	1.784	1.801	1.802	1.812	1.814	1.783	1.781	1.780	1.790
Therapeutisches Gebiet: Urogenital (Dummy)	3.988	3.545	3.684	3.545	3.619	3.601	3.633	3.777	3.814	3.764	3.484
	3.395	2.985	3.095	2.896	2.941	2.890	2.910	3.302	3.322	3.126	2.856
Therapeutisches Gebiet: Verdauungssystem (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	--	--	--	--	--	--	--	--	--	--	--
Therapeutisches Gebiet: Onkologie (Dummy)	2.232	1.889	1.923	1.823	1.833	1.800	1.803	2.240	2.251	2.062	2.087
	1.588	1.618	1.617	1.640	1.640	1.654	1.655	1.617	1.615	1.619	1.625
Therapeutisches Gebiet: Psychische Gesundheit (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	--	--	--	--	--	--	--	--	--	--	--
Therapeutisches Gebiet: Stoffwechselsystem (Dummy)	1.973	1.394	1.437	1.435	1.446	1.482	1.483	1.961	1.994	2.001	1.393
	1.790	1.829	1.828	1.850	1.848	1.861	1.861	1.798	1.807	1.837	1.850
Therapeutisches Gebiet: Sonstige (Dummy)	0.519	2.044	2.254	1.891	1.999	1.806	1.848	1.166	1.194	0.755	0.500
	1.835	2.008	2.056	2.003	2.028	2.007	2.023	1.964	1.973	1.870	1.887

Tabelle A2: (Fortsetzung)

Variable	IQWiG / Logit-Modell										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
Potenzieller Budget Impact: Minimum (Log)		0.254**	0.776								
		0.116	1.158								
Potenzieller Budget Impact: Minimum (Log) <sup>2</sup>			-0.015								
			0.032								
Potenzieller Budget Impact: Durchschnitt (Log)				0.278**	0.714				0.418		
				0.129	1.377				1.392		
Potenzieller Budget Impact: Durchschnitt (Log) <sup>2</sup>				-0.012					-0.002		
				0.037					0.038		
Potenzieller Budget Impact: Maximum (Log)						0.279**	0.501				
						0.130	1.419				
Potenzieller Budget Impact: Maximum (Log) <sup>2</sup>							-0.006				
							0.038				
Inkrementeller Budget Impact: Durchschnitt (Log)								0.022	-0.167		
								0.217	0.232		
Jahrestherapiekosten insgesamt: pU-Dossier (Log)										0.126	-2.116
										0.121	1.425
Jahrestherapiekosten insgesamt: pU-Dossier (Log) <sup>2</sup>											0.062
											0.039
Konstante	-6.739	-11.071***	-15.728	-12.068***	-16.091	-12.257***	-14.331	-7.678***	-14.087	-9.552***	10.607
	5.288	3.017	10.731	3.363	13.120	3.452	13.638	2.152	13.094	3.236	13.181
Beobachtungen	488	470	470	471	471	471	471	443	443	471	471

Regressionsergebnisse sind statistisch signifikant bei: \*\*\* p < 0.01; \*\* p < 0.05; \* p < 0.1.

Die Logit-Modelle (10) und (11) bilden die Regressionsergebnisse unter Heranziehung der jeweils (zu den einzelnen IQWiG-Nutzenbewertungen) zugehörigen Daten aus den pU-Dossiers ab, die folglich (weitestgehend) konsistent sind (=> signifikant positiver Effekt von sowohl Studienevidenz (RCTs; p < 0.01) als auch der vergleichenden klinischen Wirksamkeit (patientenrelevante Endpunkte; p < 0.01) auf die IQWiG-Bewertungsempfehlungen; vgl. hierzu auch Tabelle 13), wobei die Jahrestherapiekosten insgesamt - da ausschließlich für (Teil-)Populationen mit einem potenziell vorhandenen Zusatznutzen angegeben - nur bedingt aussagekräftig sind. Anmerkung: Tabelle A2 stellt - außer für Logit-Modelle (10) und (11) - eine Ergänzung zu Tabelle 13 dar und bildet die Regressionsergebnisse der Dummy-Variablen Veröffentlichungsjahr und Therapeutisches Gebiet ab.

**Tabelle A3:** Ergänzende Resultate der Regressionsanalyse für GBA und IQWiG auf Verfahrensebene  
(Eigene Darstellung)

Variable	GBA / Ordinales Logit-Modell					IQWiG / Ordinales Logit-Modell				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
<i>[Weitere Indikatorvariablen, vgl. hierzu Tabelle 14 und 15]</i>										
Veröffentlichungsjahr 2011 (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Veröffentlichungsjahr 2012 (Dummy)	2.448	2.496	2.502	2.013	1.798	0.774	0.896	1.027	-0.312	0.492
Veröffentlichungsjahr 2013 (Dummy)	1.856	1.859	1.860	1.913	1.924	2.023	2.056	2.087	2.169	2.081
Veröffentlichungsjahr 2014 (Dummy)	1.183	1.144	1.123	0.649	0.448	1.081	0.552	0.583	0.580	0.833
Veröffentlichungsjahr 2015 (Dummy)	1.874	1.876	1.878	1.938	1.949	2.021	2.057	2.082	2.103	2.085
Veröffentlichungsjahr 2016 (Dummy)	0.915	0.885	0.855	0.437	0.199	1.171	0.525	0.778	0.440	0.870
Veröffentlichungsjahr 2017 (Dummy)	1.886	1.892	1.896	1.952	1.964	2.036	2.082	2.109	2.127	2.089
Veröffentlichungsjahr 2018 (Januar-Juni) (Dummy)	1.262	1.211	1.183	1.030	0.753	0.953	0.495	0.722	0.493	0.782
Veröffentlichungsjahr 2019 (Dummy)	1.815	1.817	1.820	1.870	1.879	1.963	2.002	2.032	2.041	2.011
Veröffentlichungsjahr 2020 (Dummy)	0.504	0.489	0.463	0.266	-0.113	0.919	0.536	0.694	0.512	0.821
Veröffentlichungsjahr 2021 (Dummy)	1.812	1.813	1.816	1.863	1.881	1.973	2.005	2.033	2.057	2.024
Veröffentlichungsjahr 2022 (Dummy)	0.212	0.191	0.164	0.069	-0.335	1.322	1.033	1.302	0.818	1.126
Veröffentlichungsjahr 2023 (Dummy)	1.839	1.840	1.844	1.895	1.904	2.009	2.050	2.080	2.098	2.061
Veröffentlichungsjahr 2024 (Januar-Juni) (Dummy)	1.248	1.201	1.177	0.943	0.613	1.567	1.251	1.433	0.948	1.495
Veröffentlichungsjahr 2025 (Dummy)	1.874	1.876	1.879	1.924	1.936	2.026	2.066	2.095	2.112	2.081
Therapeutisches Gebiet: Augen (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Therapeutisches Gebiet: Haut (Dummy)	3.907***	3.711**	3.663**	3.638**	3.830***	4.470***	3.162**	3.168**	4.732***	4.380***
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	1.473	1.486	1.496	1.478	1.467	1.596	1.586	1.585	1.632	1.609
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	2.725*	2.608*	2.586*	1.811	2.058	2.643*	1.837	1.670	2.155	2.458
Therapeutisches Gebiet: Augen (Dummy)	1.449	1.446	1.447	1.531	1.500	1.545	1.503	1.503	1.617	1.567

Tabelle A3: (Fortsetzung)

Variable	GBA / Ordinales Logit-Modell					IQWiG / Ordinales Logit-Modell				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
Therapeutisches Gebiet: Infektionen (Dummy)	1.772	1.598	1.521	1.518	1.763	1.897	0.699	0.909	2.245	1.798
	1.336	1.346	1.379	1.334	1.326	1.409	1.412	1.406	1.461	1.421
Therapeutisches Gebiet: Atmungssystem (Dummy)	2.545*	2.367	2.328	2.428*	2.562*	1.581	0.557	0.410	1.755	1.472
	1.438	1.449	1.456	1.435	1.434	1.543	1.527	1.529	1.580	1.559
Therapeutisches Gebiet: Blut und blutbildende Organe (Dummy)	-11.905	-12.876	-12.919	-13.096	-12.414	-12.343	-13.674	-13.845	-12.095	-12.475
	1183.586	1727.547	1728.517	1742.529	1343.071	1761.500	1991.594	1993.050	1762.270	1756.521
Therapeutisches Gebiet: Muskel-Skelett-System (Dummy)	2.419	2.202	2.157	1.797	1.892	2.845*	1.499	1.495	3.135*	2.759
	1.622	1.633	1.639	1.674	1.681	1.698	1.682	1.688	1.742	1.694
Therapeutisches Gebiet: Nervensystem (Dummy)	1.227	1.185	1.170	1.062	1.207	1.299	0.550	0.319	1.595	1.217
	1.493	1.495	1.493	1.483	1.488	1.655	1.618	1.623	1.702	1.670
Therapeutisches Gebiet: Urogenital (Dummy)	2.871	2.755	2.718	2.373	2.728	2.853	2.447	2.664	3.058	2.811
	1.827	1.826	1.834	1.838	1.822	1.895	1.922	1.914	1.907	1.899
Therapeutisches Gebiet: Verdauungssystem (Dummy)	-12.253	-13.404	-13.699	-15.692	-55.435	-12.155	-13.019	-13.253	-12.079	-12.216
	1902.410	2751.930	2715.879	2433.041	1358.088	2797.754	3171.295	3170.879	2789.497	2793.574
Therapeutisches Gebiet: Onkologie (Dummy)	2.878**	2.776**	2.762**	2.457*	2.785**	2.832**	2.241*	2.043	3.082**	2.877**
	1.309	1.301	1.301	1.331	1.312	1.388	1.335	1.333	1.417	1.399
Therapeutisches Gebiet: Psychische Gesundheit (Dummy)	-13.962	-14.853	-14.875	-15.482	-14.497	-14.883	-16.094	-16.427	-14.776	-15.063
	1200.365	1744.089	1743.296	1781.744	1392.149	1919.591	2146.946	2109.022	1902.551	1933.885
Therapeutisches Gebiet: Stoffwechselsystem (Dummy)	1.375	1.206	1.176	0.363	0.806	2.215	1.331	1.307	2.312	2.133
	1.384	1.387	1.390	1.436	1.402	1.474	1.426	1.417	1.500	1.486
Therapeutisches Gebiet: Sonstige (Dummy)	2.177	2.186	2.164	1.711	1.979	2.729	2.978*	3.190*	3.097*	2.927*
	1.723	1.702	1.704	1.736	1.711	1.723	1.655	1.639	1.779	1.766

Tabelle A3: (Fortsetzung)

Variable	GBA / Ordinales Logit-Modell					IQWiG / Ordinales Logit-Modell				
	(1)	(2)	(3)	(4)	(5)	(1)	(2)	(3)	(4)	(5)
<i>Potenzieller Budget Impact: Durchschnitt (Log)</i>		0.065	-0.180		0.105		0.394***	4.841**		4.795**
		0.097	0.979		1.099		0.137	2.127		2.158
<i>Potenzieller Budget Impact: Durchschnitt (Log)<sup>2</sup></i>			0.007		-0.001			-0.112**		-0.110**
			0.027		0.029			0.053		0.054
<i>Inkrementeller Budget Impact: Durchschnitt (Log)</i>				0.469**	0.459**				0.125	-0.016
				0.216	0.217				0.201	0.209
<i>Konstante 1</i>	3.658	4.603*	2.356	4.129*	5.564	5.954**	12.162***	55.984***	5.091**	55.326**
	2.347	2.741	9.348	2.285	10.671	2.639	3.527	21.410	2.490	21.681
<i>Konstante 2</i>	5.915**	6.862**	4.616	6.570***	8.012	8.761***	15.106***	59.033***	8.025***	58.545***
	2.385	2.778	9.356	2.333	10.685	2.712	3.627	21.498	2.571	21.786
Beobachtungen	209	207	207	198	198	205	204	204	194	194

Regressionsergebnisse sind statistisch signifikant bei: \*\*\* p < 0.01; \*\* p < 0.05; \* p < 0.1.

Anmerkung: Tabelle A3 stellt eine Ergänzung zu Tabelle 14 / 15 dar und bildet die Regressionsergebnisse der Dummy-Variablen Veröffentlichungsjahr und Therapeutisches Gebiet ab.

**Tabelle A4:** Ergänzende Resultate der Regressionsanalyse für NICE auf Subgruppenebene  
(Eigene Darstellung)

Ordinales Logit-Modell														
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
<i>[Weitere Indikatorvariablen, vgl. hierzu Tabelle 20]</i>														
Veröffentlichungsjahr 2011 (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Veröffentlichungsjahr 2012 (Dummy)	0.499	1.464	0.857	0.762	0.520	0.336	0.460	0.279	0.501	0.364	0.503	0.470	0.383	0.352
Veröffentlichungsjahr 2013 (Dummy)	0.710	1.285	0.726	0.732	0.719	0.727	0.719	0.728	0.718	0.727	0.718	0.711	0.723	0.714
Veröffentlichungsjahr 2014 (Dummy)	-0.661	-12.760	-0.159	-0.334	-0.604	-0.935	-0.733	-1.051	-0.645	-0.901	-0.607	-0.618	-0.604	-0.616
Veröffentlichungsjahr 2015 (Dummy)	0.764	1626.519	0.798	0.823	0.799	0.830	0.802	0.832	0.794	0.826	0.778	0.783	0.779	0.785
Veröffentlichungsjahr 2016 (Dummy)	1.420**	2.091	1.763**	1.626**	1.448**	1.181	1.378*	1.119	1.425*	1.237*	1.490**	1.265*	1.488**	1.248*
Veröffentlichungsjahr 2017 (Dummy)	0.721	1.718	0.746	0.756	0.733	0.744	0.732	0.745	0.727	0.739	0.726	0.724	0.732	0.730
Veröffentlichungsjahr 2018 (Dummy)	1.177**	1.841	1.587***	1.457**	1.213**	0.949	1.122*	0.871	1.185*	1.013	1.262**	1.263**	1.384**	1.397**
Veröffentlichungsjahr 2019 (Dummy)	0.585	1.139	0.616	0.630	0.609	0.626	0.609	0.627	0.609	0.623	0.597	0.600	0.602	0.606
Veröffentlichungsjahr 2020 (Dummy)	0.745	0.569	1.084*	0.994*	0.766	0.560	0.679	0.484	0.741	0.622	0.850	0.904	0.857	0.925
Veröffentlichungsjahr 2021 (Dummy)	0.551	0.962	0.569	0.577	0.571	0.583	0.575	0.588	0.581	0.589	0.573	0.574	0.585	0.586
Veröffentlichungsjahr 2022 (Dummy)	0.299	-0.335	0.636	0.553	0.341	0.150	0.264	0.080	0.319	0.222	0.427	0.360	0.495	0.434
Veröffentlichungsjahr 2023 (Dummy)	0.513	0.937	0.531	0.538	0.531	0.543	0.535	0.547	0.531	0.537	0.529	0.532	0.535	0.537
Veröffentlichungsjahr 2024 (Januar-Juni) (Dummy)	0.357	0.205	0.661	0.623	0.379	0.272	0.297	0.198	0.356	0.294	0.363	0.450	0.384	0.485
Veröffentlichungsjahr 2025 (Juli-Dezember) (Dummy)	0.608	1.054	0.624	0.625	0.624	0.628	0.627	0.632	0.628	0.630	0.641	0.646	0.645	0.652
Therapeutisches Gebiet: Augen (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Therapeutisches Gebiet: Haut (Dummy)	-2.464**		-2.190**	-2.394**	-2.408**	-2.798***	-2.513**	-2.901***	-2.434**	-2.506**	-2.391**	-2.513**	-2.461**	-2.591**
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	1.028	0.000	1.051	1.077	1.046	1.076	1.047	1.078	1.084	1.089	1.073	1.084	1.079	1.091
Therapeutisches Gebiet: Krebs (Dummy)	-0.454	0.000	-0.140	-0.343	-0.400	-0.792	-0.503	-0.891	-0.432	-0.700	-0.357	-0.343	-0.329	-0.315
Therapeutisches Gebiet: Neurologie (Dummy)	0.775	--	0.797	0.830	0.798	0.835	0.798	0.837	0.797	0.830	0.813	0.813	0.814	0.814

Tabelle A4: (Fortsetzung)

Ordinales Logit-Modell														
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Therapeutisches Gebiet: Infektionen (Dummy)	-2.875***		-2.539***	-2.650***	-2.818***	-3.040***	-2.950***	-3.182***	-2.861***	-3.058***	-2.844***	-2.849***	-2.946***	-2.955***
Therapeutisches Gebiet: Atmungssystem (Dummy)	0.828		0.864	0.877	0.862	0.877	0.863	0.879	0.852	0.871	0.876	0.879	0.881	0.884
Therapeutisches Gebiet: Blut/-bildende Organe (Dummy)	-1.601		-1.312	-1.548	-1.548	-2.025*	-1.680	-2.128*	-1.587	-1.877	-1.496	-1.630	-1.540	-1.677
Therapeutisches Gebiet: Muskel-Skelett-System (Dummy)	1.121		1.221	1.252	1.144	1.182	1.149	1.186	1.149	1.176	1.173	1.182	1.178	1.189
Therapeutisches Gebiet: Nervensystem (Dummy)	-1.625		-1.486	-1.658	-1.596	-2.041*	-1.637	-2.105*	-1.599	-1.978	-1.579	-1.645	-1.596	-1.673
Therapeutisches Gebiet: Urogenital (Dummy)	1.187		6241.893	1.214	1.191	1.224	1.189	1.229	1.218	1.263	1.220	1.229	1.225	1.234
Therapeutisches Gebiet: Verdauungssystem (Dummy)	-1.585**		-1.137	-1.332	-1.519*	-1.874**	-1.677**	-2.020**	-1.568*	-1.815**	-1.426*	-1.532*	-1.420	-1.518*
Therapeutisches Gebiet: Onkologie (Dummy)	0.796		0.850	0.877	0.848	0.876	0.848	0.877	0.854	0.878	0.859	0.859	0.870	0.870
Therapeutisches Gebiet: Psychische Gesundheit (Dummy)	-0.858		-0.725	-0.908	-0.798	-1.159	-0.923	-1.277	-0.835	-1.062	-0.784	-0.614	-0.705	-0.520
Therapeutisches Gebiet: Sonstige (Dummy)	1.211		1.271	1.287	1.230	1.248	1.235	1.253	1.235	1.248	1.245	1.237	1.252	1.241
Therapeutisches Gebiet: Stoffwechselsystem (Dummy)	-1.192		-0.746	-0.631	-1.125	-0.983	-1.298	-1.201	-1.177	-0.907	-1.024	-1.105	-1.020	-1.088
Therapeutisches Gebiet: Atmungssystem (Dummy)	2.152		2.204	2.212	2.176	2.191	2.175	2.188	2.188	2.205	2.208	2.241	2.227	2.263
Therapeutisches Gebiet: Verdauungssystem (Dummy)	-1.654		-1.304	-1.369	-1.601	-1.705	-1.767	-1.852*	-1.651	-1.706	-1.778	-1.802	-1.838	-1.856
Therapeutisches Gebiet: Urogenital (Dummy)	1.054		1.099	1.105	1.096	1.103	1.100	1.105	1.104	1.108	1.174	1.184	1.178	1.188
Therapeutisches Gebiet: Psychische Gesundheit (Dummy)	-1.389**		-1.156*	-1.382*	-1.341*	-1.750**	-1.433**	-1.828**	-1.369**	-1.616**	-1.349*	-1.497**	-1.361*	-1.523**
Therapeutisches Gebiet: Stoffwechselsystem (Dummy)	0.669		4413.686	0.689	0.734	0.690	0.691	0.735	0.692	0.723	0.737	0.740	0.738	0.741
Therapeutisches Gebiet: Atmungssystem (Dummy)	0.212		0.303	0.217	0.216	0.030	0.219	-0.006	0.216	0.175	0.404	0.424	0.337	0.365
Therapeutisches Gebiet: Blut/-bildende Organe (Dummy)	1.891		1.927	2.003	1.889	2.004	1.896	1.986	1.893	2.001	1.949	1.948	1.963	1.964
Therapeutisches Gebiet: Muskel-Skelett-System (Dummy)	-2.227**		-1.847*	-1.881**	-2.164**	-2.268**	-2.315**	-2.445**	-2.212**	-2.335**	-2.116**	-2.098**	-2.175**	-2.148**
Therapeutisches Gebiet: Verdauungssystem (Dummy)	0.907		0.945	0.946	0.947	0.948	0.950	0.951	0.944	0.948	0.956	0.958	0.960	0.962
Therapeutisches Gebiet: Urogenital (Dummy)	-1.047		-1.079	-1.203	-1.040	-1.269	-1.071	-1.279	-1.048	-1.181	-0.709	-0.714	-0.144	-0.133
Therapeutisches Gebiet: Sonstige (Dummy)	0.980		0.986	0.999	0.981	0.991	0.980	0.989	0.982	0.985	1.070	1.070	1.092	1.089

Tabelle A4: (Fortsetzung)

Ordinales Logit-Modell														
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Potenzieller Budget Impact: Minimum (Log)			-0.137	0.672										
Potenzieller Budget Impact: Minimum (Log) <sup>2</sup>			0.088	0.863										
Potenzieller Budget Impact: Durchschnitt (Log)					-0.020	1.508*						1.639*		
Potenzieller Budget Impact: Durchschnitt (Log) <sup>2</sup>					0.092	0.887						0.914		
Potenzieller Budget Impact: Maximum (Log)						-0.046*						-0.049*		
Potenzieller Budget Impact: Maximum (Log) <sup>2</sup>						0.026						0.027		
Potenzieller Budget Impact: Maximum (Log)						0.030	1.473*							
Potenzieller Budget Impact: Maximum (Log) <sup>2</sup>						0.093	0.874							
Potenzieller Budget Impact: NHS-Population (Log)							-0.043*							
Potenzieller Budget Impact: NHS-Population (Log) <sup>2</sup>							0.026							
Potenzieller Budget Impact: NHS-Population (Log)									-0.004	1.032				0.800
Potenzieller Budget Impact: NHS-Population (Log) <sup>2</sup>									0.089	0.837				0.883
Potenzieller Budget Impact: NHS-Population (Log)										-0.030				-0.024
Potenzieller Budget Impact: NHS-Population (Log) <sup>2</sup>										0.024				0.025
Inkrementeller Budget Impact: Durchschnitt (Log)											0.116	0.126		
Inkrementeller Budget Impact: NHS-Population (Log)											0.123	0.125		
Intercept 1	-2.347	-17.652	-3.409	2.672	-2.590	9.701	-2.018	9.656	-2.412	6.379	-3.060	9.878	-3.021	3.540
Intercept 2	2.414	4413.685	2.554	6.938	2.626	7.582	2.640	7.525	2.615	7.536	2.551	7.706	2.572	7.812
Intercept 2	1.276	-14.967	0.288	6.381	1.031	13.359*	1.606	13.313*	1.209	10.020	0.684	13.673*	0.770	7.345
Intercept 2	2.407	4413.685	2.540	6.950	2.620	7.613	2.639	7.557	2.609	7.559	2.541	7.744	2.561	7.829
Beobachtungen	255	85	251	251	254	254	254	254	254	254	235	235	232	232

ICER, Incremental Cost-Effectiveness Ratio; NHS, National Health Service; QALY, Quality-Adjusted Life Year; RCT, Randomized Controlled Trial.

Regressionsergebnisse sind statistisch signifikant bei: \*\*\* p < 0.01; \*\* p < 0.05; \* p < 0.1.

Anmerkung: Tabelle A4 stellt eine Ergänzung zu Tabelle 20 dar und bildet die Regressionsergebnisse der Dummy-Variablen Veröffentlichungsjahr und Therapeutisches Gebiet ab.



**Tabelle A5:** Ergänzende Resultate der Regressionsanalyse für NICE auf Indikationsebene – exklusive und inklusive CDF  
(Eigene Darstellung)

Variable	Ordinales Logit-Modell						Ordinales Logit-Modell unter Berücksichtigung des Cancer Drugs Fund					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
<i>[Weitere Indikatorvariablen, vgl. hierzu Tabelle 21 und 23]</i>												
Veröffentlichungsjahr 2011 (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Veröffentlichungsjahr 2012 (Dummy)	0.158	2.219	0.241	0.053	0.171	0.106	0.320	2.408	0.413	0.216	0.288	0.250
Veröffentlichungsjahr 2013 (Dummy)	0.805	1.555	0.810	0.821	0.809	0.805	0.798	1.580	0.802	0.818	0.809	0.807
Veröffentlichungsjahr 2014 (Dummy)	-1.064	-12.927	-0.850	-1.154	-1.132	-1.149	-0.721	-15.162	-0.528	-0.839	-1.011	-1.015
Veröffentlichungsjahr 2015 (Dummy)	0.878	2665.806	0.898	0.913	0.897	0.900	0.872	9635.131	0.889	0.912	0.897	0.897
Veröffentlichungsjahr 2016 (Dummy)	0.952	2.851	1.035	0.840	1.021	0.739	1.264	3.582*	1.350*	1.128	1.189	1.040
Veröffentlichungsjahr 2017 (Dummy)	0.798	1.955	0.807	0.810	0.803	0.803	0.812	2.023	0.819	0.825	0.820	0.821
Veröffentlichungsjahr 2018 (Januar-Juni) (Dummy)	0.730	2.589*	0.830	0.642	0.813	0.786	1.008	2.207	1.105*	0.914	0.958	0.943
Therapeutisches Gebiet: Augen (Dummy)	0.660	1.412	0.667	0.677	0.672	0.676	0.660	1.420	0.666	0.679	0.672	0.671
Therapeutisches Gebiet: Haut (Dummy)	0.219	1.344	0.318	0.126	0.360	0.373	0.714	2.495*	0.841	0.647	0.992	1.002
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	0.610	1.293	0.619	0.629	0.631	0.631	0.615	1.332	0.622	0.638	0.642	0.639
Therapeutisches Gebiet: Krebs (Dummy)	-0.287	0.496	-0.158	-0.298	-0.149	-0.256	0.938	2.963**	1.007	0.882	1.041	1.008
Therapeutisches Gebiet: Diabetes (Dummy)	0.606	1.266	0.615	0.622	0.624	0.629	0.611	1.320	0.619	0.630	0.635	0.637
Therapeutisches Gebiet: Infektionskrankheiten (Dummy)	-0.114	1.196	-0.029	-0.122	-0.139	-0.102	0.438	2.429*	0.551	0.453	0.200	0.220
Therapeutisches Gebiet: Neurologie (Dummy)	0.652	1.331	0.658	0.663	0.677	0.682	0.651	1.365	0.656	0.666	0.678	0.679
Therapeutisches Gebiet: Onkologie (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Therapeutisches Gebiet: Schmerzmittel (Dummy)	-2.189**		-1.999*	-2.393**	-2.295**	-2.378**	-2.403**		-2.252**	-2.674**	-2.903***	-2.934***
Therapeutisches Gebiet: Spezialpräparate (Dummy)	1.024		1.038	1.061	1.072	1.079	1.040		1.051	1.088	1.094	1.095
Therapeutisches Gebiet: Spezialpräparate (Dummy)	-0.230	0.000	-0.025	-0.395	-0.344	-0.316	0.131	0.000	0.308	-0.078	-0.382	-0.365
Therapeutisches Gebiet: Spezialpräparate (Dummy)	0.849	--	0.870	0.892	0.888	0.885	0.862	--	0.879	0.913	0.903	0.901

Tabelle A5: (Fortsetzung)

Variable	Ordinales Logit-Modell					Ordinales Logit-Modell unter Berücksichtigung des Cancer Drugs Fund						
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
Therapeutisches Gebiet: Infektionen (Dummy)	-2.181*		-1.606	-1.480	-2.261*	-2.301**	-1.883*		-1.376	-1.260	-2.413**	-2.423**
	1.120		1.215	1.223	1.162	1.166	1.139		1.238	1.250	1.194	1.194
Therapeutisches Gebiet: Atmungssystem (Dummy)	-1.333		-1.093	-1.597	-1.422	-1.531	-1.434		-1.184	-1.722	-1.944	-1.989*
	1.114		1.136	1.168	1.167	1.175	1.133		1.153	1.200	1.186	1.186
Therapeutisches Gebiet: Blut/ -bildende Organe (Dummy)	-1.505	1.267	-1.524	-1.814	-1.692	-1.696	-1.292	0.124	-1.304	-1.632	-1.756	-1.755
	1.296	8755.920	1.309	1.320	1.324	1.324	1.270	3404.052	1.278	1.295	1.304	1.297
Therapeutisches Gebiet: Muskel-Skelett-System (Dummy)	-1.530*		-1.174	-1.481	-1.519	-1.634*	-1.100		-0.762	-1.088	-1.603	-1.647*
	0.907		0.957	0.968	0.967	0.970	0.927		0.977	1.005	0.981	0.981
Therapeutisches Gebiet: Nervensystem (Dummy)	-0.986		-0.758	-1.173	-1.116	-0.914	-0.461		-0.262	-0.680	-0.927	-0.819
	1.244		1.257	1.280	1.282	1.270	1.276		1.289	1.321	1.317	1.312
Therapeutisches Gebiet: Urogenital (Dummy)	-1.168		-0.757	-0.702	-1.058	-1.190	-0.412		-0.023	0.026	-0.643	-0.689
	1.986		2.019	2.033	2.031	2.054	2.003		2.043	2.060	2.070	2.071
Therapeutisches Gebiet: Verdauungssystem (Dummy)	-1.601		-1.246	-1.377	-1.920	-1.926	-1.586		-1.184	-1.311	-1.606	-1.594
	1.069		1.115	1.121	1.182	1.189	1.106		1.155	1.172	1.200	1.202
Therapeutisches Gebiet: Onkologie (Dummy)	-1.242*	-17.928	-1.062	-1.482*	-1.439*	-1.516*	-0.345	-15.669	-0.170	-0.606	-0.928	-0.954
	0.746	6191.371	0.766	0.794	0.811	0.807	0.760	2407.029	0.774	0.820	0.813	0.810
Therapeutisches Gebiet: Psychische Gesundheit (Dummy)	-0.238		-0.355	-0.331	-0.233	-0.227	0.141		0.057	0.050	-0.216	-0.211
	1.808		1.769	1.981	1.861	1.865	1.663		1.639	1.774	1.704	1.703
Therapeutisches Gebiet: Stoffwechselsystem (Dummy)	-1.445		-1.056	-1.026	-1.539	-1.543	-1.015		-0.643	-0.635	-1.443	-1.430
	1.090		1.132	1.133	1.139	1.140	1.109		1.150	1.152	1.158	1.156
Therapeutisches Gebiet: Sonstige (Dummy)	-0.763		-0.684	-0.943	-0.584	-0.513	-0.857		-0.778	-1.045	-0.609	-0.568
	1.089		1.108	1.102	1.198	1.200	1.123		1.122	1.124	1.193	1.191

**Tabelle A5:** (Fortsetzung)

Variable	Ordinales Logit-Modell						Ordinales Logit-Modell unter Berücksichtigung des Cancer Drugs Fund					
	(1)	(2)	(3)	(4)	(5)	(6)	(1)	(2)	(3)	(4)	(5)	(6)
<i>Potenzieller Budget Impact: Durchschnitt (Log)</i>			-0.132	2.024*		2.326**			-0.117	2.051*		2.707**
			0.108	1.104		1.136			0.110	1.099		1.148
<i>Potenzieller Budget Impact: Durchschnitt (Log)<sup>2</sup></i>				-0,063*		-0,071**				-0,063**		-0,082**
				0,032		0,033				0,032		0,033
<i>Inkrementeller Budget Impact: Durchschnitt (Log)</i>					0.138	0.149					0.061	0.073
					0.126	0.127					0.127	0.129
<i>Intercept 1</i>	-2.415	-17.761	-4.131	13.851	-2.956	15.795	-5.178**	-18.914	-6.527**	11.408	-4.256*	17.633*
	2.301	6191.370	2.698	9.533	2.435	9.721	2.375	2407.029	2.738	9.433	2.465	9.799
<i>Intercept 2</i>	0.597	-15.234	-1.110	16.925*	0.180	19.013*	-2.114	-16.236	-3.442	14.558	-1.004	20.997**
	2.292	6191.370	2.681	9.567	2.422	9.764	2.349	2407.029	2.706	9.473	2.443	9.860
Beobachtungen	198	79	197	197	180	180	198	79	197	197	180	180

ICER, Incremental Cost-Effectiveness Ratio; QALY, Quality-Adjusted Life Year; RCT, Randomized Controlled Trial.

Regressionsergebnisse sind statistisch signifikant bei: \*\*\* p < 0.01; \*\* p < 0.05; \* p < 0.1.

Anmerkung: Tabelle A5 stellt eine Ergänzung zu Tabelle 21 / 23 dar und bildet die Regressionsergebnisse der Dummy-Variablen Veröffentlichungsjahr und Therapeutisches Gebiet ab.

**Tabelle A6:** Ergänzende Resultate der Regressionsanalyse für NICE auf Subgruppenebene – inklusive CDF  
(Eigene Darstellung)

Ordinales Logit-Modell unter Berücksichtigung des Cancer Drugs Fund (CDF)														
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
<i>[Weitere Indikatorvariablen, vgl. hierzu Tabelle 22]</i>														
Veröffentlichungsjahr 2011 (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Veröffentlichungsjahr 2012 (Dummy)	0.815	1.940	1.063	0.972	0.834	0.692	0.812	0.678	0.956	0.781	0.753	0.721	0.630	0.595
Veröffentlichungsjahr 2013 (Dummy)	0.718	1.330	0.733	0.740	0.728	0.737	0.728	0.738	0.731	0.743	0.729	0.725	0.734	0.728
Veröffentlichungsjahr 2014 (Dummy)	-0.323	-14.870	-0.062	-0.223	-0.323	-0.570	-0.364	-0.598	-0.127	-0.478	-0.662	-0.664	-0.659	-0.661
Veröffentlichungsjahr 2015 (Dummy)	0.776	6605.532	0.807	0.825	0.808	0.828	0.809	0.830	0.802	0.829	0.800	0.801	0.803	0.804
Veröffentlichungsjahr 2016 (Dummy)	1.666**	3.249*	1.877**	1.742**	1.678**	1.461*	1.653**	1.448*	1.786**	1.511**	1.637**	1.509**	1.618**	1.466*
Veröffentlichungsjahr 2017 (Dummy)	0.741	1.828	0.762	0.770	0.755	0.766	0.755	0.767	0.750	0.761	0.750	0.746	0.754	0.750
Veröffentlichungsjahr 2018 (Januar-Juni) (Dummy)	1.740***	1.620	1.997***	1.873***	1.755***	1.549**	1.723***	1.532**	1.930***	1.692***	1.735***	1.734***	1.823***	1.830***
Veröffentlichungsjahr 2019 (Dummy)	0.597	1.163	0.624	0.636	0.622	0.638	0.622	0.639	0.625	0.638	0.608	0.607	0.615	0.614
Veröffentlichungsjahr 2020 (Dummy)	1.302**	2.140**	1.561***	1.478**	1.337**	1.180**	1.309**	1.163*	1.518**	1.362**	1.579***	1.605***	1.610***	1.650***
Veröffentlichungsjahr 2021 (Dummy)	0.562	1.029	0.578	0.585	0.582	0.594	0.586	0.599	0.592	0.602	0.592	0.590	0.604	0.602
Veröffentlichungsjahr 2022 (Dummy)	1.458***	2.628***	1.651***	1.575***	1.436***	1.293**	1.411**	1.276**	1.559***	1.452***	1.574***	1.556***	1.527***	1.512***
Veröffentlichungsjahr 2023 (Januar-Juni) (Dummy)	0.528	1.008	0.544	0.551	0.544	0.555	0.548	0.560	0.544	0.552	0.554	0.554	0.559	0.560
Veröffentlichungsjahr 2024 (Januar-Juni) (Dummy)	0.929	1.760	1.173*	1.145*	0.969	0.892	0.941	0.870	1.125*	1.061*	0.698	0.738	0.717	0.770
Therapeutisches Gebiet: Augen (Dummy)	0.608	1.099	0.623	0.625	0.626	0.631	0.630	0.635	0.629	0.634	0.647	0.647	0.652	0.653
Therapeutisches Gebiet: Haut (Dummy)	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	-2.587**	-2.542**	-2.757**	-2.617**	-2.939***	-2.649**	-2.961***	-2.961***	-2.286**	-2.385**	-3.193***	-3.240***	-3.221***	-3.279***
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	1.052	1.072	1.098	1.098	1.069	1.096	1.069	1.097	1.104	1.109	1.116	1.118	1.123	1.126
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	-0.044	0.000	0.049	-0.162	-0.063	-0.383	-0.097	-0.405	0.120	-0.269	-0.655	-0.641	-0.627	-0.612
Therapeutisches Gebiet: Herz-Kreislauf (Dummy)	0.786	--	0.809	0.839	0.809	0.841	0.809	0.843	0.808	0.841	0.838	0.837	0.838	0.837

Tabelle A6: (Fortsetzung)

Ordinales Logit-Modell unter Berücksichtigung des Cancer Drugs Fund (CDF)														
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Therapeutisches Gebiet: Infektionen (Dummy)	-2.759***		-2.657***	-2.773***	-2.755***	-2.936***	-2.798***	-2.982***	-2.567***	-2.864***	-3.462***	-3.452***	-3.513***	-3.505***
Therapeutisches Gebiet: Atmungssystem (Dummy)	0.843		0.881	0.892	0.877	0.889	0.877	0.890	0.863	0.885	0.915	0.915	0.919	0.919
Therapeutisches Gebiet: Blut/-bildende Organe (Dummy)	-1.479		-1.384	-1.632	-1.472	-1.868	-1.516	-1.878	-1.225	-1.663	-2.114*	-2.169*	-2.118*	-2.179*
Therapeutisches Gebiet: Muskel+Skelett-System (Dummy)	1.138		1.234	1.263	1.162	1.197	1.167	1.200	1.168	1.200	1.203	1.203	1.211	1.213
Therapeutisches Gebiet: Nervensystem (Dummy)	-1.201		-1.155	-1.328	-1.235	-1.592	-1.248	-1.616	-0.966	-1.510	-1.775	-1.798	-1.803	-1.838
Therapeutisches Gebiet: Urogenital (Dummy)	1.187		2559.538	1.205	1.190	1.214	1.189	1.221	1.222	1.275	1.225	1.224	1.232	1.231
Therapeutisches Gebiet: Verdauungssystem (Dummy)	-1.027		-0.848	-1.044	-1.015	-1.299	-1.068	-1.335	-0.714	-1.074	-1.674*	-1.713*	-1.656*	-1.695*
Therapeutisches Gebiet: Onkologie (Dummy)	0.815		0.868	0.892	0.865	0.888	0.864	0.887	0.871	0.895	0.889	0.885	0.900	0.896
Therapeutisches Gebiet: Psychische Gesundheit (Dummy)	-0.208		-0.292	-0.477	-0.222	-0.508	-0.263	-0.537	0.008	-0.310	-0.758	-0.653	-0.678	-0.553
Therapeutisches Gebiet: Stoffwechselsystem (Dummy)	1.246		1.309	1.324	1.268	1.284	1.270	1.286	1.264	1.276	1.294	1.291	1.298	1.292
Therapeutisches Gebiet: Sonstige (Dummy)	-0.311		-0.131	-0.003	-0.281	-0.155	-0.342	-0.255	0.139	0.563	-0.596	-0.621	-0.615	-0.632
Therapeutisches Gebiet: Sonstige (Dummy)	2.205		2.246	2.257	2.236	2.251	2.234	2.247	2.258	2.288	2.291	2.294	2.315	2.321
Therapeutisches Gebiet: Sonstige (Dummy)	-1.582		-1.430	-1.479	-1.528	-1.591	-1.584	-1.635	-1.207	-1.285	-1.708	-1.705	-1.755	-1.747
Therapeutisches Gebiet: Sonstige (Dummy)	1.101		1.139	1.141	1.140	1.139	1.144	1.142	1.147	1.148	1.210	1.211	1.214	1.216
Therapeutisches Gebiet: Sonstige (Dummy)	-0.046		-14.665	-0.225	-0.061	-0.392	-0.092	-0.403	0.112	-0.248	-0.789	-0.843	-0.832	-0.899
Therapeutisches Gebiet: Sonstige (Dummy)	0.676		1809.867	0.699	0.739	0.697	0.697	0.735	0.698	0.727	0.751	0.751	0.752	0.752
Therapeutisches Gebiet: Sonstige (Dummy)	0.650		0.707	0.632	0.636	0.499	0.639	0.471	0.571	0.550	0.162	0.171	0.127	0.143
Therapeutisches Gebiet: Sonstige (Dummy)	1.772		1.802	1.876	1.768	1.846	1.770	1.829	1.756	1.892	1.828	1.823	1.846	1.843
Therapeutisches Gebiet: Sonstige (Dummy)	-1.697*		-1.572	-1.609*	-1.688*	-1.770*	-1.737*	-1.837*	-1.438	-1.589	-2.265**	-2.238**	-2.303**	-2.266**
Therapeutisches Gebiet: Sonstige (Dummy)	0.931		0.968	0.966	0.969	0.966	0.971	0.969	0.963	0.967	0.997	0.994	1.000	0.997
Therapeutisches Gebiet: Sonstige (Dummy)	-0.792		-0.835	-0.962	-0.794	-0.984	-0.798	-0.971	-0.759	-0.970	-0.771	-0.751	-0.171	-0.143
Therapeutisches Gebiet: Sonstige (Dummy)	1.033		1.031	1.040	1.031	1.031	1.031	1.030	1.036	1.030	1.090	1.089	1.101	1.098

Tabelle A6: (Fortsetzung)

Ordinales Logit-Modell unter Berücksichtigung des Cancer Drugs Fund (CDF)														
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
<i>Potenzieller Budget Impact: Minimum (Log)</i>			-0.054	0.868										
			0.090	0.893										
<i>Potenzieller Budget Impact: Minimum (Log)<sup>2</sup></i>				-0.028										
				0.027										
<i>Potenzieller Budget Impact: Durchschnitt (Log)</i>					-0.006	1.333						1.920**		
					0.095	0.901						0.951		
<i>Potenzieller Budget Impact: Durchschnitt (Log)<sup>2</sup></i>						-0.040						-0.056**		
						0.027						0.028		
<i>Potenzieller Budget Impact: Maximum (Log)</i>							0.011	1.236						
							0.095	0.892						
<i>Potenzieller Budget Impact: Maximum (Log)<sup>2</sup></i>								-0.036						
								0.026						
<i>Potenzieller Budget Impact: NHS-Population (Log)</i>									-0.095	1.519*				1.680*
									0.093	0.859				0.931
<i>Potenzieller Budget Impact: NHS-Population (Log)<sup>2</sup></i>										-0.047*				-0.051*
										0.025				0.027
<i>Inkrementeller Budget Impact: Durchschnitt (Log)</i>											0.055	0.058		
											0.124	0.125		
<i>Inkrementeller Budget Impact: NHS-Population (Log)</i>													0.056	0.089
													0.125	0.128
<i>Intercept 1</i>	-4.796*	-17.392	-5.099*	1.844	-4.815*	5.953	-4.618*	5.282	-5.844**	7.855	-4.265	11.048	-4.311	9.496
	2.483	1809.867	2.607	7.168	2.705	7.675	2.719	7.652	2.716	7.723	2.646	8.000	2.672	8.285
<i>Intercept 2</i>	-1.038	-14.522	-1.303	5.662	-1.047	9.760	-0.850	9.083	-2.057	11.701	-0.312	15.071*	-0.300	13.59
	2.463	1809.867	2.585	7.190	2.686	7.709	2.702	7.685	2.686	7.763	2.627	8.055	2.653	8.331
Beobachtungen	255	85	251	251	254	254	254	254	254	254	235	235	232	232

ICER, Incremental Cost-Effectiveness Ratio; NHS, National Health Service; QALY, Quality-Adjusted Life Year; RCT, Randomized Controlled Trial.

Regressionsergebnisse sind statistisch signifikant bei: \*\*\* p < 0.01; \*\* p < 0.05; \* p < 0.1.

Anmerkung: Tabelle A6 stellt eine Ergänzung zu Tabelle 22 dar und bildet die Regressionsergebnisse der Dummy-Variablen Veröffentlichungsjahr und Therapeutisches Gebiet ab.

## 11 LEBENSLAUF

### PERSONALIEN

Name und Vorname:	Schäfer, Ramon
Geburtsdatum:	05. Dezember 1988
Geburtsort:	Pforzheim
Familienstand:	ledig
Vater:	Helmut Wilhelm Walter Schäfer
Mutter:	Claudia Schäfer-Mischung geb. Görzner

### SCHULISCHER WERDEGANG

1995 – 1999	Südstadtschule Pforzheim
1999 – 2008	Theodor-Heuss Gymnasium Pforzheim
19. Juni 2008	Allgemeine Hochschulreife / Abitur

### UNIVERSITÄRER WERDEGANG

WS 2009/10	Beginn des Bachelorstudiums in Betriebswirtschaftslehre (BWL) an der Hochschule Pforzheim
2013	Bachelorarbeit ( <i>Titel</i> : Transparency International im Fokus moralischer Akteure für eine korruptionsfreie Wirtschaft in der VR China)
19. August 2013	Bachelor of Science (Abschlussnote: 1,9)
WS 2013/14	Beginn des Masterstudiums in Innovation Management an der Hochschule Ludwigshafen am Rhein
2015	Masterarbeit ( <i>Titel</i> : Health Technology Assessment (HTA): Eine vergleichende Untersuchung der Implementierung bei NICE in England und IQWiG in Deutschland)
12. Oktober 2015	Master of Arts (Abschlussnote: 1,5)

## 12 DANKSAGUNG

Meinem Doktorvater Prof. Dr. Michael Schlander danke ich einerseits für die Überlassung des Themas dieser Arbeit und andererseits für die zeitintensive Betreuung sowie die damit einhergehenden, ausgezeichneten Möglichkeiten, den inhaltlichen Schwerpunkt – trotz beziehungsweise gerade (auch) wegen meiner beruflichen Tätigkeit(en) – ausführlich bearbeiten zu können. Bei Prof. Dr. Joachim E. Fischer und Prof. Dr. Marteen J. Postma bedanke ich mich ferner, als Zweit- und Drittgutachter der Dissertation zu fungieren.

Besonderen Dank schulde ich an dieser Stelle Dr. Diego Hernández Carreno, der mich über das gesamte Promotionsprojekt hinweg wissenschaftlich begleitet und mich insbesondere bei der Durchführung der statistischen Analysen und Untersuchungen umfassend unterstützt hat. Ebenso danke ich Prof. Dr. Peter Kolominsky-Rabas und Prof. Dr. Dr. Till W. Bärnighausen für die fachliche Unterstützung meiner Arbeit und die vielen konstruktiven Hinweise, die zur Vollendung derselben notwendig waren. Darüber hinaus möchte ich mich bei den Studierenden Lorenz Selberg und Clémence Perraud bedanken, die mich bei der Extraktion und Zusammenführung der zugrunde liegenden Daten unterstützt haben.

Abschließend bedanke ich mich von Herzen bei meiner Partnerin Anne-Marie Summer für ihre Geduld und fortwährende Unterstützung während der gesamten Bearbeitungsdauer des Promotionsprojekts, sowie bei 'Fred', der mir in manchem schwierigen Moment geholfen hat, einfach einmal abzuschalten und den Kopf wieder frei zu bekommen.