

# Surrogatendpunkte als Parameter der Nutzenbewertung

Sandra Mangiapane, Marcial Velasco Garrido





**Schriftenreihe  
Health Technology Assessment (HTA)  
in der Bundesrepublik Deutschland**

---

**Surrogatendpunkte als Parameter der Nutzenbewertung**

---

**Sandra Mangiapane  
Marcial Velasco Garrido**

Fachgebiet Management im Gesundheitswesen, Technische Universität Berlin

## **Wir bitten um Beachtung**

Dieser HTA-Bericht ist publiziert in der DAHTA-Datenbank des DIMDI ([www.dimdi.de](http://www.dimdi.de) – HTA) und in der elektronischen Zeitschrift GMS Health Technology Assessment ([www.egms.de](http://www.egms.de)).

Die HTA-Berichte des DIMDI durchlaufen ein unabhängiges, grundsätzlich anonymisiertes Gutachterverfahren. Potentielle Interessenkonflikte bezüglich der HTA-Berichte werden dem DIMDI von den Autoren und den Gutachtern offengelegt. Die durchgeführte Literaturrecherche erhebt keinen Anspruch auf Vollständigkeit. Die Literatúrauswahl erfolgt nach den Kriterien der evidenzbasierten Medizin. Die Verantwortung für den Inhalt des Berichts obliegt den jeweiligen Autoren.

Die Erstellung des vorliegenden HTA-Berichts des Deutschen Instituts für Medizinische Dokumentation und Information (DIMDI) erfolgte gemäß gesetzlichem Auftrag nach Artikel 19 des GKV-Gesundheitsreformgesetzes 2000. Das Thema stammt aus dem öffentlichen Vorschlagsverfahren beim DIMDI, durch das Kuratorium HTA priorisiert und vom DIMDI beauftragt. Der Bericht wurde mit Mitteln des Bundes finanziert.

---

## **Herausgeben vom Deutschen Institut für Medizinische Dokumentation und Information (DIMDI), Köln**

Das DIMDI ist ein Institut im Geschäftsbereich des Bundesministeriums für Gesundheit (BMG)

### **Kontakt**

DAHTA  
Deutsche Agentur für Health Technology Assessment des  
Deutschen Instituts für Medizinische Dokumentation und Information  
Waisenhausgasse 36-38a  
50676 Köln  
Tel: +49 221 4724-525  
Fax: +49 2214724-340  
E-Mail: [dahta@dimdi.de](mailto:dahta@dimdi.de)  
[www.dimdi.de](http://www.dimdi.de)

Schriftenreihe Health Technology Assessment, Bd. 91

ISSN: 1864-9645

1. Auflage 2009

DOI: 10.3205/hta000074L

URN: urn:nbn:de:0183-hta000074L8

# Inhaltsverzeichnis

<b>1</b>	<b>Verzeichnisse</b>	<b>V</b>
1.1	Tabellenverzeichnis	V
1.2	Abbildungsverzeichnis	VI
1.3	Abkürzungsverzeichnis	VI
1.4	Glossar	X
<b>2</b>	<b>Zusammenfassung</b>	<b>1</b>
<b>3</b>	<b>Abstract</b>	<b>2</b>
<b>4</b>	<b>Kurzfassung</b>	<b>3</b>
4.1	Gesundheitspolitischer Hintergrund	3
4.2	Wissenschaftlicher Hintergrund	3
4.3	Forschungsfragen	4
4.4	Methodik	4
4.5	Ergebnisse	4
4.5.1	Literaturrecherche	4
4.5.2	Analyse von Methodenpapieren aus HTA-Agenturen	5
4.5.3	Analyse von HTA-Berichten	5
4.6	Diskussion	6
4.6.1	Diskussion der Methodik	6
4.6.2	Diskussion der Fragestellungen	6
4.7	Schlussfolgerung	7
<b>5</b>	<b>Summary</b>	<b>8</b>
5.1	Health policy background	8
5.2	Scientific background	8
5.3	Research questions	9
5.4	Methods	9
5.5	Results	9
5.5.1	Literature review	9
5.5.2	Analysis of methodological guidance from HTA agencies	10
5.5.3	Survey of HTA reports	10
5.6	Discussion	10
5.6.1	Methods of the report	10
5.6.2	Results of the report	11
5.7	Conclusions	11
<b>6</b>	<b>Hauptdokument</b>	<b>12</b>
6.1	Gesundheitspolitischer Hintergrund	12
6.1.1	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) und Gemeinsamer Bundesausschuss (G-BA)	12
6.1.2	Richtlinien des Gemeinsamen Bundesausschusses (G-BA)	12
6.1.3	Nutzen und Nutzenbewertung im Kontext des Sozialgesetzbuchs (SGB V)	13
6.1.4	Endpunkte in der Nutzenbewertung im Kontext des Sozialgesetzbuchs (SGB V)	13
6.2	Wissenschaftlicher Hintergrund	14
6.2.1	Surrogatendpunkte – Definitionen	14
6.2.2	Surrogatendpunkte – Problematik	16
6.3	Forschungsfragen	18
6.4	Methodik	19
6.4.1	Recherche in Literaturdatenbanken - Kriterien zur Beurteilung und zur Validierung von Surrogatendpunkten	19
6.4.1.1	Informationsquellen und Recherchen	19
6.4.1.2	Selektion	19
6.4.1.3	Einschlusskriterien	20
6.4.1.4	Ausschlusskriterien	20
6.4.1.5	Synthese	20

6.4.2	Review der Methodenpapiere internationaler Agenturen aus dem Bereich HTA oder Arzneimittelnutzenbewertung .....	20
6.4.2.1	Identifizierung relevanter Institutionen .....	20
6.4.2.2	Dokumentenanalyse .....	21
6.4.2.3	Synthese.....	21
6.4.3	Analyse der Anwendung von Surrogatendpunkten in HTA-Berichten.....	21
6.4.3.1	Quellen .....	21
6.4.3.2	Einschlusskriterien für HTA-Berichte .....	22
6.4.3.3	Ausschlusskriterien für HTA-Berichte .....	22
6.4.3.4	Stichprobenziehung .....	22
6.4.3.5	Definition verwendeter Endpunkte.....	22
6.4.3.6	Synthese.....	23
6.5	Ergebnisse .....	23
6.5.1	Recherche in Literaturdatenbanken - Kriterien zur Beurteilung und Validierung von Surrogatendpunkten .....	23
6.5.1.1	Quantitative Ergebnisse .....	23
6.5.1.2	Qualitative Ergebnisse.....	25
6.5.1.2.1	Konzeptionelle Aspekte .....	25
6.5.1.2.2	Statistische Methoden zur Validierung der Surrogatendpunkten.....	32
6.5.2	Aussagen zu Surrogaten in Methodenpapieren von HTA-Agenturen und Arzneimittelbewertungsinstitutionen .....	38
6.5.2.1	Quantitative Ergebnisse .....	38
6.5.2.2	Qualitative Ergebnisse – Aussagen zu Surrogaten in Methodenpapieren .....	40
6.5.3	Anwendung von Surrogatendpunkten in HTA-Berichten.....	42
6.5.3.1	Internationale HTA-Berichte .....	42
6.5.3.2	Deutsche Berichte .....	48
6.6	Diskussion.....	53
6.6.1	Diskussion der Methodik.....	53
6.6.2	Diskussion und Beantwortung der Forschungsfragen.....	55
6.6.2.1	Welche Kriterien muss ein Surrogatparameter erfüllen, um als valider Endpunkt angesehen werden zu können?.....	55
6.6.2.2	Welche Methoden werden zur Validierung von Surrogatendpunkten in der Literatur diskutiert? .....	56
6.6.2.3	Welche methodischen Vorgaben machen internationale Agenturen aus dem Bereich HTA oder Arzneimittelnutzenbewertung hinsichtlich des Einsatzes von Surrogatendpunkten?.....	56
6.6.2.4	Welchen Stellenwert haben Surrogatendpunkte in HTA-Berichten internationaler HTA-Agenturen und in Berichten, die in Deutschland produziert werden?.....	56
6.7	Schlussfolgerungen .....	57
<b>7</b>	<b>Literatur .....</b>	<b>58</b>
<b>8</b>	<b>Anhang .....</b>	<b>63</b>
8.1	Dokumentation der Literaturrecherche .....	63
8.2	Aus der Methodenübersicht ausgeschlossene Literatur.....	63
8.3	Dokumentation der Methodenpapiere aus HTA-Institutionen.....	68
8.4	HTA-Agenturen, deren Berichte in der HTA-Datenbank enthalten sind.....	81
8.5	HTA-Berichte aus der HTA-Datenbank .....	83
8.6	Deutsche HTA-Berichte .....	89

# 1 Verzeichnisse

## 1.1 Tabellenverzeichnis

Tabelle 1: Beispiele von Surrogatzielgrößen für verschiedene Indikationen .....	15
Tabelle 2: Fehler bei der Bewertung von Arzneimitteln auf der Grundlage von Surrogatendpunkten	17
Tabelle 3: Eingeschlossene Artikel aus der elektronischen Literaturrecherche.....	24
Tabelle 4: Surrogatendpunkte – Definitionen .....	26
Tabelle 5: Konstellationen für das Zusammenspiel zwischen Therapie, Surrogat und Endpunkt .....	28
Tabelle 6: INAHTA-Mitglieder mit veröffentlichtem Methodenpapier .....	39
Tabelle 7: Berücksichtigte Institutionen zur Arzneimittelnutzenbewertung und zu Methodenpapieren.....	40
Tabelle 8: Von HTA-Agenturen und Arzneimittelbewertungsinstitutionen bevorzugte Ergebnisparameter .....	41
Tabelle 9: Länder- und Jahresverteilung in der Grundgesamtheit, in der 5 %-Zufalls- und in der endgültigen Stichprobe.....	43
Tabelle 10: Internationale HTA-Berichte - Bewertete Technologien (n = 140) .....	44
Tabelle 11: Prospektive Festlegung der für die Bewertung relevanten Ergebnisparameter in internationalen HTA-Berichten (n = 140) .....	44
Tabelle 12: Art der verwendeten Endpunkte in internationalen HTA-Berichten (n = 140).....	45
Tabelle 13: Verwendete Endpunkten in internationalen HTA-Berichten – Vergleich diagnostische versus therapeutische Technologien .....	45
Tabelle 14: Verwendete Endpunkte in internationalen HTA-Berichten – prospektiv versus nicht- prospektiv festgelegt .....	45
Tabelle 15: Überblick über Surrogatendpunkte in internationalen HTA-Berichten – nicht-diagnostische Technologien.....	46
Tabelle 16: Berücksichtigte deutsche HTA-Berichte .....	49
Tabelle 17: Prospektive Festlegung der für die Bewertung relevanten Ergebnisparameter in deutschen HTA-Berichten (n = 131) .....	49
Tabelle 18: Art der verwendeten Endpunkte in deutschen HTA-Berichten (n = 131).....	50
Tabelle 19: Verwendete Endpunkte in deutschen HTA-Berichten – Vergleich diagnostische versus therapeutische Technologien.....	50
Tabelle 20: Verwendete Endpunkte in deutschen HTA-Berichten – prospektiv versus nicht-prospektiv festgelegt.....	51
Tabelle 21: Überblick über Surrogatendpunkte in deutschen HTA-Berichten – nicht diagnostische Technologien.....	52
Tabelle 22: Aussagen zu Surrogaten in Methodenpapieren der Mitgliedinsitutionen von INAHTA.....	68
Tabelle 23: Aussagen zu Surrogaten in Methodenpapieren der Agenturen zur Arzneimittelnutzenbewertung im Rahmen von Kostenerstattungsentscheidungen .....	77

## 1.2 Abbildungsverzeichnis

Abbildung 1: Quelle: Boissel et al. <sup>12</sup> .....	26
Abbildung 2: Quelle: Fleming und de Mets <sup>30</sup> .....	29
Abbildung 3: Verschiedene Konstellationen der Beziehung zwischen Surrogatparametern und Endpunkten .....	30
Abbildung 4: Study-Flow internationaler HTA-Berichte .....	43

## 1.3 Abkürzungsverzeichnis

ACMD	Average clinically meaningful difference
ACR	American College of Rheumatology
AETMIS	Agence d'Evaluation des Technologies, Kanada
AETS	Agencia de Evaluación de Tecnologías Sanitarias, Spanien
AETSA	Agencia de Evaluación de Tecnologías Sanitarias de Andalucía, Spanien
AHFMR	Alberta Heritage Foundation for Medical Research, Kanada
AHRQ	Agency for Healthcare Research and Quality, USA
AHTA	Adelaide Health Technology Assessment, Australien
AHTA-Pol	Agency for Health Technology Assessment in Poland, Polen
AIDS	Acquired immuno deficiency syndrome
APEO	Average prediction error of other true endpoits
APEP	Average prediction error of the estimated prediction effect
AR	Argentinien
ASERNIP-S	Australian Safety and Efficacy Register of New Interventional Procedures - Surgical, Australien
AT	Österreich
AU	Australien
AVALIA-T	Galician Agency for Health Technology Assessment, Spanien
BCBS	BlueCross BlueShield Technology Evaluation Center, USA
BCOHTA	British Columbia Office of Health Technology Assessment, Kanada
BE	Belgien
BMG	Bundesministerium für Gesundheit
BR	Brasilien
BUA	Bundesausschuss der Ärzte und Krankenkassen
CA	Kanada
CADTH	Canadian Agency for Drugs and Technologies in Health, Kanada
CAHTA	Catalan Agency for Health Technology Assessment and Research, Spanien
CAST	Cardiac arrythmia suppression trial (Studie)
CCE	The Centre for Clinical Effectiveness, Australien
CED	Coverage with evidence development
CEA	Carcino-embryonales Antigen
CEDIT	Comité d'Evaluation et de Diffusion des Innovations Technologiques Assistance Publique Hôpitaux de Paris, Frankreich
CENETEC	Centro Nacional de Excelencia Tecnológica en Salud, Brasilien

**Fortsetzung: Abkürzungsverzeichnis**

CH	Schweiz
CHUM	Centre Hospitalier de l'Université de Montréal, Kanada
CMT	Centre for Medical Technology Assessment, Schweden
CRD	Centre for Reviews and Dissemination, Vereinigtes Königreich
CT	Computertomogramm
CTFPHC	Canadian Task Force on Preventive Health Care, Kanada
CVZ	College voor Zorgverzekeringen, Niederlande
DACEHTA	Danish Centre for Evaluation and Health Technology Assessment, Dänemark
DAHTA	Deutsche Agentur für Health Technology Assessment
DE	Deutschland
DECIT-CGATS	Department of Science and Technology, Brazilian Health Technology Assessment General Coordination, Brasilien
DIMDI	Deutsches Institut für Medizinische Dokumentation und Information, Deutschland
DMP	Disease Management Programme
DK	Dänemark
DSI	Danish Institute for Health Services Research and Development, Dänemark
E	Klinisch relevanter Endpunkt
ECRI	Emergency Care Research Institute
EDSS	Expanded disability status scale
EMA	European Medicines Agency
EPO	Erythropoietin
ES	Spanien
ETESA	Unidad De Tecnologias De Salud, Chile
EU	Europäische Union
FDA	Food and Drug Administration, USA
FEV1	Forced expiratory volumen in one second
FI	Finnland
FinOHTA	Finnish Office for Health Care Technology Assessment, Finnland
FR	Frankreich
FVC1	Einsekunden-Ausatmungsvolumen, Lungenfunktionstest
G-BA	Gemeinsamer Bundesausschuss
GKV	Gesetzliche Krankenversicherung
GKV-WSG	Gesetz zur Stärkung des Wettbewerbs in der Gesetzlichen Krankenversicherung
GMG	Gesetz zur Modernisierung der Gesetzlichen Krankenversicherung
GR	Health Council of the Netherlands, Niederlande
HAS	Haute Autorité de Santé/French National Authority for Health, Frankreich
HbA1c	Glykosiliertes Hämoglobin
HIV	Human Immuno Deficiency Virus
HNO	Hals-Nase-Ohren
HSURC	Health Services Utilization and Research Committee, Kanada

**Fortsetzung: Abkürzungsverzeichnis**

HTA	Health Technology Assessment
HTAC	Health Technology Advisory Committee, USA
HTAi	Health Technology Assessment international (Internationale Gesellschaft für HTA)
HU	Ungarn
HunHTA	Unit of Health Economics and Health Technology Assessment, Ungarn
IAHS	Institute of Applied Health Sciences, UK
ICES	Institute for Clinical Evaluative Sciences, Kanada
ICH	International Conference on Harmonisation
ICSI	The Institute for Clinical Systems Improvement, USA
ICTAHC	Israel Centre for Technology Assessment in Health Care, Israel
IECS	Institute for Clinical Effectiveness and Health Policy, Argentinien
IgE	Immunoglobulin E
IHE	Institute of Health Economics, Kanada
IL	Israel
INAHTA	International Network of Agencies for Health Technology Assessment
IQWiG	Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen, engl.: Institute for Quality and Efficiency in Health Care
J.	Jahre
KBV	Kassenärztliche Bundesvereinigung
KCE	Belgian Health Care Knowledge Centre, Belgien
KHK	Koronare Herzerkrankung
LBI@ITA	Ludwig Boltzmann Institute of Health Technology Assessment, Österreich
LDL	Low density lipoprotein
LRF	Likelihood reduction factor
LV	Lettland
LVEF	Linksventrikuläre Ejektionsfraktion
MAS	Medical Advisory Secretariat, Ontario Ministry of Health and Long-Term Care, Kanada
MHH	Medizinische Hochschule Hannover
MHTAU	Malaysian Health Technology Assessment Unit, Malaysia
MRT	Magnetresonanztomografie
MSAC	Medical Services Advisory Committee, Australien
MTU-SFOPH	Medical Technology Unit, Swiss Federal Office of Public Health, Schweiz
MUHC	Technology Assessment Unit of the McGill University Health Centre, Kanada
MX	Mexiko
NCCHTA	National Coordinating Centre for Health Technology Assessment, UK
NHSC	National Horizon Scanning Centre, UK
NHS-QIS	National Health Service - Quality Improvement Scotland, UK
NICE	National Institute for Health and Clinical Excellence, Großbritannien
NL	Niederlande
NO	Norwegen

**Fortsetzung: Abkürzungsverzeichnis**

NOKC	Norwegian Knowledge Centre for the Health Services, Norwegen
NZ	Neuseeland
NZHTA	New Zealand Health Technology Assessment, Neuseeland
OECD	Organisation für wirtschaftliche Zusammenarbeit und Entwicklung
OSTEBA	Basque Office for Health Technology Assessment, Spanien
OTA	Office of Technology Assessment, USA
P	Wahrscheinlichkeit
PASI	Psoriasis area and severity index
PBAC	Pharmaceutical Benefits Advisory Committee
PEFR	Peak expiratory flow rate
PHARMAC	Pharmaceutical Management Agency
PL	Polen
PRO	Patient Reported Outcomes
PROFILE	(Studie)
PSA	Prostata-spezifisches Antigen
PTE	Proportion of treatment effect explained
RCT	Randomisierte kontrollierte Studie, engl.: randomized controlled trials
RE	Relativer Effekt
RNA	Ribonucleinsäure
S	Surrogat
SBK	Spitzenverband der Krankenkassen
SBU	Swedish Council on Technology Assessment in Health Care, Schweden
SCORAD	Scoring atopic dermatitis
SE	Schweden
SGB V	Sozialgesetzbuch V
STE	Surrogate threshold effects
SVR	Sustained viral response
SHPIC	Scottish Health Purchasing Information Centre, UK
T	Therapie
TA-Swiss	Swiss Science and Technology Council/Technology Assessment, Schweiz
TNO	TNO Prevention and Health, Niederlande
U	Ursache
UETS	Unidad de Evaluacion de Tecnologias Sanitarias, Spanien
UHC	University HealthSystem Consortium, USA
UK	Großbritannien
USA	Vereinigte Staaten von Amerika
VATAP	Veteran Affairs Technology Assessment Program, USA
VRF	Variance reduction factor
WHO	World Health Organisation, dt.: Weltgesundheitsorganisation
WIHRD	Wessex Institute for Health Research and Development
WMHTAC	West Midlands Health Technology Assessment Collaboration, UK

**Fortsetzung: Abkürzungsverzeichnis**

WorkSafe BC	Workers' Compensation Board of BC, Kanada
WSDLI	Washington State Department of Labor and Industries, USA
UVB	Ultraviolett-B
ZonMW	The Netherlands Organisation for Health Research and Development, Niederlande
$\gamma_T$	Adjusted association

**1.4 Glossar**

ACR-Response	Kriterien des American College of Rheumatology (ACR) für die Überprüfung des Ansprechens auf antirheumatische Therapien. Sie kombinieren Angaben der Patienten zu Symptomatik, Behinderungsgrad und Laborwerte.
Deutsches HTA-Projekt	Projekt gefördert vom Bundesministerium für Gesundheit (BMG), das zwischen 1995 und 2001 von der Medizinischen Hochschule Hannover (MHH) koordiniert wird. Im Rahmen des Projekts werden HTA-Berichte für das Deutsche Institut für Medizinische Dokumentation und Information (DIMDI) erstellt bis zur Gründung der Deutschen Agentur für Health Technology Assessment (DAHTA).
Diagnostische Genauigkeit	Bei diagnostischen Tests, Anteil der Patienten, die durch den Test richtig klassifiziert werden.
Endpunkt	Siehe Outcome.
Inzidenz	Anzahl von Neuerkrankungen in einer definierten Bevölkerung während eines bestimmten Zeitraums.
Kosten-Effektivität	Verhältnis zwischen den Kosten einer Intervention und dem Nutzen derselben.
Kostenerstattungsentscheidungen	Entscheidungen über die Kostenübernahme von Technologien in Rahmen des Gesundheitssystems.
Metaanalyse	Quantitative Zusammenfassung der Ergebnisse mehrerer Studien, die ein einzelner Effektschätzer ergibt.
Outcome	Ergebnis der Progression der Krankheit bzw. der Behandlung.
Psoriasis Area and Severity Index (PASI)	Score, der die Ergebnisse einer körperlichen Untersuchung zur Beurteilung der Ausdehnung und Intensität einer Psoriasis in einer Zahl zusammenfasst. Beurteilung durch Kliniker aber auch durch Patienten möglich (Selbst-PASI).
Prädiktiver Wert	Bei diagnostischen Tests, Wahrscheinlichkeit das nach einem positiven bzw. negativen Test, der Getestete tatsächlich erkrankt oder gesund ist.
Primärstudie	Studie, bei der Originaldaten analysiert werden. Die Daten können für die Studie gezielt erhoben sein oder aus einer routinemäßigen Datensammlung (Register, administrative Daten) stammen.

**Fortsetzung: Glossar**

Peak flow (PEFR)	Lungenfunktionstest, kann auch von Patienten selbst durchgeführt werden, um das weitere Vorgehen bei einem Asthmaanfall zu entscheiden.
Randomisierte kontrollierte Studie (RCT)	Studie, bei der Patienten nach dem Zufallprinzip (Randomisierung) einer von zwei oder mehrere Gruppen zugeteilt werden. Die Gruppen unterscheiden sich in der Intervention (z. B. eine Therapie versus Placebo, oder zwei verschiedene Therapien).
Scoring atopic dermatitis (SCORAD)	Score, der die Ergebnisse einer körperlichen Untersuchung zur Beurteilung der Ausbreitung und Intensität einer Atopischen Dermatitis (Neurodermitis) in einer Zahl zusammenfasst.
Sensitivität	Bei diagnostischen Tests, Anteil von Erkrankten, die durch den Test als solche erkannt werden (Test = Krank/Alle Erkrankten).
Spezifizität	Bei diagnostischen Tests beschreibt Spezifität den Anteil von Gesunden, die durch den Test als solche erkannt werden (Test = Gesund/Alle Gesunden).
Übersichtstudie	Studie, die andere Studien (Primärstudien) bewertet und zusammenfasst.

## 2 Zusammenfassung

Die Frage nach den Ergebnisparametern, die für eine valide Bewertung des Nutzens medizinischer Technologien verwendet werden können, beschäftigt alle an Entscheidungsfindungen und an der Regulierung im Gesundheitssystem beteiligten Akteure.

Während die klinisch relevanten Endpunkte jene sind, die für den Patienten belangvoll sind (z. B. Morbidität, Mortalität), stellen Surrogatendpunkte im besten Fall Vorstufen der eigentlichen klinisch relevanten Endpunkte dar, die für den Patienten in der Regel nicht unmittelbar spürbar sind. Surrogatparameter werden nicht nur in Studien über die Wirksamkeit von Arzneimitteln eingesetzt, sondern auch in Studien zu anderen Technologien. Der Einsatz von Surrogatendpunkten in der Bewertung des Nutzens von Gesundheitstechnologien ist jedoch problematisch.

In diesem Bericht wird folgenden Fragen nachgegangen:

- Welche Kriterien muss ein Surrogatparameter erfüllen, um als valider Endpunkt angesehen werden zu können?
- Welche Methoden werden zur Validierung von Surrogatendpunkten in der Literatur diskutiert?
- Welche methodischen Vorgaben machen internationale Agenturen aus dem Bereich Health Technology Assessment (HTA) oder Arzneimittelnutzenbewertung hinsichtlich des Einsatzes von Surrogatendpunkten?
- Welchen Stellenwert haben Surrogatendpunkte in HTA-Berichten internationaler HTA-Agenturen und in Berichten, die in Deutschland produziert werden?

Entsprechend dieser Fragestellungen werden drei verschiedene methodische Herangehensweisen gewählt: Reviews der einschlägigen methodischen Literatur zu Surrogatendpunkten und ihre Validierung, Analyse der aktuellen Methodenpapiere von HTA-Institutionen, Analyse von abgeschlossenen sowie öffentlich zugänglichen internationalen und deutschen HTA-Berichten.

Zusammenfassend zeigen die Empfehlungen der hier berücksichtigten Institutionen dahingehend eine kritische Einstellung zur Verwendung von Surrogatendpunkten in HTA. Es lässt sich zudem feststellen, dass Surrogatendpunkte einen geringen Stellenwert in HTA-Berichten haben. In keinem der untersuchten HTA-Berichte wird die Bewertung von nicht diagnostischen Technologien ausschließlich auf die Ergebnisse von Surrogatendpunkten gestützt. Die Ergebnisse der Analyse deutscher HTA-Berichte sind annähernd identisch zu denen internationaler.

Die Validierung eines Surrogatendpunkts erfordert ausführliche Forschung, einschließlich der Durchführung von randomisierten kontrollierten Studien (RCT) mit klinisch relevanten Endpunkten. Die Validität eines Surrogatendpunkts ist eher technologie- als krankheitsspezifisch, so dass die Ergebnisse der Validierung eines Surrogatendpunkts nicht auf andere Technologien übertragen werden können (auch nicht bei einem angeblich ähnlichen Wirkmechanismus). Um ein Höchstmaß an Sicherheit zu erreichen, muss die Validität eines Surrogatendpunkts bei jeder Technologie bzw. jedem Wirkstoff einzeln geprüft werden.

Nach wie vor ist der Einsatz von Surrogatendpunkten bei der Nutzenbewertung sehr kritisch zu betrachten.

### 3 Abstract

The different actors involved in health system decision-making and regulation have to deal with the question which are valid parameters to assess the health value of health technologies.

So called surrogate endpoints represent in the best case preliminary steps in the casual chain leading to the relevant outcome (e. g. mortality, morbidity) and are not usually directly perceptible by patients. Surrogate endpoints are not only used in trials of pharmaceuticals but also in studies of other technologies. Their use in the assessment of the benefit of a health technology is however problematic.

In this report we intend to answer the following research questions: Which criteria need to be fulfilled for a surrogate parameter to be considered a valid endpoint? Which methods have been described in the literature for the assessment of the validity of surrogate endpoints? Which methodological recommendations concerning the use of surrogate endpoints have been made by international HTA agencies? Which place has been given to surrogate endpoints in international and German HTA reports?

For this purpose, we choose three different approaches. Firstly, we conduct a review of the methodological literature dealing with the issue of surrogate endpoints and their validation. Secondly, we analyse current methodological guidelines of HTA agencies members of the International network of agencies for Health Technology Assessment (INAHTA) as well as of agencies concerned with assessments for reimbursement purposes. Finally, we analyse the outcome parameter used in a sample of HTA reports available for the public.

The analysis of methodological guidelines shows a very cautious position of HTA institutions regarding the use of surrogate endpoints in technology assessment. Surrogate endpoints have not been prominently used in HTA reports. None of the analysed reports based its conclusions solely on the results of surrogate endpoints. The analysis of German HTA reports shows a similar pattern.

The validation of a surrogate endpoint requires extensive research, including randomized controlled trials (RCT) assessing clinical relevant endpoints. The validity of a surrogate parameter is rather technology-specific than disease-specific. Thus – even in the case of apparently similar technologies – it is necessary to validate the surrogate for every single technology (i. e. for every single active agent).

The use of surrogate endpoints in the assessment of the benefit of health technologies is still to be seen very critically.

## 4 Kurzfassung

### 4.1 Gesundheitspolitischer Hintergrund

Die Diskussionen, welche Ergebnisparameter für eine valide Bewertung des Nutzens medizinischer Technologien verwendet werden können, sind spätestens seit der Gründung des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) in Deutschland wieder aufgelebt. Diese Frage beschäftigt alle Akteure, die an die Entscheidungsfindung und Regulierung im Gesundheitssystem beteiligt sind.

### 4.2 Wissenschaftlicher Hintergrund

Während klinisch relevante Endpunkte Einschnitte bezeichnen, die für den Patienten bedeutend sind (z. B. Morbidität, Mortalität), stellen Surrogatendpunkte im besten Fall Vorstufen der eigentlichen klinisch relevanten Endpunkte dar, die für den Patienten in der Regel nicht unmittelbar spürbar sind. Surrogatendpunkte weisen folgende Charakteristika auf:

- Es handelt sich um Zielgrößen, die anstelle von (*in lieu*) den eigentlichen Endpunkten gemessen werden.
- Es handelt sich häufig um physiologische oder biochemische Marker bzw. subklinische Endpunkte, die für den Patienten in der Regel nicht unmittelbar von Bedeutung bzw. in irgendeiner Weise bemerkbar sind, jedoch stellvertretend für wichtige klinische Endpunkte stehen können (z. B. erhöhter Blutdruck verstärkt das Risiko für Schlaganfall, erhöhte LDL-Cholesterinwerte (LDL = Low density lipoprotein) sind ein Risikofaktor für Myokardinfarkt, der Anzahl der CD4-Zellen für AIDS-Mortalität (AIDS = Acquired Immuno Deficiency Syndrome)).
- Veränderungen dieser Parameter sind im Vergleich zum Auftreten der relevanten Endpunkte (Morbidität, Tod) häufiger und nach kürzerer Beobachtungszeit feststellbar.
- Sie stellen Zwischenschritte auf dem Weg von der Erkrankung hin zum klinischen Endpunkt dar und werden deshalb mitunter auch als „intermediate“ oder „intermediary outcomes“ bezeichnet.
- Es gibt einen (statistisch nachweisbaren) Zusammenhang zwischen dem Surrogatendpunkt und dem eigentlichen klinischen Endpunkt im Sinne einer guten Prognosekraft.
- Der Zusammenhang zwischen Surrogat und klinisch relevantem Endpunkt muss biologisch und pathophysiologisch (d. h. bzgl. der Funktion des Organismus unter Krankheit) plausibel sein.

Surrogatparameter werden nicht nur in Studien über die Wirksamkeit von Arzneimitteln eingesetzt. Auch Studien über die Effekte von medizinischen und chirurgischen Prozeduren berichten häufig über Ergebnisparameter, die nicht das ultimative, patientennahe Ziel der Intervention wiedergeben. Bei nicht-klinischen Interventionen (z. B. Patientenschulungen, bevölkerungsbezogenen Interventionen etc.) werden auch häufig Zielgrößen gemessen, die einen intermediären Charakter haben.

Die Verwendung von Surrogatendpunkten ist jedoch nicht unproblematisch. In der Vergangenheit gab es viele Situationen, in denen das Vertrauen auf Surrogatendpunkte trotz starker Korrelation mit dem klinischen Endpunkt irreführend war bzw. fatale Folgen hatte. Diese Problematik ist mittlerweile mehr als 30 Jahre bekannt. Prominentes und mittlerweile klassisches Beispiel stellen die Klasse I-Antiarrhythmika dar. Einige Produkte, die auf der Basis von Surrogatendpunkten zugelassen wurden, mußten später vom Markt genommen werden, da sich in Studien mit Mortalitäts- bzw. Morbiditätseindpunkten, die Nutzen-Risiko-Abwägung umkehrte. Aber ebenso kann das Festhalten an Surrogatendpunkten auch dazu führen, dass therapeutische Chancen nicht genutzt werden. So galten beispielsweise Betablocker über lange Zeit bei Patienten mit Herzinsuffizienz als kontraindiziert, da sie zu einer Verringerung der Herzfrequenz führen, was unter pathophysiologischen Überlegungen zu negativen Effekten bei diesen Patienten führen sollte.

## 4.3 Forschungsfragen

1. Welche Kriterien muss ein Surrogatparameter erfüllen, um als valider Endpunkt angesehen werden zu können?
2. Welche Methoden werden zur Validierung von Surrogatendpunkten in der Literatur diskutiert?
3. Welche methodischen Vorgaben machen internationale Agenturen aus den Bereichen Health Technology Assessment (HTA) oder Arzneimittelnutzenbewertung hinsichtlich des Einsatzes von Surrogatendpunkten?
4. Welchen Stellenwert haben Surrogatendpunkte in HTA-Berichten internationaler HTA-Agenturen und in Berichten, die in Deutschland produziert werden?

## 4.4 Methodik

Entsprechend der Fragestellungen dieses Berichts, werden drei verschiedene methodische Herangehensweisen gewählt.

Um die erste und die zweite Frage bzgl. der methodischen Diskussionen um die Anwendung von Surrogatendpunkten zu beantworten, wird ein Review der einschlägigen methodischen Literatur durchgeführt. Folgende Suchbegriffe werden verwendet:

SURROGATE END POINT; SURROGATE END POINTS; SURROGATE ENDPOINT; SURROGATE ENDPOINTS; ENDPOINT, SURROGATE; ENDPOINTS, SURROGATE; END POINT, SURROGATE; END POINTS,SURROGATE; BIOLOGICAL MARKER; BIOLOGICAL MARKERS; VALIDATION; STATISTICS; BIOMETRY; DECISION SUPPORT TECHNIQUES; ENDPOINT DETERMINATION; CAUSALITY

Die methodische Literatur wird in Form eines narrativen Reviews zusammengefasst. Sie besteht aus einem Teil, in dem die Anforderungen dargestellt werden, die ein Surrogat erfüllen muss, um als valide zu gelten. Im zweiten Teil wird ein Überblick der in der Literatur diskutierten statistischen Methoden für die Surrogatendpunkt-Validierung gegeben.

Für die Beantwortung der dritten Fragestellung bzgl. des Umgangs von internationalen Agenturen mit dem Thema Surrogatendpunkte werden die Methodenpapiere dieser Institutionen untersucht. Dabei werden alle Mitglieder der INAHTA berücksichtigt, sowie eine Gruppe von Agenturen, die in der Arzneimittelnutzenbewertung im Rahmen von Kostenerstattung- und Preisentscheidungen involviert ist.

Für die Beantwortung der vierten Frage bzgl. der Anwendung von Surrogatendpunkten in HTA-Berichten werden die in abgeschlossenen und veröffentlichten internationalen sowie deutschen HTA-Berichten berücksichtigten Outcomeparameter analysiert. Dabei wird eine Stichprobe aus der HTA-Berichtsdatenbank der INAHTA untersucht. Für Deutschland werden alle Berichte mit nicht-methodischen Inhalten berücksichtigt, die in der HTA-Datenbank der Deutschen Agentur für HTA (DAHTA) gesammelt sind.

## 4.5 Ergebnisse

### 4.5.1 Literaturrecherche

Insgesamt werden mit der Recherchestrategie N = 1.109 Treffer erzielt. Von diesen werden im ersten Selektionsschritt n = 2 Dubletten entfernt, und n = 1.007 Literaturstellen ausgeschlossen, da sie für das Ziel des Berichts nach Durchsicht von Titel und Abstract als nicht-relevant eingestuft werden. Insgesamt werden n = 100 Literaturstellen ausgewählt und für eine weitere Überprüfung im Volltext bestellt. Eine weitere Studie wird aus den Literaturverzeichnissen der Publikationen ausgewählt und ebenfalls im Volltext bestellt. Nach der Selektion werden aus der elektronischen Literaturrecherche 25 Studien für das Review zu Kriterien zur Beurteilung von Surrogatendpunkten und Methoden zur Validierung von Surrogatendpunkten eingeschlossen.

Die in der Literatur diskutierten Kriterien, die ein Surrogatendpunkt erfüllen muss, damit dieser als valider Endpunkt eingestuft werden kann, lassen sich wie folgt zusammenfassen:

- **Biologische Plausibilität:** Es liegt Evidenz aus tierexperimentellen oder epidemiologischen Studien vor, die auf einen kausalen Zusammenhang zwischen dem Surrogatparameter und dem klinischen Endpunkt hinweisen. Der Surrogatparameter liegt demnach auf dem pathophysiologischen Kausalpfad zum klinisch relevanten Endpunkt.
- **Ausmaß der Beziehung zwischen Surrogatendpunkt und klinischem Endpunkt:** Epidemiologische Studien zeigen wiederholt, dass durch eine Veränderung des Surrogatendpunkts eine qualitativ und quantitativ gleichgerichtete Veränderung des klinisch relevanten Endpunkts erreicht werden kann. Das Surrogat sollte den Effekt der Intervention vollständig abbilden.
- **Wirkungsnachweis im Rahmen von RCT:** Es liegen Ergebnisse aus RCT vor, die zeigen, dass durch eine therapiebedingte Veränderung des Surrogatendpunkts eine gleichgerichtete Veränderung des klinisch relevanten Endpunkts erreicht werden kann. Der Surrogatendpunkt muss den Effekt der Intervention vollständig abbilden. Dabei ist zu beachten, dass es nicht nur bei pharmakologisch unterschiedlichen Wirkstoffen, sondern auch bei Stoffen der gleichen Wirkstoffgruppe zu abweichenden Wirkmechanismen kommen kann und daher die Übertragbarkeit der Studienergebnisse von einer Substanz auf eine andere zusätzlich geprüft werden muss.

In der Literatur werden verschiedene statistische Methoden zur Validierung von Surrogatendpunkten beschrieben, die im Bericht ausführlich dargestellt werden. Insgesamt kann festgehalten werden, dass es derzeit keinen Goldstandard für die Validierung von Surrogatendpunkten gibt. Aufgrund der Fehleranfälligkeit, die mit der Verallgemeinerung von Ergebnissen aus Einzelstudien verbunden ist, sind allerdings Ansätze auf Basis mehrerer Studien, wie dies z. B. bei Metaanalysen der Fall ist, zu bevorzugen.

#### 4.5.2 Analyse von Methodenpapieren aus HTA-Agenturen

Es werden 23 Methodenpapiere von 14 INAHTA-Mitgliedern (elf Länder) identifiziert, deren Aussagen bzgl. der in der Bewertung zu verwendenden Ergebnisparameter im Allgemeinen und zur Anwendung von Surrogatendpunkten im Besonderen zusammengefasst werden.

Darüber hinaus werden weitere elf von Arzneimittelbewertungsagenturen herausgegebene methodologische Leitlinien identifiziert.

Insgesamt geben 13 der 23 analysierten Methodenpapiere der INAHTA-Mitglieder und sieben der elf analysierten Leitlinien von Arzneimittelbewertungsinstitutionen Empfehlungen zur Auswahl der in der Bewertung der Technologie zu verwendenden Ergebnisparameter. Dabei sind sich alle Institutionen darüber einig, dass in HTA-Berichten und Nutzenbewertungen patientenrelevanter Ergebnisparameter vorrangig berücksichtigt werden müssen. Alle Agenturen betonen in ihren Methodenpapieren, dass grundsätzlich die Verwendung von harten klinischen Endpunkten gegenüber Surrogatendpunkten zu bevorzugen ist, wobei die Mehrzahl der Institutionen beschreibt, dass bei der Erfüllung bestimmter Voraussetzungen die Verwendung von Surrogatendpunkten als Ausnahme akzeptabel sein kann. Dazu ist es im Wesentlichen erforderlich, dass die oben beschriebene Validität des verwendeten Endpunkts nachgewiesen werden kann. Dies soll i. d. R. durch die Vorlage zusätzlicher Evidenz erfolgen, die den kausalen Zusammenhang zwischen Surrogat und klinischem Endpunkt belegt. In keinem der hier berücksichtigten Methodenpapiere wird eine Liste von anerkannten Surrogatparametern formuliert.

#### 4.5.3 Analyse von HTA-Berichten

Insgesamt werden 140 internationale und 131 HTA-Berichte von deutschen Institutionen untersucht.

Die internationalen Berichte decken verschiedene Technologien ab, wobei das Assessment von medizinischen und chirurgischen Prozeduren in der Mehrheit der Berichte Thema ist. In weniger als der Hälfte der internationalen HTA-Berichte werden die im Assessment zu extrahierenden Ergebnisparameter bei der Formulierung der Fragestellung bzw. im Methodenteil explizit und detailliert festgelegt. Insgesamt wird in 87 (62 %) der internationalen HTA-Berichte die Verwendung von Surrogatparametern

tern geplant bzw. Surrogatparameter aus den für das Assessment vorliegenden Studien extrahiert und berichtet. Fast alle Berichte betrachten einen klinisch relevanten, patientennahen Endpunkt. Nur in fünf Berichten werden ausschließlich Surrogatendpunkte betrachtet. In diesen Berichten werden diagnostische Technologien bewertet; hier sind Surrogate Testcharakteristika.

Nicht in allen deutschen HTA-Berichten werden die im Assessment zu extrahierenden Ergebnisparameter bei der Formulierung der Fragestellung bzw. im Methodenteil explizit und detailliert festgelegt. In ca. einem Drittel der analysierten deutschen HTA-Berichte wird lediglich erwähnt, dass der Bericht sich beispielweise mit dem „Nutzen und Risiken“, „Wirksamkeit und Sicherheit“ beschäftigt, ohne genauer darzulegen ob und welche Ergebnisparameter prospektiv festgelegt werden. Insgesamt wird in 74 (56 %) der deutschen HTA-Berichte die Verwendung von Surrogatparametern geplant bzw. Surrogatparameter aus den für das Assessment vorliegenden Studien extrahiert und berichtet. Fast alle Berichte betrachten einen klinisch relevanten, patientennahen Endpunkt. Nur in sechs der 74 Berichte, die Surrogatparameter verwenden, werden keine weiteren Parameter berücksichtigt. Das heißt, dass von den 131 deutschen HTA-Berichten nur 4,5 % ausschließlich auf Surrogatendpunkten beruhen. In diesen Berichten werden diagnostische Technologien bewertet. Die verwendeten Surrogate waren die Testcharakteristika.

## **4.6 Diskussion**

### **4.6.1 Diskussion der Methodik**

Die Beantwortung der diesen Bericht leitenden Fragestellungen erfordert verschiedene methodische Ansätze. Den ersten zwei Fragen wird im Rahmen eines Literaturreviews nachgegangen. Es werden im Literaturreview drei aktuelle Übersichtsarbeiten identifiziert, die neben den durch die Recherche gefundenen Arbeiten auch weitere relevante Artikel zusammenfassen. Dadurch ist ein umfassender Überblick zu den Anforderungen an Surrogatparameter und zu Validierungsmethoden möglich.

Die Beantwortung der dritten Frage beruht auf der Analyse von methodischen Leitlinien internationaler Institutionen, die in HTA involviert sind. Damit stellt diese Arbeit einen repräsentativen Überblick dar, welchen Stellenwert internationale HTA-Organisationen Surrogatendpunkten einräumen.

Für die Beantwortung der vierten Frage werden deutsche HTA-Berichte sowie eine Stichprobe von HTA-Berichten internationaler HTA-Agenturen analysiert. Der Ansatz ermöglicht eine Abschätzung des Stellenwerts von Surrogatendpunkten in HTA-Berichten auf der Basis einer quantitativen Analyse.

### **4.6.2 Diskussion der Fragestellungen**

Damit ein Surrogatendpunkt als valider Endpunkt für die Bewertung des Nutzens einer Technologie berücksichtigt werden kann, müsste dieser mehrere Kriterien erfüllen. Aus diesem Grund reichen die Ergebnisse statistischer Validierungsverfahren allein nicht aus, um die Eignung eines Surrogatendpunkts zu beurteilen. Die Einbeziehung von Informationen über biologische und klinische Faktoren ist unverzichtbar. Die Validierung eines Surrogatendpunkts für die Bewertung des Nutzens ist in diesem Sinn technologiespezifisch. Es hängt vom Wirkmechanismus der Technologie ab, ob das Surrogat deren Effekte abbilden kann oder nicht. Dies ist unabhängig davon, ob es ein starker und konsistenter Zusammenhang zwischen Veränderungen im Surrogat sowie Veränderungen in den klinisch relevanten Endpunkten statistisch nachgewiesen worden ist.

Zusammenfassend zeigen die Empfehlungen der hier berücksichtigten Institutionen dahingehend eine kritische Einstellung zur Verwendung von Surrogatendpunkten in HTA. Klinisch relevante Endpunkte wie Mortalität oder Morbidität haben Vorrang, weil sie eine valide Beurteilung der Wirksamkeit und der Sicherheit ermöglichen, die die Grundlage jeglicher Nutzenbewertung darstellt. Die von den HTA-Institutionen gestellten Anforderungen und Überlegungen zur Akzeptanz eines Surrogatendpunkts im Rahmen von HTA-Berichten, sind im Einklang mit denen, die in der theoretischen Literatur diskutiert werden.

Aus unserer quantitativen Analyse lässt sich feststellen, dass Surrogatendpunkten ein sehr geringer Stellenwert in HTA-Berichten eingeräumt wird. Die Ergebnisse der Analyse der deutschen HTA-Berichte sind annähernd identisch zu denen aus der Stichprobe internationaler HTA-Berichte. Es liegt

kein Hinweis dafür vor, dass in Deutschland ein strengerer Standard als in den Ländern des internationalen Umfeldes beim Umgang mit Surrogatendpunkten praktiziert wird.

## **4.7 Schlussfolgerung**

Die Validierung eines Surrogatendpunkts erfordert ausführliche Forschung, einschließlich der Durchführung von RCT mit klinisch relevanten Endpunkten. Die Validität eines Surrogatendpunkts ist eher technologie- als krankheitsspezifisch, so dass die Ergebnisse der Validierung eines Surrogatendpunkts nicht auf andere Technologien übertragen werden können (auch nicht bei angeblich ähnlichem Wirkmechanismus). Um ein Höchstmaß an Sicherheit zu erreichen, muss die Validität eines Surrogatendpunktes bei jeder Technologie bzw. jedem Wirkstoff einzeln geprüft werden.

Nach wie vor ist der Einsatz von Surrogatendpunkten bei der Nutzenbewertung sehr kritisch zu betrachten.

## 5 Summary

### 5.1 Health policy background

The issue which are valid and acceptable parameters for the assessment of the health benefit of the application of health technologies is recurrently discussed among the different actors and stakeholders in the health system. In Germany, with the establishment of the Institute for Quality and Efficiency in Health Care (IQWiG) in 2004 the discussions on the methods for health technology assessment in general and on the use of surrogate endpoints in particular have recently gained actuality und publicity.

### 5.2 Scientific background

So called surrogate endpoints represent in the best case preliminary steps in the casual chain leading to the relevant outcome (e. g. mortality, morbidity) and are not usually directly perceptible by patients. Characteristics of surrogate endpoints are:

- They are measured in lieu of the actually relevant outcome of interest.
- They are usually biochemical markers, physiological parameters or subclinical endpoints which for the patient are not directly perceptible. However, they are correlated with relevant clinical endpoints (e. g. high blood pressure is associated with higher risk of stroke, high LDL-cholesterine (LDL = Low density lipoprotein) is a risk factor for a heart attack, the CD4-cell count is associated with AIDS mortality).
- Changes in the surrogate are easier to observe than changes in the related relevant endpoint (i. e. the occur earlier and more commonly).
- Surrogates are sometimes named intermediate or intermediary outcomes, since they represent an intermediate step in the casual chain leading to the clinical relevant endpoint.
- Surrogate parameters are statistically associated with the clinical relevant outcome and have prognostic power.
- The association between the surrogate and the relevant endpoint is plausible from a biological and pathophysiological point of view.

Surrogate endpoints are not only used in trials of pharmaceuticals but also in studies of other clinical technologies. Parameters with an intermediary character are also applied in the field of community and public health interventions.

Their use in the assessment of the benefit of a health technology is however problematic. In the past, reliance on surrogate outcomes has led to false conclusions concerning the effects of a technology on the relevant health outcome. In many situations, relying on the strong correlation observed between surrogate and relevant endpoint to find an intervention has had fatal consequences (i. e. positive effects on the surrogate but increased mortality with the intervention in question). The problematic is known since around 30 years. A classical example of the potential for fatal consequences as a result of reliance on surrogate is the case of class I antiarrhythmic drugs. Some drugs have been removed from market after the observation of an increased mortality or morbidity with their use, contrary to the expectations raised by the observation of positive effects on a surrogate endpoint. In other occasions, reliance on surrogate has also led to withholding effective therapies. For example, for many years betablockers – due to the bradycardic effect – were considered to be contraindicated in patients with heart failure, since following pathophysiological reasoning a reduction of the heart rate was thought to have deleterious effects in this patients.

## 5.3 Research questions

1. Which criteria need to be fulfilled for a surrogate parameter to be considered a valid endpoint?
2. Which methods have been described in the literature for the assessment of the validity of surrogate endpoints?
3. Which methodological recommendations concerning the use of surrogate endpoints have been made by international HTA agencies?
4. Which place has been given to surrogate endpoints in international and German HTA reports?

## 5.4 Methods

According to the above mentioned research questions, we follow different methodological approaches. In order to answer research questions 1 and 2, related to the concepts and the methods of surrogate validation we conduct a systematic review of methodological papers. Electronic databases are searched with the following terms:

SURROGATE END POINT; SURROGATE END POINTS; SURROGATE ENDPOINT; SURROGATE ENDPOINTS; ENDPOINT, SURROGATE; ENDPOINTS, SURROGATE; END POINT, SURROGATE; END POINTS,SURROGATE; BIOLOGICAL MARKER; BIOLOGICAL MARKERS; VALIDATION; STATISTICS; BIOMETRY; DECISION SUPPORT TECHNIQUES; ENDPOINT DETERMINATION; CAUSALITY

The methodological literature is summarised in a narrative review consisting of two parts: 1. an overview of the criteria to be fulfilled by a surrogate in order to be considered acceptable. 2. an overview of the statistical methods proposed in the literature for the validation of surrogates.

In order to answer research question 3 we analyse the methodological guidelines and recommendations of HTA agencies, member of the International network of agencies for Health Technology Assessment (INAHTA), and of agencies involved in pharmaceutical pricing and reimbursement decisions.

In order to answer research question 4 we analyse a random sample of HTA reports from the HTA database. We extracted the type of outcome parameter used and reported in the HTA reports. We analyse in the same way the full sample of HTA reports procured in Germany and registered in the database of the German Agency for HTA.

## 5.5 Results

### 5.5.1 Literature review

The literature search yields a total of N = 1,109 hits. After checking title and abstract, n = 2 duplicates and n = 1,007 references lacking references are excluded. A total of n = 100 papers is retrieved for more detailed analysis. At the end n = 25 methodological papers are summarised in the review on criteria and validation methods.

The criteria that a surrogate parameter need to fulfil in order to be recognized an acceptable and valid endpoint can be summarised as follows:

- **Biological plausibility:** There is evidence from animal models and epidemiological studies of a causal relationship between the surrogate parameter and the clinical relevant endpoint. The surrogate is part of the pathophysiological causal path leading to the health outcome.
- **Magnitude of the association between surrogate and relevant endpoint:** Epidemiological evidence has shown repeatedly and consistently that changes in the surrogate are qualitative and quantitative associated with changes in the relevant health outcome.
- **Evidence of effect form randomized controlled trials (RCT):** There is evidence from RCT showing that the changes induced by an intervention in the surrogate lead to changes in the relevant outcome in the same direction. The effect of the intervention is fully captured by the surrogate. Even in the case of very similar active principles, the mechanism of action may differ. Thus, the transferability of conclusions on the validity of a surrogate from one technology to another needs to be carefully assessed.

In the full report, we summarise the different statistical methods discussed in the literature for the validation of surrogate endpoints. In summary, we conclude that there is no goldstandard for the validation of surrogate endpoints. Since the generalisation of results from single studies is more prone to produce fallacies, approaches summarising results from several studies (i. e. meta-analysis) are preferred.

## **5.5.2 Analysis of methodological guidance from HTA agencies**

A total of 23 methodological papers from 14 INAHTA members (eleven countries) is identified. In addition, eleven further methodological guidelines from agencies involved in pricing and reimbursement decisions are found. We extract their recommendations concerning the selection of outcome parameters in general and the use of surrogate endpoints in particular.

A total of 13 from 23 analysed INAHTA member methodological papers' and seven of eleven from "fourth-hurdle agencies" provide information on how to choose outcome parameters for the assessment. All institutions agree, that patient relevant outcome parameters are strongly preferred in the assessment of the benefit of a health technology. All agencies underline that hard outcome parameters are to be preferred to surrogate endpoints. Nevertheless, the majority of agencies describes that under some circumstances surrogate endpoints may exceptionally be accepted – provided the validity of the surrogate is well established. In order to accept a surrogate, HTA agencies require the presentation of evidence which supports the causal relationship between surrogate and clinical relevant endpoint.

None of the methodological guidance papers from HTA agencies provided a list of well established/generally accepted surrogate endpoints.

## **5.5.3 Survey of HTA reports**

A total of 140 HTA reports from INAHTA members and of 131 HTA reports from German institutions is analysed.

The reports cover different types of technologies, although the assessment of medical and surgical interventions represent the majority. A prospective description (e. g. in the research questions or in the methods section) of the outcome parameters in which the assessment would be based is present in less than half of the analysed HTA reports. Surrogate endpoints are extracted and reported in 87 (62 %) HTA reports from the HTA database. Almost all HTA reports include also a clinical and patient relevant outcome. Only five reports use exclusively surrogate parameters, all of them assessing a diagnostic technology and being the surrogate test characteristics.

Similar results are obtained for the sample of German HTA-reports. Approximately one third of the German HTA reports describe to assess benefits and risks, or effectiveness and safety of the technology, without further describing how these terms were operationalised into outcome parameters. Surrogate endpoints are extracted and reported in 74 (56 %) German HTA reports. Almost all German HTA reports also consider a clinical and patient relevant parameter. Only six reports exclusively use surrogate parameters, all of them assessing a diagnostic technology and being the surrogate test characteristics.

## **5.6 Discussion**

### **5.6.1 Methods of the report**

According to the different nature of our research questions, we follow several approaches in this report.

The literature review represents a good overview of the field. Besides original works, we also identify three recent systematic reviews which summarise additional methodological papers.

We also provide a representative overview of the methodological guidance regarding surrogate endpoints form HTA agencies worldwide.

In addition, we also provide a representative picture of the actual consideration of surrogate endpoints in HTA reports from international and German HTA agencies. The representative survey allows to understand the value surrogate endpoints have in the field of HTA.

### **5.6.2 Results of the report**

In order to be considered valid and acceptable a surrogate needs to fulfil several criteria. Thus, favourable results from statistical validation approaches are not a sufficient condition to conclude on the validity of a surrogate endpoint. Information on biological and pathophysiological factors is also required. In addition, the validity of a surrogate is to be seen as technology-specific. Whether a surrogate is able to capture the full effect of a technology depends on the mechanism of action of the technology in question. This irrespective of whether a strong and consistent association between surrogate and relevant health outcome has been well established.

In summary, HTA agencies show a cautious position regarding the reliability of surrogate endpoints in HTA. Clinical relevant endpoints such as mortality and morbidity are preferred, since they allow a sound assessment of effectiveness and safety of the interventions. The considerations and recommendations provided by HTA agencies regarding the use of surrogates are conform to the discussions being held in the theoretical and methodological literature.

The quantitative analysis of the sample of HTA reports allows to conclude that surrogate endpoints have not been prominently used in HTA reports. The results are similar for the international sample and for the German sample, indicating that the handling of surrogate endpoints in Germany is not more stringent than in the international context.

## **5.7 Conclusions**

The validation of a surrogate endpoint requires extensive research, including RCT assessing clinical relevant endpoints. The validity of a surrogate parameter is rather technology-specific than disease-specific. Thus – even in the case of apparently similar technologies – it is necessary to validate the surrogate for every single technology (i. e. for every single active agent), in order to avoid false conclusions potentially leading to fatal consequences.

The use of surrogate endpoints in the assessment of the benefit of health technologies is still to be seen very critically.

## 6 Hauptdokument

### 6.1 Gesundheitspolitischer Hintergrund

Die Diskussionen, welche Ergebnisparameter für eine valide Bewertung des Nutzens medizinischer Technologien verwendet werden können, sind in Deutschland spätestens seit der Gründung des Instituts für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) in Deutschland wieder aufgelebt. Diese Frage beschäftigt alle an Entscheidungsfindungen und Regulierung beteiligten Akteure im Gesundheitssystem.

Auch auf internationaler Ebene wird eine ähnliche Diskussion geführt, die das Policy Forum der internationalen HTA-Gesellschaft (Health Technology Assessment international (HTAi)) im September 2008 veranlasst hat, eine Arbeitsgruppe zu Surrogatoutcomes zu gründen. Ziel dieser Arbeitsgruppe ist, eine Debatte über die Anwendung von Surrogaten in Health Technology Assessment (HTA) zu führen<sup>44</sup>. Es liegen zurzeit keine Ergebnisse von dieser Gruppe vor.

#### 6.1.1 Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen (IQWiG) und Gemeinsamer Bundesausschuss (G-BA)

Das IQWiG wird im Zuge des Gesetzes zur Modernisierung der Gesetzlichen Krankenversicherung (GMG) 2004 gegründet. Zu den Aufgaben des Instituts gehört u. a. die Bewertung medikamentöser und nicht-medikamentöser Verfahren im Zusammenhang mit deren Einschluss in den Leistungskatalog bzw. Erstattungsfähigkeit zu Lasten der Gesetzlichen Krankenversicherung (GKV). Das Bewertungsverfahren beschränkt sich gemäß Sozialgesetzbuch V (SGB V) bis 2007 auf die ausschließliche Untersuchung des Nutzens einer Technologie, wobei die Darstellung des Schadenpotenzials wesentlicher und gleichberechtigter Bestandteil der Nutzenbewertung einer Intervention ist<sup>51</sup>. Mit der im Gesetz zur Stärkung des Wettbewerbs in der Gesetzlichen Krankenversicherung (GKV-WStG) vorgenommenen Änderung des § 35b SGB V und des § 139a SGB V wird das Aufgabengebiet des IQWiG erweitert. Seither gehört die Bewertung des Kosten-Nutzen-Verhältnisses von Arzneimitteln auch zu den Aufgaben des Instituts. Diese Bewertungen können für jedes erstmals verordnungsfähige Arzneimittel mit patentgeschützten Wirkstoffen sowie für andere Arzneimittel, die von Bedeutung sind, erstellt werden. Das IQWiG handelt hauptsächlich im Auftrag des G-BA gemäß § 139b SGB V, wobei grundsätzlich die den G-BA bildenden Institutionen, das Bundesministerium für Gesundheit (BMG), die für die Wahrnehmung der Interessen der Patientinnen und Patienten sowie der Selbsthilfe chronisch kranker und behinderter Menschen maßgeblichen Organisationen sowie die oder der Beauftragte der Bundesregierung für die Belange der Patientinnen und Patienten die Beauftragung des Instituts beim G-BA beantragen können. Darüber hinaus ist auch eine direkte Beauftragung des IQWiG durch das BMG möglich. Das IQWiG leitet die Arbeitsergebnisse der Aufträge dem G-BA in Form von Abschlussberichten mit einer Bewertung des Nutzens zu. Der G-BA ist angehalten, diese Ergebnisse im Rahmen seiner Aufgabenstellung bei seinem Entscheidungsfindungsprozess zu berücksichtigen.

Darüber hinaus führt der G-BA selbst Bewertungen neuer Untersuchungs- und Behandlungsmethoden für den ambulanten und stationären Sektor durch, ohne das IQWiG zu beauftragen. Wesentliche Bewertungskriterien sind dabei die gesetzlich vorgegebenen Aspekte *diagnostischen oder therapeutischen Nutzens, medizinische Notwendigkeit* und *Wirtschaftlichkeit* (§ 92 Abs. 1 SGB V und § 19 der Verfahrensordnung).

#### 6.1.2 Richtlinien des Gemeinsamen Bundesausschusses (G-BA)

Auf Basis der Ergebnisse der vom IQWiG oder G-BA durchgeführten Nutzenbewertungen entscheidet der G-BA in Form von Richtlinien gemäß § 92 SGB V, welche Leistungen im ambulanten und stationären Sektor zu Lasten der GKV erbracht oder verordnet werden können bzw. welche Leistungen vollständig oder in Teilen ausgeschlossen werden. Darüber hinaus stellen zukünftig die Ergebnisse der Kosten-Nutzen-Bewertung von Arzneimitteln, die nicht in eine Festbetragsgruppe gemäß § 35 SGB V einzubeziehen sind (sogenannte *Innovationen*), die Grundlage für die Festlegung eines Höchstbetrags dar, bis zu dem die GKV die Arzneimittelkosten tragen muss. Dieser Höchstbetrag wird vom Spitzen-

verband der Krankenkassen (SBK) festgelegt, wobei auch die Möglichkeit besteht, dass sich die Hersteller und der SBK ohne die Durchführung einer Kosten-Nutzen-Bewertung auf einen Preis verständigen können (§ 31 Abs. 2a SGB V). Grundsätzlich sind Arzneimittel, deren Kosten-Effektivität erwiesen ist oder für die eine Kosten-Nutzen-Bewertung nur im Vergleich zur Nichtbehandlung erstellt werden kann, weil eine zweckmäßige Therapiealternative fehlt, von der Festsetzung eines Höchstbetrags ausgenommen. Außerdem kann eine Kosten-Nutzen-Bewertung als Grundlage für die Festsetzung eines Höchstbetrags erst dann erstellt werden, wenn hinreichende Erkenntnisse über die Wirksamkeit des Arzneimittels nach den Grundsätzen der evidenzbasierten Medizin vorliegen können.

### 6.1.3 Nutzen und Nutzenbewertung im Kontext des Sozialgesetzbuchs (SGB V)

Sowohl der G-BA als auch das IQWiG arbeiten im Rahmen ihrer Bewertungen auf Grundlage eigener Verfahrensordnungen<sup>36, 51</sup>. Die Begrifflichkeit des Nutzens einer medizinischen Intervention wird dabei in der Verfahrensordnung des G-BA nicht explizit definiert. Es wird allerdings in § 17 der Verfahrensordnung erklärt, dass für die Überprüfung des Nutzens einer Methode Unterlagen zum Nachweis der *Wirksamkeit* bei den beanspruchten Indikationen, zum Nachweis der *therapeutischen Konsequenz* einer diagnostischen Methode, zur *Abwägung des Nutzens gegen die Risiken*, zur Bewertung der *erwünschten und unerwünschten Folgen (outcomes)* und zum *Nutzen im Vergleich zu anderen Methoden gleicher Zielsetzung* zugrunde gelegt werden müssen<sup>36</sup>. Das IQWiG liefert in seinem Methodenpapier hingegen sowohl eine Definition für den Begriff der *Nutzenbewertung* als auch des *Nutzens* bzw. *Zusatznutzens* einer Technologie. Unter einer **Nutzenbewertung** wird demnach der gesamte Prozess der Evaluation medizinischer Interventionen hinsichtlich ihrer kausal begründeten positiven und negativen Effekte – im Vergleich zu einer klar definierten anderen Therapie, einem Placebo oder keiner Behandlung – verstanden<sup>51</sup>. Der Begriff **Nutzen** definiert dabei kausal begründete positive Effekte einer medizinischen Intervention auf patientenrelevante Endpunkte<sup>51</sup>. Es wird dann von einem kausal begründeten Effekt ausgegangen, wenn ausreichend Sicherheit besteht, dass beobachtete Effekte allein auf die zu prüfende Intervention zurückgeführt werden können. Während Nutzen per se immer im Vergleich zu Placebo oder keiner Behandlung festgestellt wird, wird im Fall eines Vergleichs der zu evaluierenden medizinischen Intervention mit einer anderen eindeutig definierten medizinischen Intervention geprüft, ob ein **Zusatznutzen** oder ein **geringerer bzw. vergleichbarer Nutzen** nachgewiesen werden kann<sup>51</sup>.

### 6.1.4 Endpunkte in der Nutzenbewertung im Kontext des Sozialgesetzbuchs (SGB V)

Gemäß § 35b SGB V soll das IQWiG für die Bewertung des Nutzens von Arzneimitteln insbesondere untersuchen, inwiefern durch die zu bewertende Therapie eine Verbesserung des Gesundheitszustands, eine Verkürzung der Krankheitsdauer, eine Verlängerung der Lebensdauer, eine Verringerung der Nebenwirkungen sowie eine Verbesserung der Lebensqualität der Patienten erreicht werden können. Darüber hinaus haben sowohl der G-BA als auch das IQWiG in ihren Verfahrensmethoden festgelegt, dass für die Nutzenbewertung einer medizinischen Intervention wissenschaftliche Untersuchungen zur Beeinflussung der **patientenrelevanten Endpunkte Mortalität, Morbidität und gesundheitsbezogener Lebensqualität** zugrunde gelegt werden müssen<sup>36, 51</sup>. Ergänzend berücksichtigt das IQWiG auch den interventions- und erkrankungsbezogenen Aufwand sowie die Patientenzufriedenheit im Sinn sekundärer Zielgrößen. Dabei wird gefordert, dass die verwendeten Endpunkte zuverlässig und direkt konkrete Änderungen des Gesundheitszustands abbilden müssen und dass für die Erfassung der Lebensqualität und der Patientenzufriedenheit nur Instrumente eingesetzt werden sollen, die für den Einsatz in klinischen Studien geeignet und entsprechend evaluiert sind. **Surrogatendpunkte**, also Endpunkte, die in der medizinischen Forschung häufig als Ersatz für patientenrelevante Endpunkte verwendet werden, um Aussagen zum patientenrelevanten Nutzen bzw. Zusatznutzen früher und einfacher zu erhalten, werden vom IQWiG nur unter bestimmten Voraussetzungen für die Nutzenbewertung berücksichtigt. So muss beispielsweise anhand geeigneter statistischer Methoden gezeigt worden sein, dass der Effekt von Interventionen mit einem vergleichbarem Wirkmechanismus auf

den zu ersetzenden patientenrelevanten Endpunkt durch den Effekt auf den Surrogatendpunkt in einem ausreichenden Ausmaß erklärt wird<sup>51</sup>.

Auch eine auf Einladung des BMG im November 2007 im Anschluss an eine Fachtagung zur Einführung der Kosten-Nutzen-Bewertung durchgeführte wissenschaftliche Klausurtagung mit deutschen sowie internationalen Wissenschaftlerinnen und Wissenschaftlern kam grundsätzlich zu dem Schluss, dass im Rahmen von Nutzenbewertungen „direkte Endpunkte“, wie z. B. Myokardinfarkt oder Mortalität, Surrogatendpunkten vorzuziehen sind<sup>3</sup>. Dennoch wird darauf hingewiesen, dass es Situationen geben kann, in denen keine hinreichende Evidenz für den gesundheitlich relevanten Endpunkt vorliegt und, dass die Frage, ob ein Surrogatendpunkt Verwendung finden kann, für jede medizinische Situation neu entschieden werden muss. Die Validität eines solchen Endpunkts könne man dabei nicht ausschließlich unter statistischen Gesichtspunkten beurteilen. Vielmehr sind auch Kriterien wie die Qualität der verfügbaren Studien, pharmakologisches Wissen und der gewählte Komparator zu berücksichtigen. Im Ergebnis ist es aus Sicht der Workshop-Teilnehmer nicht vertretbar, Evidenz auf der Basis von Surrogatendpunkten unbeachtet zu lassen<sup>3</sup>. Vielmehr schlagen sie in Bezug auf die kommenden Kosten-Nutzen-Bewertungen von Arzneimitteln vor, bei starker Evidenz für einen Surrogatendpunkt und gleichzeitig fehlender oder schwacher Evidenz für den gesundheitlich relevanten Endpunkt zu einer vorläufigen positiven Kosten-Nutzen-Einschätzung zu kommen und parallel zu fordern, dass nach einer festgelegten Zeit belastbare Studien und Daten für den relevanten Endpunkt nachgeliefert werden müssen<sup>3</sup>. Diese vorläufige positive Erstattungsentscheidungen mit der Auflage in einer angemessenen Zeitrahmen weitere Evidenz zu generieren werden in der internationalen Diskussion - als „coverage with evidence development“ (CED) bezeichnet - kontrovers diskutiert als eine Möglichkeit, den Zugang zu vielversprechenden Technologien in Situationen, in denen wichtige Evidenzlücken bestehen (z. B. bzgl. ihrer Kosten-Effektivität)<sup>47</sup> zu erlauben.

Vor dem Hintergrund dieser laufenden Diskussion ist es von Bedeutung, einen umfassenden Überblick zu erhalten, welche wissenschaftlichen Kriterien der Validierung von Surrogatendpunkten zugrunde gelegt werden und wie international mit Surrogatendpunkten bei der Bewertung von medizinischen Technologien im Rahmen von HTA-Berichten umgegangen wird. Das Thema fand daher Eingang in den Themenfindungs- und Prioritätensetzungsprozess der Deutschen Agentur für Health Technology Assessment beim Deutschen Institut für Medizinische Dokumentation und Information (DAHTA beim DIMDI), die einen entsprechenden Bericht in Auftrag gab.

## 6.2 Wissenschaftlicher Hintergrund

Auch wenn der ideale Endpunkt einer methodisch hochwertigen Studie anerkanntermaßen klinisch relevant sein sollte, führen die für ihre Durchführung benötigten hohen Fallzahlen, bzw. die in der Regel langen Beobachtungszeiten dieser Studien dazu, dass in der medizinischen Forschung stattdessen vielfach sogenannte Surrogatparameter als Endpunkte verwendet werden<sup>5, 30, 70</sup>. Darüber hinaus, kann der Effekt auf Surrogatparameter nicht nur schneller sondern viel stärker nachweisbar sein, als bei der Untersuchung klinisch relevanter Endpunkte. Das Vorziehen von Surrogatendpunkten kann deshalb auch darauf beruhen, dass Hersteller bzw. Befürworter einer Technologie befürchten, dass bei einem Vergleich anhand von klinisch relevanten Endpunkten kein zusätzlicher Nutzen nachweisbar wäre. Während der Einsatz von Surrogatendpunkten in Phase I und II der klinischen Prüfung bei der Entwicklung von Arzneimitteln selbstverständlich ist, wird ihre Verwendung in der Nutzenbewertung vielfach kritisiert und die Verlässlichkeit bzw. die Prognosekraft dieser Parameter häufig angezweifelt<sup>37, 38</sup>.

### 6.2.1 Surrogatendpunkte – Definitionen

Während die klinisch relevanten Endpunkte jene sind, die für den Patienten von Bedeutung sind (z. B. Morbidität, Mortalität), stellen Surrogatendpunkte im besten Fall Vorstufen des eigentlichen klinisch relevanten Endpunkts dar, die für den Patienten in der Regel nicht unmittelbar spürbar sind.

Es gibt zahlreiche Definitionen für diese Art von Endpunkten, allen gemeinsam sind die folgenden Aspekte:

- Es handelt sich um Zielgrößen, die anstelle von (*in lieu*) den eigentlichen Endpunkten gemessen werden.
- Es handelt sich häufig um physiologische oder biochemische Marker bzw. subklinische Endpunkte, die für den Patienten in der Regel nicht unmittelbar von Bedeutung bzw. in irgendeiner Weise bemerkbar sind, jedoch stellvertretend für wichtige klinische Endpunkte stehen können (z. B. erhöhter Blutdruck steigert das Risiko für einen Schlaganfall, erhöhte LDL-Cholesterinwerte (LDL = Low density lipoprotein) stellen ein Risikofaktor für Myokardinfarkt, der Anzahl der CD4-Zellen für AIDS-Mortalität (AIDS = Acquired Immuno Deficiency Syndrome)).
- Veränderungen dieser Parameter sind im Vergleich zum Auftreten der relevanten Endpunkte (Morbidity, Tod) häufiger und nach kürzerer Beobachtungszeit feststellbar.
- Sie stellen Zwischenschritte auf dem Weg von der Erkrankung zum klinischen Endpunkt dar und werden deshalb mitunter auch als intermediate oder intermediary outcomes bezeichnet.
- Es gibt einen (statistisch nachweisbaren) Zusammenhang zwischen dem Surrogatendpunkt und dem eigentlichen klinischen Endpunkt im Sinne einer guten Prognosekraft.
- Der Zusammenhang zwischen Surrogat und klinisch relevantem Endpunkt muss biologisch und pathophysiologisch (d. h. bzgl. der Funktion des Organismus unter Krankheit) plausibel sein.

Surrogatparameter werden nicht nur in Studien über die Wirksamkeit von Arzneimitteln eingesetzt. Auch Studien über die Effekte von medizinischen und chirurgischen Prozeduren berichten häufig über Ergebnisparameter, die nicht das ultimative, patientennahe Ziel der Intervention wiedergeben.

Bei nicht-klinischen Interventionen (z. B. Patientenschulungen, bevölkerungsbezogenen Interventionen etc.) werden auch häufig Zielgrößen gemessen, die einen intermediären Charakter haben. Dabei wird das Erreichen von Zwischenstufen des theoretischen Modells, das der Intervention zugrunde liegt, abgebildet. Die tatsächlichen Gesundheitseffekte bleiben jedoch ungemessen. Beispielsweise bei der Intervention „Patientenschulung“, würde die Messung des „Wissensgrades über die eigene Krankheit“ einen Effekt erfassen, von dem erwartet wird, dass er zur Verbesserung des Gesundheitszustands beiträgt. Diese Annahme basiert auf dem der Intervention zugrunde liegenden theoretischen Modell, dass durch bessere Kenntnisse über die Krankheit der Patient befähigt wird das Verhalten so zu ändern, dass das Risiko von Komplikationen oder Verschlechterungen – das sind die relevanten Gesundheitsoutcomes – reduziert wird. Im Rahmen von Public health-Maßnahmen weisen Prozessindikatoren einige Charakteristika von Surrogatparametern auf: Sie werden anstelle der eigentlich relevanten Endpunkte gemessen, stellen Zwischenschritte bzw. Voraussetzungen für das Erreichen der relevanten Endpunkte dar und treten schneller auf, so dass Veränderungen schneller erfassbar sind.

Tabelle 1 zeigt einige Beispiele von Surrogatparametern bzw. intermediären Zielgrößen aus verschiedenen Gebieten.

**Tabelle 1: Beispiele von Surrogatzielgrößen für verschiedene Indikationen**

Kondition (Interventionen)	Surrogatzielgröße (Quelle)	Relevante, definitive Endpunkte
AIDS/HIV (antivirale Therapie)	Anzahl CD-4 Zellen <sup>46</sup> Nachweis von Virus RNA <sup>46</sup>	Überleben
Arterieller Hypertonus (medikamentöse Therapien, körperliche Betätigung)	Blutdruck <sup>64, 84</sup>	Inzidenz von Schlaganfall Inzidenz von Myokardinfarkt Überleben
Hyperlipidämie (medikamentöse Therapien)	Lipidspiegel <sup>84</sup>	Inzidenz von Myokardinfarkt Überleben
Diabetes (medikamentöse Therapien, DMP)	Blutzuckerspiegel <sup>64, 84</sup> HbA1c <sup>64, 84</sup>	Inzidenz von Komplikationen Mortalität Lebensqualität
Chronische Erkrankungen (Patientenschulungen, DMP)	Wissen über die Krankheit <sup>64</sup> Verhaltensänderungen <sup>64</sup> Patienteneinstellungen <sup>64</sup>	Inzidenz von Komplikationen Inzidenz von Exazerbationen Mortalität Lebensqualität

**Fortsetzung Tabelle 1: Beispiele von Surrogatzielgrößen für verschiedene Indikationen**

Krebs (alle Formen der Therapie, Screening)	Tumor response/Tumorreduktion <sup>13</sup> PSA <sup>23</sup> Histologische Veränderungen <sup>77</sup>	Krankheitfreiesüberleben Überleben Inzidenz von Krebs
Dengue (Interventionen auf Gemeindeebene zur Dengue-Kontrolle)	Messungen des Grads an Infestation mit Moskitoarven (z. B. House-Index, Breteau-Index, Conatainer-Index) <sup>40</sup>	Inzidenz von Dengue Mortalität durch Dengue
Verkehrssicherheit (z. B. Kindersitzpflicht, Verkehrserziehung)	Besitz von Kindersitz <sup>92</sup> Benutzung des Kindersitzes <sup>92</sup> Besitz von Fahrradhelm <sup>75</sup> Benutzung des Fahrradhelms <sup>75</sup>	Mortalität bei Verkehrsunfällen Verletzungen bei Verkehrsunfällen Behinderung nach Verkehrsunfällen

**Fettdruck:** Zielgröße die in der Arzneimittelmarktzulassung akzeptiert wird, nach Angaben der FDA<sup>32</sup> und Temple<sup>84</sup>.

AIDS = Acquired Immuno Deficiency Syndrome. CEA = Carcino-embryonales Antigen. DMP = Disease Management Programme. FDA = Food and Drug Administration. HbA1c = Glykosiliertes Hämoglobin. HIV = Human Immuno Deficiency Virus. PSA = Prostataspezifisches Antigen. RNA = Ribonucleinsäure.

Quelle: Eigene Darstellung mit Information aus den zitierten Arbeiten

Nach dem von Fryback und Thornbury vorgeschlagenen hierarchischen Modell zur Evaluierung von diagnostischen Tests, stellt die Bestimmung von Testcharakteristika eine Zwischenstufe in der Bewertung von diagnostischen Technologien dar<sup>34</sup>. In diesem Sinn, sind Parameter wie Sensitivität, Spezifität oder die diagnostische Genauigkeit eine Art Surrogatendpunkt im Vergleich zu den endgültigen Zielen der diagnostischen Technologie, nämlich der Verbesserung der Überlebenschancen bzw. der Lebensqualität der Patienten<sup>88</sup>. Der Wert einer diagnostischen Technologie liegt darin, dass durch eine korrekte Diagnose bzw. Abschätzung des Schweregrades oder eines Stadiums einer Krankheit, die Auswahl der angemesseneren Therapie bzw. weiteres Management ermöglicht wird, und dieses sich letztendlich in besseren patientenrelevanten Outcomes niederschlägt.

## 6.2.2 Surrogatendpunkte – Problematik

Die Anwendung von Surrogatendpunkten bei der Bewertung einer Therapie ist insbesondere bei Erkrankungen mit hoher Letalität und geringen therapeutischen Möglichkeiten häufig befürwortet worden. In diesen Situationen, entsteht ein Handlungsdruck, der das Warten auf die langwierige Entwicklung neuer Therapieoptionen unangemessen erscheinen lässt<sup>43</sup>. Ein klassisches Beispiel hierfür ist das Auftreten der Humanen Immundefizienz Virus (HIV)-Infektion und des dazugehörigen erworbenen Immundefizienz-Syndroms. Die Beurteilung der Wirksamkeit der retroviralen Therapie anhand der Anzahl von CD-4 – die mit der Progression der Erkrankung und die Mortalität korreliert – ist effektiver, als auf das Auftreten von opportunistischen Infektionen bzw. Todesfällen zu warten<sup>28</sup>.

Diesen Situationen wird in den Marktzulassungsanforderungen Rechnung getragen. Die amerikanische Food and Drug Administration (FDA) veröffentlicht 1993 eine Richtlinie, um eine schnelle Zulassung (*accelerated approval*) von Arzneimitteln zu ermöglichen, deren Ziel die Behandlung von schweren, lebensbedrohlichen Erkrankungen ist, für die es bisher keine Therapie gibt bzw. die eine Behandlung von spezifischen Patientengruppen (z. B. bei Unverträglichkeit der existierenden Behandlung) ermöglichen sollen<sup>31</sup>. Diese Richtlinie, die später (1997) gesetzlich festgeschrieben wird, ermöglicht die Zulassung auf Basis von Surrogatendpunkten, die nicht als etabliert gelten, was eine Herabsetzung der bis dahin geltenden Zulassungsstandard in besonderen Fällen bedeutet<sup>84</sup>. Die frühere Prozedur sah die Möglichkeit der Marktzulassung auf der Grundlage der akzeptierten Surrogatendpunkte vor, während der *accelerated approval* die Möglichkeit einräumt, die Zulassung auf der Basis von Evidenz über Effekte auf ein Surrogatendpunkt, der „*reasonably suggests clinical benefit*“ oder auf der Basis von Evidenz über Effekte auf andere klinische Zielgrößen als Mortalität oder unumkehrbarer Morbidität<sup>31</sup>. In einem Papier zur Orientierung für Arzneimittelhersteller bzw. Antragsteller, gibt die FDA zwei Beispiele von akzeptierten Surrogatendpunkten, nämlich Blutdruck und Cholesterinspiegel, sowie zwei Beispiele von Surrogatendpunkten, die einen klinischen Effekt suggerieren, nämlich die Anzahl von CD-4 und die virale Last bei HIV<sup>32</sup>. Letztere ermöglichen eine Zulassung mittels der beschleunigten Prozedur, die allerdings die Durchführung von weiteren Studien zur Überprüfung der relevanten klinischen Effekte als Bedingung stellt<sup>84</sup> (sogenannten *conditional approval*). Die FDA

warnen aber auch vor den Risiken von Surrogatendpunkten und weist darauf hin, dass ihr Nutzen größer ist, wenn Erfahrungen mit der Wirkstoffklasse vorliegen<sup>32</sup>.

Die European Medicines Agency (EMA) erwähnt die Anwendung von Surrogatparametern im Zulassungsprozess in den International Conference on Harmonisation (ICH) Richtlinien ICH 8 und ICH 9 ebenfalls. Dort wird festgelegt, dass „*Surrogate endpoints may be used as primary endpoints when appropriate (when the surrogate is reasonably likely or well known to predict clinical outcome)*“<sup>49</sup>. Surrogatendpunkte dürfen verwendet werden, wenn die Durchführung von Studien mit anderen Endpunkten nicht praktikabel (practical) erscheint<sup>50</sup>. Allerdings verweisen auch diese Leitlinien auf die Limitationen von Surrogatendpunkten<sup>50</sup>:

- Surrogatendpunkte sagen eventuell nicht wirklich viel über das relevante, echte (*true*) Ziel der Behandlung aus, wobei hier sowohl eine Über- als eine Unterschätzung des Effekts möglich ist, da der Surrogat nur einen Effektmechanismus abbildet, nicht jedoch alle potentiellen Effekte der Intervention.
- Surrogatendpunkte sind schwer in einer Nutzen-Risiko-Abwägung einzubauen.

Die Verwendung von Surrogatendpunkten in Marktzulassungsverfahren ist jedoch nicht unproblematisch. Es ist somit erlaubt, Arzneimittel auf den Markt zu bringen, ohne dass ihre Wirksamkeit anhand bedeutsamer Zielgrößen belegt sei und es beschränkt die Sicherheitsprüfung auf eine kleinere Anzahl von Patienten<sup>84</sup>. In der Vergangenheit gab es viele Situationen, in denen das Vertrauen auf Surrogatendpunkte trotz starker Korrelation mit dem klinischen Endpunkt irreführend war bzw. auch fatale Folgen hatte. Diese Problematik ist mittlerweile mehr als 30 Jahre bekannt. Der erwartete Nutzen einer Senkung der Cholesterinspiegel aufgrund eines zwischen erhöhten Cholesterinwerten und koronarer Herzerkrankung (KHK) beobachteten Zusammenhangs wird 1978 in der WHO-Studie (WHO = Weltgesundheitsorganisation) zur Prävention des Herzinfarkts mittels Clofibrat widerlegt: diese Arznei führt zu einer 44 % höheren Gesamtmortalität als Placebo<sup>11, 78</sup>. Prominentes und mittlerweile auch klassisches Beispiel sind die Klasse I-Antiarrhythmika<sup>14</sup>. Obwohl ventrikuläre Arrhythmien mit einer schlechteren Prognose bei Myokardinfarktpatienten verbunden sind und Klasse I-Antiarrhythmika diese Arrhythmien effektiv verhindern, zeigt sich in klinischen Studien eine erhöhte Mortalität unter dieser Medikation<sup>27</sup>. Einige Produkte, die auf der Basis von Surrogatendpunkten zugelassen wurden, mußten später vom Markt genommen werden, wenn in Studien mit Mortalitäts- bzw. Morbiditätendpunkten, die Nutzen-Risiken-Abwägung sich umkehrte<sup>30, 71</sup>. Aber ebenso kann das Festhalten an Surrogatendpunkten auch dazu führen, dass therapeutische Chancen nicht genutzt werden. So galten beispielsweise Betablocker für lange Zeit bei Patienten mit Herzinsuffizienz als kontraindiziert, da sie zu einer Verringerung der Herzfrequenz führen, was unter pathophysiologischen Überlegungen zu negativen Effekten bei diesen Patienten führen sollte. Erst Ende der 1990er Jahre wird deutlich, dass Betablocker trotz dieser Eigenschaft die Mortalität bei Herzinsuffizienzpatienten verringern<sup>27, 42, 85, 87</sup>. Weitere Beispiele für Trugschlüsse, die auf dem Vertrauen zu Surrogatparametern beruhen, sind von verschiedenen Autoren zusammengetragen (siehe Tabelle 2, ohne Anspruch auf Vollständigkeit).

**Tabelle 2: Fehler bei der Bewertung von Arzneimitteln auf der Grundlage von Surrogatendpunkten**

Kondition	Intervention	Surrogatergebnis (+/-)*	Klinisch relevanter Endpunkt Ergebnis
Kardio- vaskuläre Erkrankungen	Lipidsenkung mit Clofibrat	Senkung von Serumcholesterin (+)	Reduktion der Inzidenz von KHK Erhöhung der Gesamtmortalität
	Antiarrhythmika	Suppression von Extrasystolen (+)	Erhöhung der Gesamtmortalität und des Herzstillstands
	Hormonbehandlung in der Menopause**	Senkung des Serumcholesterins (+) Senkung des Blutzuckerspiegels (+)	Erhöhung des Herzinfarkt- und Schlaganfallrisikos
Granulomatose bei Kindern	Interferon	Keine Verbesserung der Produk- tion von Superoxid und der Makrophagenfähigkeit Erreger zu beseitigen (-)	Vermindert die Inzidenz von schweren Infektionen
Osteoporose	Natrium-Fluorid	Verbesserung der Knochendichte (+)	Erhöhte Frakturinzidenz

**Fortsetzung Tabelle 2: Fehler bei der Bewertung von Arzneimitteln auf der Grundlage von Surrogatendpunkten**

Demenz	Hormonbehandlung in der Menopause**	Verbesserung des zerebralen Blutfluss (u. a.)	Verschlechterung kognitiver Fähigkeiten
Kolorektaler Krebs	Chemotherapie (5-Fluoruracil + Leucovorin)	Verbesserung des Ansprechens des Tumors auf Therapie (komplette und partielle Remission)	Keine Unterschiede im Überleben
HIV/AIDS	Zidovudine bei HIV+ asymptomatisch	Verlust von CD4-Zellen verlangsamt (+)	Progression zu AIDS keine Unterschiede Höhere Mortalität
Lungenkrebs	Vitamine	Erhöhung des Vitaminspiegels	Erhöhte Lungenkrebsinzidenz Erhöhte Gesamtmortalität
Inkontinenz	Hormonbehandlung in der Menopause**	Verbesserung der Zellfunktionen im Urogenitalbereich	Verschlechterung der Inkontinenz
Intensivmedizin (Hypovolämie)	Albumininfusionen	Erhöhung der Albuminkonzentration (+)	Keine Reduktion der Mortalität, Hinweis auf erhöhte Mortalität
Anämie bei Nierenerkrankungen	EPO	Erhöhung von Hämoglobin (+)	Keine Reduktion der Mortalität, Hinweise auf erhöhte Mortalität

\* (+) positiver Effekt, (-) negativer Effekt.

\*\* Sogenannte Hormonersatztherapie (Hormone Replacement Therapy).

AIDS = Acquired Immuno Deficiency Syndrome. EPO = Erythropoietin. HIV = Human Immuno Deficiency Virus. KHK = Koronare Herzerkrankung.

Quellen: Böger RH<sup>11</sup>. Endpunkte klinischer Studien: Surrogatparameter oder „harte klinische Endpunkte“? Internist 2002; 43(4): 493-497.

Fleming TR, de Mets DL<sup>30</sup>. Surrogate end points in clinical trials: are we being misled? Annals of Internal Medicine 1996; 125(7):605-613.

Gotzsche PC, Liberati A, Torri V, Rossetti L<sup>37</sup>. Beware of surrogate outcome measures. International Journal Technology Assessment in Health Care 1996; 12(2):238-246.

Mühlhauser I, Berger M<sup>60</sup>. Surrogat-Marker: Trugschlüsse. Deutsches Ärzteblatt 1996; 93(49): A3280-A3283.

Mühlhauser I, Meyer G<sup>61</sup>. Evidence Based Medicine: Widersprüche zwischen Surrogatergebnissen und klinischen Endpunkten. Psychotherapie Psychosomatik Medizinische Psychologie 2006; 56: 193-201.

Schmidt J<sup>78</sup>. Risiko und Nutzen einer Behandlung: Das Beispiel Clofibrat. Schweizerisch medizinische Wochenschrift 1987; 117(20): 773-776.

Strippoli GFM, Navaneethan SD, Craig JC.<sup>81</sup> Haemoglobin and haematocrit targets for the anaemia of chronic kidney disease. Cochrane Database of Systematic Reviews 2006, Issue 4. Art. No.: CD003967. DOI: 10.1002/14651858.CD003967.pub2.

The Albumin Reviewers<sup>85</sup> (Alderson P, Bunn F, Li Wan Po A, Li L, Roberts I, Schierhout G). Human albumin solution for resuscitation and volume expansion in critically ill patients Cochrane Database of Systematic Reviews 2004, Issue 4. Art. No.: CD001208. DOI:10.1002/14651858.CD001208.pub2.

Diese Beispiele machen deutlich, dass es offensichtlich weitere, über eine Korrelation hinausgehende Kriterien geben muss, um einen Surrogatparameter stellvertretend für einen klinisch relevanten Endpunkt einsetzen zu können. Die Auswahl eines geeigneten und aussagekräftigen Surrogatparameters sollte deshalb grundsätzlich sehr kritisch erfolgen. Starke Korrelationen zwischen einem Surrogat- und einem klinischen Parameter müssen nicht zwingend eine gute Prognosekraft für patientenrelevante Endpunkte und für die Nutzenbewertung bedeuten, nicht zuletzt weil der Surrogat nicht alle möglichen Effekte abbildet.

### 6.3 Forschungsfragen

Das Ziel dieser Arbeit besteht zum einen darin, im Rahmen eines Reviews der publizierten methodischen Arbeiten herauszuarbeiten, a) welche Kriterien ein Surrogatparameter erfüllen muss, um als valider Endpunkt anerkannt zu werden und b) welche Methoden zur Validierung von Surrogatendpunkten eingesetzt werden (im Sinn von Verwendbarkeit als Alternative zu klinisch relevanten Endpunkten). Des Weiteren soll analysiert werden, welche methodischen Vorgaben hinsichtlich des Umgangs mit Surrogatendpunkten internationale Agenturen aus dem Bereich HTA und Arzneimittelnutzenbewertung machen und welchen Stellenwert Surrogatendpunkte in HTA-Berichten haben, d. h. in wie weit diese Art von Endpunkten bei Bewertungen berücksichtigt werden.

Aus den beschriebenen Zielen der Arbeit ergeben sich die folgenden Forschungsfragen:

1. Welche Kriterien muss ein Surrogatparameter erfüllen, um als valider Endpunkt angesehen werden zu können?
2. Welche Methoden werden zur Validierung von Surrogatendpunkten in der Literatur diskutiert?
3. Welche methodischen Vorgaben machen internationale Agenturen aus dem Bereich HTA oder Arzneimittelnutzenbewertung hinsichtlich des Einsatzes von Surrogatendpunkten?
4. Welchen Stellenwert haben Surrogatendpunkte in HTA-Berichten internationaler HTA-Agenturen und in Berichten, die in Deutschland produziert werden?

## **6.4 Methodik**

Entsprechend der Fragestellungen des Berichts, werden drei verschiedene methodische Herangehensweisen gewählt, die in den folgenden Abschnitten beschrieben werden.

### **6.4.1 Recherche in Literaturdatenbanken - Kriterien zur Beurteilung und zur Validierung von Surrogatendpunkten**

Um die erste und die zweite Frage dieses Berichts bzgl. der methodischen Diskussionen um die Anwendung von Surrogatendpunkten zu beantworten (siehe 6.3 „Forschungsfragen“), wird ein Review der einschlägigen methodischen Literatur durchgeführt.

#### **6.4.1.1 Informationsquellen und Recherchen**

Folgende Literaturdatenbanken werden im Rahmen der Literaturrecherche von DAHTA beim DIMDI in Absprache mit den Autoren durchsucht: MEDLINE, MEDLINE Alert, EMBASE, EMBASE Alert, DAHTA, Database of Abstracts of Reviews of Effectiveness (DARE), NHS Economic Evaluation Database (NHS-EED), Cochrane Central Trials Register (CCTR), Cochrane Database of Systematic Reviews (CDSR), INAHTA, Ethmed, DIMDI-Superbase.

Die verwendeten Suchbegriffe lauten:

SURROGATE END POINT; SURROGATE END POINTS; SURROGATE ENDPOINT; SURROGATE ENDPOINTS; ENDPOINT, SURROGATE; ENDPOINTS, SURROGATE; END POINT, SURROGATE; END POINTS,SURROGATE; BIOLOGICAL MARKER; BIOLOGICAL MARKERS; VALIDATION; STATISTICS; BIOMETRY; DECISION SUPPORT TECHNIQUES; ENDPOINT DETERMINATION; CAUSALITY

Die Recherche wird am 08.03.2007 durchgeführt. Der Suchzeitraum umfasst den 01.01.1990 bis 08.03.2007. Die detaillierte Suchstrategie ist dem Anhang zu entnehmen (8.1 „Dokumentation der Literaturrecherche“).

Zusätzliche Literaturstellen aus Referenzlisten der eingeschlossenen Arbeiten werden ebenfalls berücksichtigt, ohne die Referenzlisten jedoch systematisch zu durchsuchen.

#### **6.4.1.2 Selektion**

Die Selektion relevanter Publikationen wird von den zwei Autoren dieses Berichts unabhängig von einander vorgenommen. Zunächst werden die Titel und Abstracts der Ergebnisse der Literaturrecherche hinsichtlich der nachfolgend genannten Ein- und Ausschlusskriterien gescreent.

In einem zweiten Schritt werden die Volltexte der Arbeiten hinsichtlich ihrer tatsächlichen Relevanz überprüft, wobei hier darauf geachtet wird, Redundanzen zu vermeiden (d. h. Artikel, die lediglich andere Artikel in unsystematischer Form zusammenfassen bzw. zitieren, ohne jedoch zusätzlichen Erkenntnisgewinn zu bringen, blieben unberücksichtigt).

#### 6.4.1.3 Einschlusskriterien

- Sprache: Deutsch, Englisch oder Spanisch.
- Publikationsjahr  $\geq$  1990.
- Publikationen, die primär aus einer theoretischen und methodischen Perspektive das Thema „Surrogatendpunkte“ behandeln. Es sind Publikationen, die sich mit Definitionen, Kriterien und Methoden zur Validierung von Surrogatendpunkten befassen.

#### 6.4.1.4 Ausschlusskriterien

- In-vitro Studien, Tierstudien.
- Klinische Studien, Phase I bis II.
- Interventionsstudien mit ausschließlich Surrogatendpunkten, die keine Validierungsstudien sind.
- Publikationen, die nicht aus dem Gebiet Medizin/Pharmakologie (z. B. Modelle aus dem Bereich Psychologie) stammen.
- Publikationen, die ausschließlich Biomarker thematisieren, die lediglich als diagnostisches Merkmal für eine Erkrankung zu verstehen sind, deren Veränderung aber nicht als potentieller Endpunkt für ein Therapieregime angesehen wird.
- Publikationen, die ausschließlich biologische Marker zu Umweltexpositionen thematisieren.

#### 6.4.1.5 Synthese

Die Literatur wird in Form eines narrativen Reviews zusammengefasst. Dabei werden zunächst die gefundenen Artikel in folgende Gruppen geteilt: Artikel die aus theoretischer Perspektive das Konzept der Surrogatparameter erläutern und diskutieren (6.5.1.2.1 „Konzeptionelle Aspekte“) sowie Artikel, die methodische Ansätze zur Validierung von Surrogaten beschreiben und diskutieren (6.5.1.2. „Statistische Methoden zur Validierung von Surrogatendpunkten“).

Für die Synthese werden gefundene systematische Übersichten als Ausgangspunkt genommen.

### 6.4.2 Review der Methodenpapiere internationaler Agenturen aus dem Bereich HTA oder Arzneimittelnutzenbewertung

Für die Beantwortung der dritten Frage bzgl. des Umgangs von internationalen Agenturen mit dem Thema *Surrogatendpunkte* (siehe 6.3 „Forschungsfragen“) werden die Methodenpapiere dieser Institutionen untersucht. Es werden zwei Typen von Institutionen eingeschlossen, die wie folgt definiert werden:

- HTA-Agenturen: öffentlich finanzierte Institutionen, deren Aufgabe (u. a.) die Bewertung von medizinischen Technologien und die Erstellung von HTA-Berichten ist, mit dem Ziel die gesundheitspolitische Entscheidungsfindung im Allgemeinen zu unterstützen.
- Institutionen, die Arzneimittelnutzenbewertungen durchführen: öffentlich finanzierte Institutionen, deren Aufgabe die (vergleichende) Bewertung von Arzneimitteln im Rahmen formalisierter Prozesse zu Kostenübernahme-/Preissetzungsentscheidungen ist und die daher auch als Bestandteil der sogenannten fourth hurdle-Systeme verstanden werden<sup>47</sup>.

#### 6.4.2.1 Identifizierung relevanter Institutionen

##### *HTA-Agenturen*

Es werden alle Mitglieder des internationalen Netzwerks von HTA-Agenturen (International network of agencies for Health Technology Assessment (INAHTA)) durch dessen Webseite ([www.inahta.org/inahta\\_web/index.asp](http://www.inahta.org/inahta_web/index.asp); Zugriff: Juni 2007) identifiziert. Auf den Webseiten der jeweili-

gen INAHTA-Mitglieder wird zunächst nach Publikationen zu den jeweiligen Methoden für die Bewertung von Technologien gesucht. Es werden alle Methodenpapiere von INAHTA-Mitgliedern eingeschlossen die

- Durch den Internetauftritt der Agentur öffentlich zugänglich sind und
- In einer den Autoren dieses Berichts verständlichen Sprache (Deutsch, Englisch, Französisch, Portugiesisch, Spanisch) verfasst werden.

#### *Agenturen zur Arzneimittelnutzenbewertung*

Darüber hinaus werden die Methodenbeschreibungen von in die Arzneimittelnutzenbewertung involvierten Institutionen berücksichtigt, die in einem vorherigen Projekt („Methoden zur vergleichenden Bewertung pharmazeutischer Produkte“)<sup>93</sup> identifiziert wurden. Dabei werden folgende Einschlusskriterien formuliert:

- Institutionen aus Ländern mit Mitgliedschaft in der Europäischen Union (EU) oder in der Organisation für wirtschaftliche Zusammenarbeit und Entwicklung (OECD),
- Institutionen, die an der Arzneimittelregulierung beteiligt sind (einschließlich Kostenübernahme/Erstattungsfähigkeit (Leistungskatalog), Preisregulierung und Zuzahlungsregelungen).

#### **6.4.2.2 Dokumentenanalyse**

In den verfügbaren Methodenpapieren der HTA-Agenturen werden die Kapitel bzw. Textstellen analysiert, die sich mit den Outcomeparametern für die Bewertung der Technologien befassen. Darüber hinaus werden diese Dokumente mit Hilfe der Suchfunktionen der Textverarbeitungs- bzw. Textanzeigeprogramme auf folgende Schlüsselwörter durchsucht:

- „surrogat“,
- „biological marker“ bzw. „biologisch\* Marker“ bzw. „marcador biológico“,
- „endpoint“ bzw. „Endpunkt“ bzw. „medidas de resultado“,
- „outcome“.

Darüber hinaus werden die Webseiten der Organisationen zusätzlich anhand dieser Suchwörter mithilfe der Suchfunktionen der Webseiten durchforstet.

Die auf diese Weise gefundenen Textstellen zu der Thematik der Surrogatendpunkte werden extrahiert und zusammengefasst.

#### **6.4.2.3 Synthese**

Die Extraktion der wesentlichen Ausführungen aus den Methodenpapieren wird in einer zusammenfassenden Tabelle dokumentiert (siehe Anhang 8.3 „Dokumentation der Methodenpapiere aus HTA-Institutionen“). Die Ergebnisse werden narrativ dargestellt.

### **6.4.3 Analyse der Anwendung von Surrogatendpunkten in HTA-Berichten**

Für die Beantwortung der vierten Frage bzgl. der Anwendung von Surrogatendpunkten in HTA-Berichten (siehe 6.3 „Forschungsfragen“), werden die in abgeschlossenen sowie veröffentlichten internationalen und deutschen HTA-Berichten berücksichtigten Outcomeparameter analysiert.

#### **6.4.3.1 Quellen**

##### **Internationale Berichte**

Als HTA-Bericht werden alle Dokumente definiert, die in der internationalen HTA-Datenbank (*HTA database*) registriert sind und nicht als *ongoing projects* bzw. *projects* bezeichnet werden. Die HTA-Datenbank wird vom Centre for Reviews and Dissemination (CRD) in York in Zusammenarbeit mit der INAHTA und der schwedischen HTA-Agentur SBU bereitgestellt ([www.crd.york.ac.uk/crdweb](http://www.crd.york.ac.uk/crdweb)). Sie beinhaltet Publikationen von Institutionen, die sich mit der Bewertung medizinischer Technologien befassen, mit dem Ziel die aus den wissenschaftlichen Studien gewonnenen Erkenntnisse in gesund-

heitspolitischen Entscheidungsprozesse einzubringen (für eine Liste der in der HTA-Datenbank berücksichtigten Institutionen siehe Anhang 8.4 „HTA-Agenturen, deren Berichte in der HTA-Datenbank enthalten sind“).

#### **Deutsche Berichte**

Als deutsche HTA-Berichte werden alle Berichte betrachtet, die in der HTA-Datenbank der DAHTA beim DIMDI ([www.dimdi.de/dynamic/de/hta/db/index.htm](http://www.dimdi.de/dynamic/de/hta/db/index.htm)) registriert sind und nicht von internationalen Agenturen herausgegeben werden. Die Datenbank beinhaltet neben den Berichten der DAHTA beim DIMDI, auch Berichte des IQWiG, vom ehemaligen Bundesausschuss sowie von der Kassenärztlichen Bundesvereinigung und privaten Institutionen. Die Berichte des IQWiG werden direkt von der Internetseite dieses Instituts heruntergeladen ([www.iqwig.de](http://www.iqwig.de)).

#### **6.4.3.2 Einschlusskriterien für HTA-Berichte**

Es werden alle Berichte berücksichtigt, die folgende Kriterien erfüllen:

- Sprache: Deutsch, Englisch, Französisch, Portugiesisch oder Spanisch.
- Abgeschlossener HTA-Bericht.

#### **6.4.3.3 Ausschlusskriterien für HTA-Berichte**

Berichte werden ausgeschlossen, wenn mindestens eines der folgenden Kriterien erfüllt ist:

- Keine kostenfreie Zugänglichkeit der Berichte über das Internet bzw. Bibliotheken.
- Bericht beinhaltet keine Nutzenbewertung einer Technologie (z. B. Methodenberichte, nur ökonomische Betrachtungen, Bewertung von Leitlinien).
- Die Autoren des HTA-Berichts extrahieren keine Ergebnisse und haben auch nicht im Methodenteil bzw. in einer Frage dargelegt, welche Outcomeparameter extrahiert werden sollen.
- Die verwendeten bzw. extrahierten Endpunkte sind unklar.

#### **6.4.3.4 Stichprobenziehung**

##### **Internationale Berichte**

Es werden die Datenbankeinträge aller die Einschlusskriterien erfüllenden HTA-Berichte in eine Reference-Manager-Datei geladen und anschließend aus dieser eine repräsentative 5-prozentige Zufallsstichprobe gezogen. Dazu werden zunächst eine entsprechend der Anzahl an Gesamtberichten (Gesamtstichprobe) große Menge an Zufallszahlen in dem Statistikprogramm SAS<sup>®</sup> generiert. Diese Zahlen werden dann in der so ermittelten Reihenfolge der zuvor nach Autoren sortierten Gesamtstichprobe an Berichten zugematcht. Anschließend werden die Berichte nach der Zufallszahl sortiert und die ersten 5 % der Berichte hinsichtlich der repräsentativen Verteilung der als wesentlich erachteten Merkmale „Herkunftsland der Agentur“ und „Jahr der Publikation“ überprüft. Sofern die Verteilung repräsentativ für die Verteilung in der Gesamtstichprobe ist, wird diese 5 %-Stichprobe für die Betrachtung der verwendeten Endpunkte genutzt. Hätte sich gezeigt, dass die Verteilung dieser Merkmale in der 5 %-Stichprobe noch nicht repräsentativ gewesen wäre, wäre der Stichprobenumfang sukzessiv vergrößert und die Stichprobenerhebung entsprechend neu durchgeführt worden.

##### **Deutsche Berichte**

Für deutsche Berichte wird keine Stichprobe gezogen, sondern alle in der DAHTA-Datenbank bzw. im Internetauftritt des IQWiG beinhalteten Berichte berücksichtigt, vorausgesetzt keines der Ausschlusskriterien trifft zu (siehe 6.4.3.3 „Ausschlusskriterien für HTA-Berichte“).

#### **6.4.3.5 Definition verwendeter Endpunkte**

Es werden alle HTA-Berichte, die die Einschluss- und keines der Ausschlusskriterien erfüllten, daraufhin durchgesehen, welche Endpunkte für die Nutzenbewertung in der Methodik des Berichts festgelegt werden. Sofern diese in der Methodik nicht genannt oder nur allgemein bezeichnet werden (z. B. Morbidität), werden die berücksichtigten Endpunkte aus den Evidenztabelle der Berichte

extrahiert. Die inhaltliche Einteilung der Endpunkte zur Untersuchung der Wirksamkeit erfolgt nach folgendem Klassifikationsschema, das sich an die Definitionen der Biomarkers Definition Group<sup>10</sup> anlehnt:

**Klinisch relevanter Endpunkt:** „A characteristic or variable that reflects how a patient feels, functions, or survives“<sup>10</sup>.

Klinisch relevante, patientennahe Endpunkte sind für den vorliegenden Bericht entweder „harte klinische Endpunkte“ oder „Patient Reported Outcomes“ (PRO).

**Harter (klinischer) Endpunkt:** Jeder patientenrelevante Endpunkt, der von Dritten objektivierbar ist, wie z. B. Mortalität oder morbiditätsbedingte Ereignisse (Myokardinfarkt etc.)

**PRO:** Jeder patientenrelevante Endpunkt, der vom Patienten selbst berichtet wird und somit nicht von Dritten zuverlässig objektivierbar ist. Dies können Messungen der Lebensqualität sowie andere subjektive Angaben, wie z. B. Symptome, Leistungsfähigkeit, Zufriedenheit etc. sein. Diese Art von Endpunkten wird mittels verschiedener Instrumente (z. B. Fragebögen zu Symptomkomplexen, Skalen zur Schmerzintensität) ermittelt bzw. erfragt. Die Instrumente können validiert oder nicht validiert, mehr oder weniger komplex sein.

Bei manchen Erkrankungen werden Scores verwendet, die eine Kombination aus objektivierbaren klinischen Zielgrößen und Patientenangaben sind, um die Verbesserungen bzw. Verschlechterungen der Krankheit darzustellen und in standardisierter Form messen zu können. Diese Parameter sind somit auch klinisch relevante Endpunkte und werden entsprechend kodiert.

**Surrogatendpunkt:** Objektivierbare Messgrößen, die – basierend auf epidemiologischen, therapeutischen oder pathophysiologischen Überlegungen – in einer Kausalkette als Vorstufen und somit als Prädiktoren für den eigentlich interessierenden klinischen Endpunkt angesehen werden. Darüber hinaus werden in diesem Bericht auch die für die Bewertung diagnostischer Maßnahmen häufig verwendeten Größen wie Sensitivität, Spezifität etc. als Surrogatendpunkte bezeichnet (in Anlehnung an das Modell zur Bewertung diagnostischer Technologien von Fryback und Thornbury<sup>34</sup>).

Weiterhin wird überprüft, ob in den Berichten neben den o. g. Endpunkten, die überwiegend für die Bewertung der Wirksamkeit einer Technologie verwendet werden, auch Risiken, potentielle Komplikationen oder unerwünschte Wirkungen der Technologie explizit betrachtet bzw. ob solche Ergebnisparameter extrahiert werden. Diese Art von Parametern wird im vorliegenden Bericht als **Sicherheitsendpunkt** betrachtet. Dabei handelt es sich immer um patienten- bzw. klinisch relevante Endpunkte.

#### 6.4.3.6 Synthese

Zunächst wird eine Übersicht über die verwendeten Ergebnisparameter erstellt. Anschließend werden die als „Surrogatendpunkte“ klassifizierten Parameter detaillierter betrachtet und qualitativ diskutiert. Dabei werden diagnostische und therapeutische Studien separat ausgewertet, da zu erwarten ist, dass im Rahmen der diagnostischen Studien im Wesentlichen die oben genannten diagnostischen Testkenngrößen verwendet werden. Ohne eine stratifizierte Darstellung hätte dies möglicherweise zu einem verzerrten Ergebnis bezüglich der Häufigkeit der Anwendung von Surrogatendpunkten (im Sinn von Vorstufen in der pathophysiologischen Kausalkette) in HTA-Berichten geführt.

## 6.5 Ergebnisse

### 6.5.1 Recherche in Literaturdatenbanken - Kriterien zur Beurteilung und Validierung von Surrogatendpunkten

#### 6.5.1.1 Quantitative Ergebnisse

Insgesamt werden mit der Recherchestrategie gemäß 8.1 „Dokumentation der Literaturrecherche“ N = 1.109 Treffer erzielt. Von diesen werden in einem ersten Selektionsschritt n = 2 Dubletten entfernt und n = 1.007 Literaturstellen ausgeschlossen, da sie für das Ziel des Berichts nach Durchsicht von Titel und Abstract als nicht relevant eingestuft werden. Insgesamt sind n = 100 Literaturstellen ausgewählt und für eine weitere Überprüfung im Volltext bestellt. Eine weitere Studie wird aus den Literaturverzeichnissen der Publikationen ausgewählt und ebenfalls im Volltext bestellt<sup>30</sup>.

Von den 101 im ersten Screeningschritt ausgewählten Publikationen werden 76 Studien im zweiten ausgeschlossen. Ein Teil der aus dem Review ausgeschlossenen Publikationen wird als Hintergrundmaterial für die Einleitung und die Diskussion verwendet. Eine Übersicht über die ausgeschlossenen Literaturstellen ist 8.2 „Aus der Methodenübersicht ausgeschlossene Literatur“ zu entnehmen. Somit werden insgesamt 25 Studien in dem Review über Kriterien zur Beurteilung von Surrogatendpunkten und Methoden zur Validierung von Surrogatendpunkten aus der elektronischen Literaturrecherche eingeschlossen.

Von diesen sind fünf der Kategorie „Konzeptionelle Aspekte“ (6.5.1.2.1 „Konzeptionelle Aspekte“) und 20 der Kategorie „Validierungsaspekte“ (6.5.1.2.2 „Statistische Methoden zur Validierung von Surrogatendpunkten“) zugeordnet. Als Grundlage für die Darstellung werden Übersichtsarbeiten verwendet, und bei Bedarf durch zusätzliche Informationen aus den primären Veröffentlichungen ergänzt. Drei Reviews dienen als Grundlage für dieses Kapitel<sup>55, 57, 90</sup>, das neben Studien, die durch die elektronische Literaturrecherche identifiziert werden (siehe Tabelle 3), auch andere Artikel berücksichtigt und zusammenfasst, die nicht durch die Recherche identifiziert werden.

**Tabelle 3: Eingeschlossene Artikel aus der elektronischen Literaturrecherche**

<b>Konzeptionelle Aspekte</b>	
<i>Übersichten</i>	
Lonn E. The use of surrogate endpoints in clinical trials: focus on clinical trials in cardiovascular diseases. <i>Pharmacoepidemiology and Drug Safety</i> 2001; 10(6): 497-508.	
<i>Primärveröffentlichungen</i>	
Biomarkers Definition Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. <i>Clinical Pharmacology and Therapeutics</i> 2001; 69(3):89-95	L Ja
Boissel JP, Collet JP, Moleur P, Haugh M. Surrogate endpoints: a basis for a rational approach. <i>European Journal of Clinical Pharmacology</i> 1992; 43(3):235-244.	Ja
Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. <i>Journal of the American Medical Association</i> 1999; 282(8):771-778.	Ja
Fleming TR, DeMets DL. Surrogate end points in clinical trials: are we being misled? <i>Annals Internal Medicine</i> 1996; 125(7):605-613.	Ja
<b>Validierungsaspekte</b>	
<i>Übersichten</i>	
Molenberghs G, Burzykowski T, Alonso A, Buyse M. A perspective on surrogate endpoints in controlled clinical trials. <i>Statistical Methods in Medical Research</i> 2004; 13(3): 177-206.	
Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. <i>Statistics in Medicine</i> 2006; 25(2): 183-203.	
<i>Primär Veröffentlichungen</i>	
Alonso A, Molenberghs G, Burzykowski T, Renard D, Geys H, Shkedy Z et al. Prentice's approach and the meta-analytic paradigm: a reflection on the role of statistics in the evaluation of surrogate endpoints. <i>Biometrics</i> 2004; 60(3):724-728.	M W Ja
Alonso A, Molenberghs G, Geys H, Buyse M, Vangeneugden T. A unifying approach for surrogate marker validation based on Prentice's criteria. <i>Statistics in medicine</i> 2006; 25(2):205-221	
Baker SG. A simple meta-analytic approach for using a binary surrogate endpoint to predict the effect of intervention on true endpoint. <i>Biostatistics (Oxford, England)</i> 2006; 7(1):58-70.	
Burzykowski T, Molenberghs G, Buyse M. The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer. <i>Journal of the Royal Statistical Society Series A-Statistisc in Society</i> 2004; 167(P1): 103-124.	Ja Ja
Burzykowski T, Molenberghs G. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. <i>Journal of the Royal Statistical Society Series C-Applied Statistics</i> 2001; 50(P4): 405-422.	Ja Ja
Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. <i>Pharmaceutical statistics</i> 2006; 5(3):173-186.	
Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. <i>Biometrics</i> 1998; 54(3):1014-1029.	Ja Ja
<i>Fortsetzung nächste Seite</i>	

Buyse M, Molenberghs G, Burzykowski T. Erratum: Criteria for the validation of surrogate endpoints in randomized experiments (Biometrics (September 1998) 54 (1014-1029)). Biometrics 2000; 56(1):324.	Ja	Ja
Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. Statistical validation of surrogate endpoints: Problems and proposals. Drug Information Journal 2000; 34(2):447-454		Ja
Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. Statistics in Medicine 1992; 11(2):167-178	Ja	Ja
Hughes MD. Evaluating surrogate endpoints. Controlled clinical trials 2002; 23(6):703-707.		Ja
Li Z, Meredith MP, Hoseyni MS. A method to assess the proportion of treatment effect explained by a surrogate endpoint. Statistics in medicine 2001; 20(21):3175-3188.		Ja
Lin DY, Fleming TR, de Gruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. Statistics in Medicine 1997; 16: 1515-1527.	Ja	Ja
Molenberghs G, Buyse M, Geys H, Renard D, Burzykowski T, Alonso A. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. Controlled clinical trials 2002; 23(6):607-625.	Ja	Ja
Molenberghs G, Geys H, Buyse M. Evaluation of surrogate endpoints in randomized experiments with mixed discrete and continuous outcomes. Statistics in medicine 2001; 20(20):3023-3038	Ja	Ja
Renard D, Geys H, Molenberghs G, Burzykowski T, Buyse M. Validation of surrogate endpoints in multiple randomized clinical trials with discrete outcomes. Biometrical Journal 2002; 44(N8):921-935.	Ja	Ja
Taylor JM, Wang Y. Surrogate markers and joint models for longitudinal and survival data. Controlled clinical trials 2002; 23(6):626-634.		Ja
Wang Y, Taylor JM. A measure of the proportion of treatment effect explained by a surrogate marker. Biometrics 2002; 58(4):803-812.		Ja
Xu J, Zeger SL. The evaluation of multiple surrogate endpoints. Biometrics 2001; 57(1):81-87.		Ja

L = In der Übersicht von Lonn berücksichtigt. M = In der Übersicht von Molenberghs et al. berücksichtigt. W = In der Übersicht von Weir und Walley berücksichtigt.

Darüber hinaus sind in der Übersicht neun weitere Arbeiten zitiert, die nicht durch die elektronische Literaturrecherche identifiziert werden, sondern aus den Referenzlisten von eingeschlossenen Artikeln stammen.

## 6.5.1.2 Qualitative Ergebnisse

### 6.5.1.2.1 Konzeptionelle Aspekte

#### Definitionen

Wie bereits in der Einleitung dieses Berichts erwähnt, weist auch Lonn in seiner Übersichtsarbeit darauf hin, dass es seit den 1980er Jahren mehrere Ansätze gibt, den Begriff „Surrogatendpunkt“ zu definieren<sup>55</sup>. Eine Übersicht über die am häufigsten zitierten Definitionen ist Tabelle 4 zu entnehmen.

Dabei sprechen die ersten Definitionsansätze noch wenig konkret von einem „irgendwie gearteten Zusammenhang“ oder einer „ausreichend guten Korrelation“ mit dem klinisch relevanten Endpunkt<sup>29, 41</sup> bzw. von einer „angemessenen Alternative“ zu diesem Endpunkt<sup>12</sup>. Was genau unter „angemessen“ zu verstehen ist, verdeutlicht Boissel anhand verschiedener Kriterien (siehe nächster Abschnitt). Während er die Begriffe „surrogat end point“ und „intermediate end point“ in ihrer Bedeutung gleichsetzt, grenzt die FDA diese klar voneinander ab<sup>84</sup>. Demnach ist ein Surrogatendpunkt lediglich ein Laborparameter oder ein körperliches Merkmal, das stellvertretend für einen patientenrelevanten Endpunkt in klinischen Studien verwendet wird. Ein intermediate end point hingegen ist bereits selbst ein wahrer und für den Patienten bedeutsamer Endpunkt, wie z. B. die Häufigkeit von Angina Pectoris-Anfällen. Dieser stellt zwar nur einen Zwischenschritt auf dem Weg zum ultimativen Endpunkt z. B. Mortalität) dar, er wird aber aufgrund seiner Relevanz für die Patienten beispielsweise von der FDA als Zulassungsgrundlage akzeptiert. Der Begriff Surrogatendpunkt wird dabei sowohl in der Definition der FDA als auch in der der Biomarkers Definitions Group<sup>10</sup> enger gefasst als in denen der Tabelle 4. So wird von einem Surrogatendpunkt verlangt, dass er in der Lage ist, den klinischen Nutzen einer Therapie zu prognostizieren, was über die Anforderung einer einfachen Korrelation deutlich hinausgeht. Die Biomarkers Definitions Group verlangt für die Prüfung dieser Voraussetzung Evidenz aus epidemiologischen, therapeutischen, pathophysiologischen oder anderen wissenschaftlichen Studien. Die Kriterien sowie Modelle zur qualitativen und quantitativen Prüfung eines Surrogatendpunkts werden in den nächsten beiden Abschnitten ausführlicher behandelt.

Tabelle 4: Surrogatendpunkte – Definitionen

Autor (Jahr)	Definition
Ellenberg und Hamilton (1989) <sup>29</sup>	[...] investigators use surrogate endpoints when the endpoint of interest is too difficult and/or expensive to measure routinely and when they can define some other, more readily measurable endpoint which is sufficiently well correlated with the first to justify its use as a substitute.
Hillis und Seigel (1989) <sup>41</sup>	An observed variable that relates in some way to the variable of primary interest which we cannot conveniently observe directly.
Boissel et al. (1992) <sup>12</sup>	An intermediate endpoint can be defined as a response variable which is statistically correlated with the clinical endpoint. It can be used as an appropriate alternative to a clinical endpoint in a clinical trial.
Food and Drug Administration (1992) <sup>31</sup>	A surrogate end point, or 'marker', is a laboratory measurement or physical sign that is used in clinical trials as a substitute for a clinically meaningful end point that is a direct measure of how a patient feels, functions, or survives and is expected to predict the effect of the therapy.
Biomarker Definitions Group (2001) <sup>10</sup>	A biomarker is a characteristic that is objectively measured and evaluated as an indicator of normal biological processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention. A surrogate endpoint is a biomarker that is intended to substitute for a clinical endpoint. A surrogate endpoint is expected to predict clinical benefit (or harm or lack of benefit or harm) based on epidemiologic, therapeutic, pathophysiologic, or other scientific evidence.

Quelle: Lonn<sup>55</sup> (Erweiterte Darstellung mit eigenen Ergänzungen)

**Anforderungen an Surrogatendpunkte (Kriterien und qualitative Modelle)**

Sowohl Boissel et al. als auch Fleming weisen darauf hin, dass häufig eine einfache statistische Korrelation zwischen Surrogat und klinischem Endpunkt bereits als Beweis für einen kausalen Zusammenhang zwischen diesen beiden Endpunkten und somit als Legitimierung für den Austausch dieser beiden Endpunkte missverstanden wird<sup>12, 30</sup>. Dieses Problem ist in Abbildung 1 dargestellt. In diesem von Boissel et al. aufgeführten Beispiel werden sowohl der klinisch relevante Endpunkt E als auch der Surrogatendpunkt S durch den Faktor U verursacht. Obwohl die beiden Wirkmechanismen voneinander unabhängig sind, gibt es eine statistische Korrelation zwischen den Endpunkten S und E. Obgleich S somit ein Risikofaktor von E ist, würde eine medizinische Intervention, die ausschließlich den Weg von U zu S modifiziert keinen Einfluss auf das Auftreten von E haben. Also ist S in diesem Kontext kein geeigneter Surrogatparameter um die Wirksamkeit von T zu untersuchen.

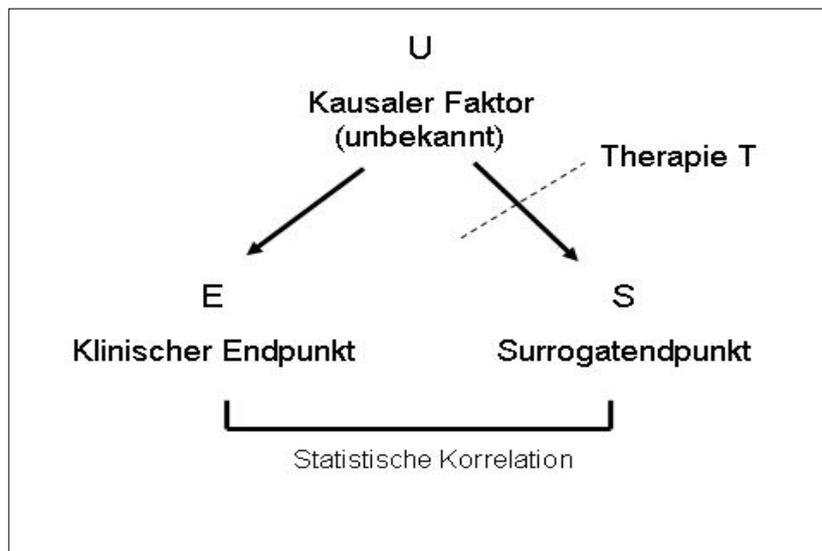


Abbildung 1

Quelle: Boissel JP, Collet JP, Moleur P, Haugh M<sup>12</sup>. Surrogate endpoints: a basis for a rational approach. European journal of clinical pharmacology 1992; 43(3): 235-244.  
E = Klinisch relevanter Endpunkt. S = Surrogat. T = Therapie. U = Ursache.

An diesem Beispiel wird verdeutlicht, dass es von großer Bedeutung ist, die Pathophysiologie der zugrunde liegenden Erkrankung genau zu kennen, um beurteilen zu können, ob ein Surrogatendpunkt als Endpunkt geeignet ist. Ohne diese Kenntnis ist es nicht möglich abzuschätzen, ob der Surrogatendpunkt die Effekte einer Intervention vollständig abbilden kann.

Ausgehend von diesen Überlegungen beschreibt Boissel ein Modell mit dessen Hilfe die Eignung von Surrogatendpunkten überprüft werden kann. Demnach müssen grundsätzlich drei Bedingungen für einen potentiellen Surrogatendpunkt erfüllt sein, damit er als valider Endpunkt angesehen werden kann:

1. Der Surrogatendpunkt (S) ist im Vergleich zum klinisch relevanten Endpunkt (E) nicht nur früher zu beobachten und einfacher zu erfassen, sondern er tritt auch häufiger auf.
2. Epidemiologische Studien zeigen sowohl einen qualitativen als auch einen quantitativen Zusammenhang zwischen Surrogat und klinischem Endpunkt.
3. Aufgrund von Ergebnissen aus randomisierten kontrollierten Studien (RCT) ist abschätzbar, mit welchem Ausmaß an klinischem Nutzen  $\delta_t(E)$  ein therapiebedingter Effekt auf dem Surrogatendpunkt  $\delta_t(S)$  einhergeht.

Dabei geht Boissel bzgl. des dritten Punkts nicht darauf ein, ob sich dieser abschätzbare Therapieeffekt nur auf die Effekte der konkret zu untersuchenden Intervention bezieht oder ob er damit auch Substanzgruppen- oder gar gruppenübergreifende Effekte meint.

Abhängig davon,

- a) ob der klinisch relevante Endpunkt auch dann auftritt, wenn der Surrogatendpunkt nicht auftritt und
- b) ob die zu untersuchende Therapie einen Effekt auf die Beziehung zwischen Surrogat und klinischem Endpunkt hat

zeigt Boissel vier verschiedene Möglichkeiten auf, wie die Größen  $\delta_t(S)$  und  $\delta_t(E)$  geschätzt werden können<sup>12</sup>.

*Möglichkeit I:* E tritt nur im Zusammenhang mit S auf und T wirkt *nicht* auf die Beziehung S-E. Wenn der qualitative und der quantitative Zusammenhang zwischen S und E aus epidemiologischen Studien bekannt und somit Bedingung 2 (siehe oben) erfüllt ist, dann ist Bedingung 3 (die Abschätzbarkeit des klinisch relevanten therapeutischen Nutzens  $\delta_t(E)$ ) automatisch auch erfüllt. Boissel räumt aber ein, dass der Nachweis, dass T nicht auf die Beziehung S-E wirkt, nur über RCT erfolgen kann.

*Möglichkeit II:* E tritt nur im Zusammenhang mit S auf und T wirkt auch auf die Beziehung S-E. In diesem Fall sind ebenfalls Daten aus RCT nötig, die den Effekt von T auf S und von T auf E zeigen (S gegeben).

*Möglichkeit III:* E tritt auch auf, wenn S nicht auftritt und T *nicht* auf die Beziehung S-E wirkt. Als Beispiel nennt Boissel hier das Auftreten von Myokardinfarkten bei Patienten mit normalen Cholesterin- oder Blutdruckwerten. Seiner Ansicht nach reicht es in diesem Fall theoretisch aus, wenn man die Wahrscheinlichkeit des Auftretens von E, gegeben S, und die Wahrscheinlichkeit des Auftretens von E ohne gleichzeitiges Auftreten von S aus epidemiologischen Studien kennt. Dennoch sind auch hier RCT notwendig, um zu zeigen, dass T nicht auf die Beziehung S-E wirkt.

*Möglichkeit IV:* E tritt auch auf, wenn S nicht auftritt und T auf die Beziehung S-E wirkt. In diesem Fall sind umfassende Erkenntnisse aus epidemiologischen Studien und RCT notwendig, um den Zusammenhang zwischen S und E zum einen und den Einfluss von T auf S, sowie T auf E in Gegenwart oder Abwesenheit von S abschätzen zu können.

Aus diesen grundsätzlichen Überlegungen leitet Boissel verschiedene konkrete Konstellationen für das Zusammenspiel zwischen T, S und E ab, für die individuell geprüft werden muss, ob die oben formulierten Bedingungen erfüllt sind (Tabelle 5).

**Tabelle 5: Konstellationen für das Zusammenspiel zwischen Therapie, Surrogat und Endpunkt**

<b>Situation (Annahme: 1. Bedingung grundsätzlich erfüllt)</b>		<b>2. und 3. Bedingung</b>
1	S und E werden durch denselben Faktor U verursacht. S ist aber nicht Bestandteil der Kausalkette U-E. Dennoch erscheint E immer zusammen mit S.  $P(E/S) = 1$ $P(E/nS) = 0$	Grundsätzlich erfüllt, wenn T auf den Pfad vor der Abteilung oder aber auf beide Pfade nach der Abteilung wirkt.
2	S und E werden durch denselben Faktor U verursacht und sind statistisch korreliert. S ist aber nicht Bestandteil der Kausalkette U-E und E tritt auch auf, wenn S nicht auftritt.  $P(E/S) \neq 1$ $P(E/nS) \neq 0$	Grundsätzlich erfüllt, wenn T auf den kausalen Pfad vor der Abteilung wirkt
3	S ist zwingender Bestandteil der Kausalkette U-E.  $P(E/S) = 1$ $P(E/nS) = 0$	Grundsätzlich erfüllt.
4	S ist zwingender Bestandteil der Kausalkette U-E. Dennoch besteht die Möglichkeit, dass der Pfad von S zu E unterbrochen wird und E somit nicht auftritt, obwohl S auftritt.  $P(E/S) \neq 1$ $P(E/nS) = 0$	Grundsätzlich erfüllt.
5	E kann durch mehrere Faktoren $U_i$ und jeweils über separate Kausalpfade verursacht werden. S ist nur Bestandteil eines Pfades.  $P(E/nS) \neq 0$	Erfüllt, wenn die anderen Surrogatendpunkte nicht dominieren oder T auf alle Pfade wirkt.
6	E kann über mehrere separate Kausalpfade ausgehend von einem gemeinsamen Faktor U verursacht werden. S ist nur Bestandteil eines Pfades.  $P(E/nS) \neq 0$	Erfüllt, wenn die anderen Surrogatendpunkte nicht dominieren oder T entweder auf alle Pfade oder aber auf U wirkt.
7	Das Auftreten von E ist nur durch das Zusammenwirken mehrerer Kausalpfade möglich, die alle von einem gemeinsamen Faktor U ausgehen. Keiner der einzelnen Pfade kann E allein verursachen, es ist immer das Zusammenwirken mehrerer Pfade notwendig, wobei es nicht erforderlich ist, dass alle Pfade beteiligt sind.	Grundsätzlich erfüllt, wenn T entweder auf alle Pfade oder aber auf U wirkt. Ansonsten abhängig davon, welchen Anteil S an $\Sigma$ hat.
8	Das Auftreten von E ist nur durch das Zusammenwirken mehrerer Kausalpfade möglich, die jeweils von verschiedenen Faktoren $U_i$ ausgehen. Keiner der einzelnen Pfade kann E allein verursachen. Es ist immer das Zusammenwirken mehrerer Pfade notwendig, wobei es nicht erforderlich ist, dass alle Pfade beteiligt sind.	Grundsätzlich erfüllt, wenn T entweder auf alle Pfade oder auf den Pfad nach $\Sigma$ wirkt. Ansonsten abhängig davon, welchen Anteil S an $\Sigma$ hat.

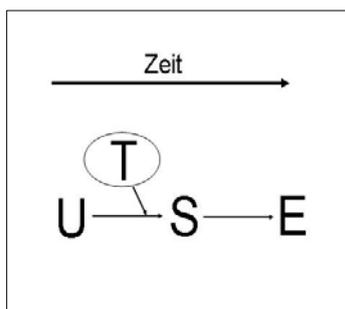
E = Klinisch relevanter Endpunkt. P = Wahrscheinlichkeit.  $P(E/nS)$  = Wahrscheinlichkeit von E, wenn S nicht gegeben ist.  $P(E/S)$  = Wahrscheinlichkeit von E gegeben S. S = Surrogat. T = Therapie. U = Ursache.

Quelle: Boissel JP, Collet JP, Moleur P, Haugh M<sup>12</sup>. Surrogate endpoints: a basis for a rational approach. European journal of clinical pharmacology 1992; 43(3): 235-244.

Generell stehen S und E hierbei entweder über parallel verlaufende (Modell 1 und 2) oder über aufeinander folgende Pfade (Modell 3 bis 8) zueinander in Beziehung, wobei der Einfluss des ursächlichen Faktors U bzw. der Faktoren  $U_i$  auf den klinischen Endpunkt E über einen einzigen kausalen Pfad (Modell 1 bis 3) oder über mehrere Pfade (Modell 4 bis 8) verlaufen kann.

Es zeigt sich deutlich, dass es für die Evaluierung eines Surrogatendpunkts nicht ausreicht, allein pathophysiologische Aspekte der zugrunde liegenden Erkrankung zu berücksichtigen. Vielmehr ist es zudem relevant zu wissen, in welcher Weise die zu untersuchende Therapie auf die Beziehung S-E wirkt. So zeigen die Modelle 3 und 4 beispielsweise, dass, die Frage, ob S ein geeigneter Surrogatendpunkt für E sein kann, davon abhängt, ob T vor oder nach dem Auftreten von S wirkt. Während sich S im ersten Fall als geeigneter Surrogatendpunkt erweisen könnte, würde es in dem Fall, dass T erst nach dem Auftreten von S direkt auf E wirkt, zu einer Unterschätzung der Wirksamkeit von T kommen, wenn S als Endpunkt verwendet wird. Modell 1 zeigt darüber hinaus, dass es Konstellationen geben kann, in denen S zwar kein Bestandteil des Kausalpfades U-E ist, S aber dennoch ein geeigneter Surrogatendpunkt sein kann, wenn T frühzeitig in den Pfad eingreift.

Fleming und de Mets stellen ähnliche Überlegungen an<sup>30</sup>. Ihrer Ansicht nach ist die Konstellation mit dem größten Potential für einen Surrogatendpunkt jene, in der es nur einen Kausalpfad von U zu E gibt, in dem S ein zwingender Zwischenschritt ist und in welcher T ausschließlich auf S und nicht auf die Beziehung zwischen S und E wirkt (siehe Abbildung 2). Dies entspricht der im Modell 3 von Boissel erläuterten Konstellation (siehe Tabelle 5) und wird auch von Bucher<sup>14</sup> als die Situation beschrieben, in der einem Surrogatendpunkt die höchste Reliabilität zugesprochen werden kann.



**Abbildung 2**

Quelle: Fleming TR, de Mets DL. 30 Surrogate end points in clinical trials: are we being misled? *Annals of Internal Medicine* 1996; 125(7):605-613.

E = Endpunkt. S = Surrogatendpunkt. T = Therapie. U = Ursache.

Um beurteilen zu können, ob diese Situation gegeben ist, fordert Bucher eine ähnliche Evidenzlage, wie sie auch Boissel beschreibt. Das heißt, zunächst müssen epidemiologische oder pathophysiologische Studien eine starke, unabhängige und konsistente Beziehung zwischen S und E gezeigt haben. Dass dies allerdings nur ein notwendiges aber noch kein ausreichendes Kriterium für einen verlässlichen Surrogatendpunkt ist, demonstriert Bucher am Beispiel der *Knochendichte* als Surrogatparameter für das Auftreten von Frakturen bei Patienten mit einer Osteoporose. Obwohl sich eine verringerte Knochendichte in Beobachtungsstudien als unabhängiger Risikofaktor für das Auftreten von Hüftfrakturen zeigt, kann in einer Studie mit postmenopausalen Osteoporosepatientinnen unter einer Natriumfluorid-Therapie zwar eine signifikante Erhöhung der Knochendichte im Vergleich zu Placebo erreicht werden, gleichzeitig steigt aber auch die Frakturinzidenz in der Verum- stärker als in der Placebogruppe<sup>73</sup>. Somit kann angenommen werden, dass es entweder keinen kausalen Zusammenhang zwischen diesen beiden Endpunkten gibt oder dass Natriumfluorid möglicherweise einen eigenständigen Risikofaktor für das Auftreten von Frakturen darstellt.

Aus diesem Grund fordert Bucher<sup>14</sup> ebenso wie Boissel<sup>12</sup>, dass es entweder Ergebnisse aus RCT mit Wirkstoffen aus *anderen* oder *derselben* Wirkstoffgruppe gibt, die zeigen, dass eine Verbesserung des Surrogatendpunkts zu einer konsistenten Verbesserung des klinisch relevanten Endpunkts führt. Letztgenannte werden häufig auch als Validierungsstudien bezeichnet. Wobei die Biomarker Definition

Group in ihrem Konsensuspapier von 2001 anmerkt, dass der Begriff valide fälschlicherweise suggeriert, dass ein Surrogatendpunkt generell und therapieübergreifend einen bestimmten klinischen Endpunkt substituieren kann<sup>10</sup>. Die Gruppe schlägt deshalb vor, nicht von Validierung sondern vielmehr von Evaluierung zu sprechen. Wie problematisch es dabei sein kann, auf die Ergebnisse aus Studien mit Substanzen aus anderen Wirkstoffgruppen zu vertrauen, schildert Bucher am Beispiel der Fibrate<sup>14</sup>. Sowohl Fibrate als auch Statine führen nachweislich zu einer Verringerung des Surrogatendpunkts *LDL* und des klinisch relevanten Endpunkts *Myokardinfarkt*. Nur für Statine kann ein positiver Einfluss auf die Gesamtmortalität bei Patienten mit KHK gezeigt werden<sup>15</sup>. Darüber hinaus kann es selbst bei Substanzen aus der gleichen Wirkstoffgruppe zu Abweichungen im Wirkmechanismus kommen. So wird beispielsweise diskutiert, dass manche Statine weitere, über die Hemmung der HMG-Co-Reduktase hinausgehende Wirkungen haben, die atherothrombotische Prozesse zusätzlich beeinflussen<sup>74</sup>.

Fleming und de Mets nennen in ihrer Übersichtsarbeit vier wesentliche Gründe, warum Surrogatendpunkte in der Vergangenheit häufig zu Trugschlüssen geführt haben, die sich bildlich darstellen lassen<sup>30</sup> (siehe Abbildung 3).

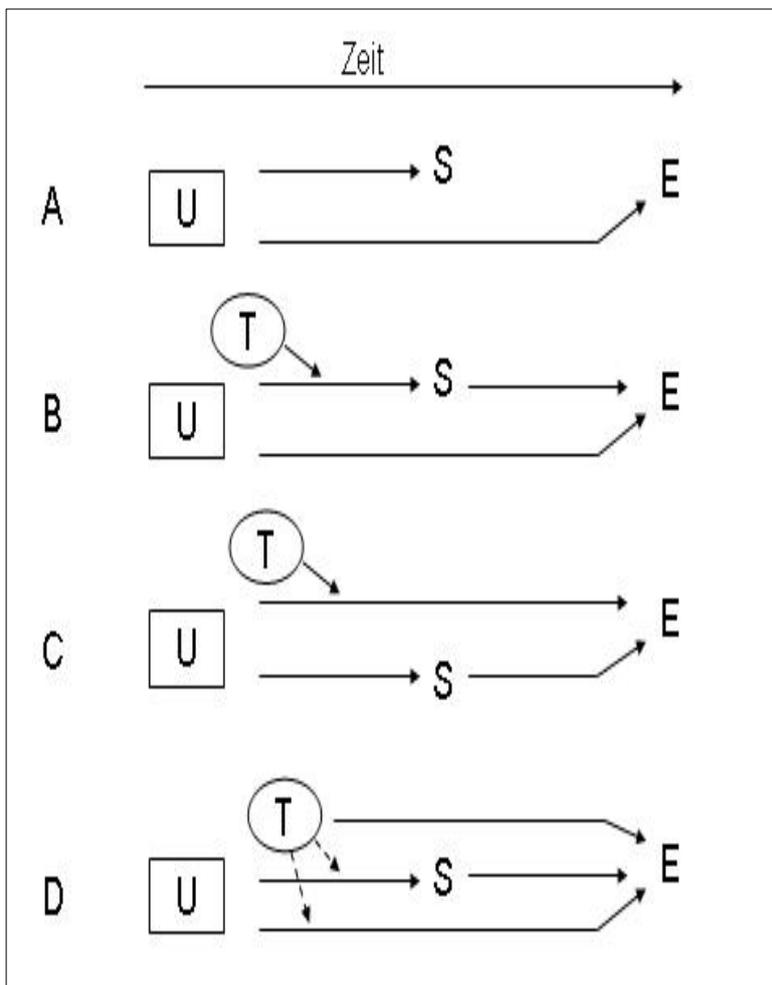


Abbildung 3: Verschiedene Konstellationen der Beziehung zwischen Surrogatparametern und Endpunkten

Quelle: Fleming TR, de Mets DL<sup>30</sup> Surrogate end points in clinical trials: are we being misled? *Annals of Internal Medicine* 1996; 125(7):605-613.

E = Endpunkt. S = Surrogat. T = Therapie. U = Ursache.

Dabei entspricht **Modell A** in Abbildung 3 bereits dem von Boissel (Abbildung 1) skizzierten Problem. Obwohl es möglicherweise eine Korrelation zwischen S und E gibt und beide auch durch denselben Faktor U verursacht werden, liegt S nicht auf dem Kausalpfad von U zu E, so dass eine therapeutische Beeinflussung von S keinerlei Auswirkung auf E hat. Als Beispiel nennen Fleming und de Mets die Verwendung von Prostata-Biopsie-Ergebnissen als Surrogatendpunkt für die klinischen Endpunkte *Symptome* und *Überlebenszeit* in Finasterid-Chemopräventionsstudien. Aufgrund der Pathophysiologie der Erkrankung und der Wirkweise von Finasterid ist es durchaus möglich, dass durch Finasterid zwar die Anzahl positiver Biopsieergebnisse verringert wird, dies aber keine Beeinflussung der klinisch relevanten Endpunkte mit sich zieht. Darüber hinaus halten die Autoren es nicht für ausgeschlossen, dass die Verwendung von *Knochendichte* als Surrogatparameter für den klinisch relevanten Endpunkt *Frakturen* bei Osteoporosepatienten diesem Problem zuzuordnen ist (siehe oben).

In einem weiteren Modell (**Modell B**) zeigt Fleming (Abbildung 3), dass allein die Tatsache, dass ein Surrogatparameter S auf einem Kausalpfad von U zu E liegt und S durch T beeinflussbar ist, noch kein hinreichendes Kriterium für die Eignung dieses Surrogatparameters ist. Vielmehr kann es weitere Kausalpfade von U zu E geben, an denen S nicht beteiligt ist und die nicht von T modifiziert werden, so dass es auch trotz einer therapiebedingten Verringerung von S zu einem Anstieg von E kommen kann. Dies könnte in der Vergangenheit laut Fleming v. a. bei Therapiestudien im Bereich der Herz-Kreislaufkrankungen zum Problem geworden sein. So führte die Verbesserung des Surrogatparameters *körperliche Belastbarkeit* bei Patienten mit Herzinsuffizienz zu einer frühzeitigen Zulassung des Vasodilatators Flosequinan in den USA. Die FDA versah die Zulassung allerdings mit der Auflage, dass eine Studie mit dem klinisch relevanten Endpunkt *Mortalität* angeschlossen wird (PRO-FILE-Studie<sup>66</sup>). Als aus dieser Studie deutlich wurde, dass es unter Flosequinan keinesfalls zu einer Verringerung, sondern vielmehr zu einer Erhöhung der Mortalität kam, wurde das Präparat sehr schnell vom Markt genommen<sup>30</sup>.

Darüber hinaus ist auch eine Situation denkbar, in der S Bestandteil eines Kausalpfades von U zu E ist, der überhaupt nicht von T beeinflusst wird (**Modell C** in Abbildung 3), so dass die Verwendung von S als Endpunkt einer Interventionsstudie zu einer Unterschätzung der Effektivität von T führen würde. Als Beispiel nennen die Autoren den Einsatz von Interferon- $\gamma$  bei Kindern mit chronisch granulomatöser Erkrankung. Es ist bekannt, dass das Immunsystem dieser Patienten aufgrund der Erkrankung stark beeinträchtigt ist, so dass Makrophagen beispielsweise nicht mehr imstande sind, Mikroorganismen abzutöten, was häufig lebensbedrohliche Infektionen zur Folge hat. Interferon- $\gamma$  ist in der Lage, Hyperoxide zu produzieren und über diesen Weg Infektionserreger zu vernichten. Aus diesem Grund wird ein RCT zur Untersuchung der Wirksamkeit von Interferon- $\gamma$  bei Kindern mit chronisch granulomatöser Erkrankung durchgeführt. Obwohl ursprünglich vorgesehen ist, nur die Surrogatendpunkte *Produktion von Hyperoxiden* und *Verringerung von Infektionserregern* als Endpunkte einzusetzen, wird noch vor Studienbeginn entschieden, doch den als klinisch relevant angesehenen Endpunkt *gravierende Infektionen* zu verwenden. Es zeigt sich, dass Interferon- $\gamma$  die Anzahl gravierender Infektionen im Vergleich zu Placebo deutlich verringert, hinsichtlich der Surrogatendpunkte allerdings keine Verbesserung erreicht werden kann. Wäre die Studie nach dem ursprünglichen Studienprotokoll durchgeführt worden, wäre das Potential dieses Wirkstoffs somit nicht erkannt worden<sup>30</sup>.

Den mit Abstand ungünstigsten Fall beschreibt **Modell D** (Abbildung 3), in dem T selbst einen Risikofaktor für E darstellt (bzw. einen schädlichen Effekt auf den klinisch relevanten Endpunkt hat). Wenn S in diesem Fall als Bestandteil eines Kausalpfades von U zu E durch T modifiziert wird, könnte es trotz einer Verringerung von S zu einer Erhöhung von E kommen. Würde S in diesem Fall als Endpunkt akzeptiert werden, hätte dies einen verheerenden Trugschluss zur Folge. Auch hier bringen Fleming und de Mets Beispiele aus dem Bereich der Herz-Kreislaufkrankungen ein. So könnte beispielsweise die bei Patienten mit ventrikulären Arrhythmien zu beobachtende Erhöhung der Mortalität unter den Antiarrhythmika Ecanide oder Flecainide im Vergleich zu Placebo (CAST-Studie, siehe auch 6.2 „Wissenschaftlicher Hintergrund“ dieses Berichts) auf dieses Problem zurückzuführen sein.

Fasst man die in diesem Abschnitt formulierten Überlegungen zusammen, so sind an einen Surrogatendpunkt folgende Kriterien zu stellen, damit dieser als valider Endpunkt eingestuft werden kann:

1. **Biologische Plausibilität:** Es liegt Evidenz aus tierexperimentellen oder epidemiologischen Studien vor, die auf einen kausalen Zusammenhang zwischen dem Surrogatparameter und dem klinischen Endpunkt hinweisen. Der Surrogatparameter liegt demnach auf dem pathophysiologischen Kausalpfad zum klinisch relevanten Endpunkt.
2. **Ausmaß der Beziehung zwischen Surrogatendpunkt und klinischem Endpunkt:** Epidemiologische Studien zeigen wiederholt, dass durch eine Veränderung des Surrogatendpunkts eine qualitativ und quantitativ gleichgerichtete Veränderung des klinisch relevanten Endpunkts erreicht werden kann. Das Surrogat sollte den Effekt der Intervention vollständig abbilden.
3. **Wirkungsnachweis im Rahmen von RCT:** Es liegen Ergebnisse aus RCT vor, die zeigen, dass durch eine therapiebedingte Veränderung des Surrogatendpunkts eine gleichgerichtete Veränderung des klinisch relevanten Endpunkts erreicht werden kann. Der Surrogatendpunkt muss den Effekt der Intervention vollständig abbilden. Dabei ist zu beachten, dass es nicht nur bei pharmakologisch unterschiedlichen Wirkstoffen, sondern auch bei Wirkstoffen der gleichen Wirkstoffgruppe zu abweichenden Wirkmechanismen kommen kann und daher die Übertragbarkeit der Studienergebnisse von einer Substanz auf eine andere zusätzlich geprüft werden muss.

Darüber hinaus sind unabhängig von der Validität auch rein pragmatische Anforderungen an einen Surrogatendpunkt zu stellen. Zu diesen gehören, dass dieser früher und häufiger auftritt als der klinisch relevante Endpunkt und zudem einfach, reliabel und reproduzierbar zu erheben ist<sup>12, 55</sup>, da er dann auch im klinischen Alltag für Therapieerfolgskontrollen Verwendung finden kann<sup>10</sup>.

#### 6.5.1.2.2 Statistische Methoden zur Validierung von Surrogatendpunkten

Seit Ende der 1980er Jahre sind eine Reihe von Verfahren zur Validierung von Surrogatendpunkten entwickelt und erprobt worden. Ziel von Validierungsstudien ist es zu ermitteln, in welchem Ausmaß eine therapiebedingte Veränderung eines Surrogatendpunkts den therapeutischen Effekt auf den klinisch relevanten Endpunkt prognostizieren kann<sup>57, 90</sup>.

Die Parameter, die in diesem Zusammenhang von größtem Interesse sind, lauten:

$\alpha$ , der Effekt einer therapeutischen Intervention T auf den Surrogatendpunkt S,

$\beta$ , der Effekt von T auf den klinisch relevanten Endpunkt E,

$\gamma$ , der Effekt von S auf E,

$\beta_s$ , der Effekt von T auf E, adjustiert für S

$\gamma_T$ , der Effekt von S auf E, adjustiert für T.

Die Beziehung zwischen T, S und E kann durch ein lineares Modell mit  $i = 1, \dots, n$  Beobachtungen folgendermaßen dargestellt werden<sup>21</sup>:

$$(1) \quad S_i | T_i = \mu_S + \alpha T_i + \varepsilon_{Si}$$

$$(2) \quad E_i | T_i = \mu_E + \beta T_i + \varepsilon_{Ei}$$

$$(3) \quad E_i | S_i = \mu + \gamma S_i + \varepsilon_i$$

Dabei sind  $\mu_S$ ,  $\mu_E$  und  $\mu$  die jeweiligen y-Achsenabschnitte;  $\alpha$ ,  $\beta$ ,  $\gamma$  die Steigungen der Regressionsgraden und somit die interessierenden Größen;  $\varepsilon_S$ ,  $\varepsilon_{Ei}$ ,  $\varepsilon_i$  normalverteilten Fehlerterme. Diese grundsätzlichen Überlegungen stellen die Ausgangssituation für die meisten der im Folgenden beschriebenen Validierungsmethoden dar.

#### **Prentice-Kriterien**

Die ersten Überlegungen zur statistischen Validierung von Surrogatendpunkten werden 1989 von Prentice angestellt<sup>70</sup>. Er fordert, dass der Nullhypothesen-Test für einen Surrogatendpunkt (Therapie hat keinen Einfluss auf den Surrogatendpunkt) auch ein valider Test für die entsprechende Nullhypothese bezogen auf den klinisch relevanten Endpunkt sein muss (Therapie hat keinen Einfluss auf den klinisch relevanten Endpunkt). *[A surrogate endpoint is response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint.]*

Aus dieser Bedingung leitet Prentice folgende Kriterien ab, die für einen Surrogatendpunkt erfüllt sein müssen, um ihn als valide bezeichnen zu können:

1. T hat einen Einfluss auf S ( $\alpha \neq 0$ ).
2. T hat einen Einfluss auf E ( $\beta \neq 0$ ).
3. Es besteht ein Zusammenhang zwischen S und dem E. Dieser Zusammenhang ist unabhängig von T ( $\gamma \neq 0$ ).
4. Der Effekt von T auf E kann vollständig über den Effekt von T auf S erklärt werden (Statistisch ausgedrückt: T hat keinen Effekt auf E, gegeben S).

Dabei spielt es bei Prentice keine Rolle, wie groß die jeweiligen Effekte ( $\alpha$ ,  $\beta$ ,  $\gamma$ ) sind. Es ist lediglich erforderlich, dass sie von 0 verschieden sind. Obwohl Prentice mit seinen Überlegungen den Grundstein für die Entwicklung von Validierungsmethoden für Surrogatendpunkte gelegt hat, werden seine Kriterien – insbesondere das letzte Kriterium – vielfach kritisiert, da sie streng genommen nur für binäre Endpunkte Gültigkeit haben und insbesondere das letzte Kriterium vielen Wissenschaftlern zu strikt bzw. unrealistisch erscheint<sup>21</sup>.

### **Proportion of treatment effect explained (PTE)**

Auch Freedman und Kollegen weisen in ihrer Arbeit von 1992 auf die konzeptionelle Schwäche des umstrittenen letzten Prentice-Kriteriums hin<sup>33</sup>. Dieses macht es erforderlich, dass der Signifikanztest für den Therapieeffekt auf den klinischen Endpunkt nach der Adjustierung für den Surrogatendpunkt nicht mehr signifikant ist. Diese Nichtsignifikanz kann aber keinesfalls als Beweis dafür angesehen werden, dass der Therapieeffekt auf den klinisch relevanten Endpunkt vollständig über den Effekt auf den Surrogatendpunkt erklärt wird. Aus diesem Grund modifizieren Freedman et al. die Überlegungen von Prentice dahingehend, dass sie von einem strikten Hypothesentest und der Forderung der hundertprozentigen Erklärungskraft eines Surrogatparameters abrücken und stattdessen eine Größe einführen, über die abgeschätzt werden kann, zu welchem Anteil der Effekt von T auf E durch den Effekt von T auf S erklärt werden kann: die sogenannte PTE. Gemäß der zugrunde liegenden Idee ist ein Surrogatendpunkt umso geeigneter, desto höher die ermittelte PTE ist. Der Zusammenhang zwischen T und E, adjustiert für S ( $\beta_s$ ) kann dabei durch ein multiples lineares Regressionsmodell dargestellt werden:

$$(4) \quad E_i | T_i S_i = \mu_E + \beta_s T_i + \gamma_T S_i + \varepsilon_{Ei}$$

Dabei ist  $\mu_E$  der y-Achsenabschnitt;  $\beta_s$  und  $\gamma_T$  sind die interessierenden adjustierten Effektschätzer;  $\varepsilon_{Ei}$  ist ein normalverteilter Fehlerterm.

Die PTE berechnet sich dann durch die Gleichung:

$$(5) \quad PTE = 1 - \beta_s / \beta$$

wobei  $\beta$  und  $\beta_s$  die Schätzer für den Effekt von T auf E ohne und mit Adjustierung für S sind, die über die Modelle (2) und (4) ermittelt werden können.

Theoretisch liegen die Werte für die PTE zwischen 0 und 1, wobei ein perfekter Surrogatendpunkt eine PTE von 1 annehmen würde und eine PTE von 0 einen ungeeigneten Surrogatendpunkt entspricht. Tatsächlich zeigt sich allerdings, dass die PTE in der Praxis häufig kleiner als 0 oder größer als 1 ist, wenn beispielsweise  $\beta_s$  und  $\beta$  unterschiedliche Vorzeichen haben<sup>59</sup>. Dies ist möglich, wenn sich die Richtung des Zusammenhangs zwischen T und E nach Adjustierung für S ändert. Darüber hinaus beschreiben Molenberghs und Kollegen mehrere Konstellationen, in denen die PTE zu Trugschlüssen führen kann<sup>59</sup>. Zudem wird die PTE instabil, wenn der geschätzte Therapieeffekt sehr klein oder Interaktionen zwischen  $T_i$  und  $S_i$  vorliegen und das Regressionsmodell (4) somit nicht ausreichend ist<sup>19, 20</sup>. Freedman et al. weisen selbst darauf hin, dass das Konfidenzintervall für die PTE häufig sehr breit ist und große Fallzahlen für eine aussagekräftige Schätzung benötigt werden. Darüber hinaus wird kritisiert, dass die Entscheidung, ab welchem PTE-Schwellenwert von einem validen Surrogatendpunkt gesprochen werden kann (z. B. untere Grenze des 95 % PTE-Konfidenzintervalls größer 0,5), dem subjektiven Empfinden eines Wissenschaftlers unterliegt<sup>19, 20</sup>.

Während Freedman und Kollegen diese Methode für binäre Endpunkte entwickeln<sup>33</sup>, erweitern Lin et al.<sup>54</sup> und Li et al.<sup>53</sup> die Berechnungen für Überlebenszeitdaten. Wang und Taylor<sup>89</sup> modifizieren die Berechnungen darüber hinaus, indem sie den tatsächlich beobachteten und einen theoretischen Therapieeffekt, der beobachtbar wäre, wenn die Verteilung des Surrogatparameters in der Kontroll- der

Verteilung in der Verumgruppe entsprechen würde, zueinander ins Verhältnis setzen. Trotz dieser Weiterentwicklungen können die genannten grundsätzlichen Schwierigkeiten dieser Methode allerdings auch hier nicht gelöst werden<sup>90</sup>.

### **Trial level und Individual level validity**

#### **Relativer Effekt (RE) und adjusted association ( $\gamma_T$ )**

Buyse und Kollegen erweitern die angestellten Überlegungen durch die Unterteilung des Begriffs Validität in eine Trial level- und eine Individual level-Validität. Während auf dem Trial level über mehrere Studien oder Studienzentren hinweg auf Basis von Durchschnittswerten quantifiziert wird, in welchem Ausmaß der Therapieeffekt auf den klinisch relevanten Endpunkt über den Therapieeffekt auf den Surrogatendpunkt erklärt werden kann, wird auf dem Individual level, also der Ebene des einzelnen Patienten, untersucht, wie stark der Zusammenhang zwischen Surrogat- und klinisch relevantem Endpunkt unabhängig von einer therapeutischen Intervention ist<sup>19, 20, 21</sup>.

Sie führen zwei weitere Größen zur Abschätzung der Eignung von Surrogatparametern ein: den RE und die  $\gamma_T$ <sup>19, 20</sup>.

Der RE ist definiert als der Quotient aus dem Effekt von T auf E ( $\beta$ ) und dem Effekt von T auf S ( $\alpha$ ):

$$(6) \quad RE = \beta / \alpha$$

wobei  $\alpha$  und  $\beta$  über die Modelle (1) und (2) ermittelt werden können. Sofern der Effekt von T auf E dem von T auf S entspricht, wird  $RE = 1$ . In diesem Fall wird von einem *perfect surrogate at the trial level* gesprochen<sup>19, 20</sup>. Buyse und Molenberghs weisen darauf hin, dass der RE wesentlich präzisere Angaben liefert als die PTE<sup>19, 20</sup>. Dies liegt daran, dass beim RE anders als bei der PTE  $\alpha$ , also der Effekt von T auf S, im Nenner steht und dieser präziser geschätzt werden kann als  $\beta$ , der Effekt von T auf E. Somit ist das Konfidenzintervall des RE in der Regel schmaler und informativer als das der PTE. Der RE kann nützlich sein, um den Effekt von T auf E zu prognostizieren, wenn der Effekt von T auf S empirisch ermittelt wird<sup>12</sup>. Sofern der RE präzise geschätzt wird, kann dabei von einer hinreichend präzisen Schätzung des interessierenden therapeutischen Effekts ausgegangen werden<sup>21</sup>. Um eine derart präzise Schätzung zu erhalten, sind große Fallzahlen und entsprechend große RCT erforderlich. Ein Nachteil der Schätzung der RE auf Basis einzelner Studien besteht darin, dass von einer generellen Gültigkeit und Übertragbarkeit der Ergebnisse auf andere Studien ausgegangen wird, ohne dass diese Annahme auf diesem Weg belegt werden kann. Einen Ausweg stellt die Kombination mehrerer Studien in Form von Metaanalysen dar<sup>21, 24, 59</sup>. Dieser Aspekt wird im nächsten Abschnitt näher erläutert (siehe unten).

In der Praxis wird der RE eines Surrogatendpunkts häufig auf Grundlage von bekannten therapeutischen Interventionen geschätzt, um sich in der Folge bei der Prüfung neuer, vergleichbarer Therapien (z. B. Wirkstoffe aus der gleichen Wirkstoffgruppe), nur noch auf Studien mit dem entsprechenden Surrogatendpunkt beschränken zu können und die Durchführung von RCT mit klinisch relevanten Endpunkten zu umgehen. Wie bereits im vorherigen Abschnitt dieses Berichts dargestellt, kann die ungeprüfte Annahme derartiger Gruppeneffekte allerdings zu gravierenden Trugschlüssen führen<sup>19, 20</sup>.

Die zweite von Buyse und Kollegen vorgeschlagene Größe ist die bereits aus der Modellgleichung (4) bekannte  $\gamma_T$ , der Effekt von S auf E nach Adjustierung für T, bestimmt auf der Ebene eines Individuums. Von einem geeigneten Surrogatparameter wird erwartet, dass er unabhängig von einer therapeutischen Intervention in hohem Maß mit dem klinisch relevanten Endpunkt assoziiert ist, da nur so ein Hinweis auf biologische und pathophysiologische Zusammenhänge gegeben ist. Je größer der Wert für  $\gamma_T$  ist, desto stärker ist die Assoziation zwischen S und E und desto höher ist die Validität von S in Bezug auf E einzuschätzen. Buyse und Molenberghs bezeichnen Surrogatendpunkte mit  $\gamma_T = \infty$  (binäre Endpunkte) bzw. mit  $\gamma_T = 1$  (kontinuierliche Endpunkte) als perfect surrogates on the individual level<sup>19, 20</sup>.

Buyse et al. zeigen, dass die von Freedman et al. vorgeschlagene PTE nichts anderes ist, als der Quotient aus der  $\gamma_T$  und dem RE:

$$(7) \quad PTE = \gamma_T / RE$$

Somit ist die PTE als eine zusammengesetzte Größe zu verstehen, die biologische bzw. pathophysiologische Informationen über den Zusammenhang zwischen S und E auf der individuellen Ebene ( $\gamma_T$ ) mit Informationen über den Effekt von T auf S sowie E auf der Studienebene (RE) kombi-

nirt. Aufgrund der Verschiedenheit dieser beiden Parameter raten Buyse und Molenberghs allerdings dazu,  $\gamma_T$  und RE getrennt und nicht als Quotienten zu betrachten, da die Interpretation dieser Verhältniszahl schwierig ist<sup>19, 20, 59</sup>.

### Metaanalysen

Wie bereits erwähnt, weisen Buyse und Kollegen darauf hin, dass es für eine zuverlässige Schätzung des RE mit kleinem Konfidenzintervall notwendig ist, über große Fallzahlen zu verfügen, und idealerweise mehrere Studien kombiniert auswerten zu können<sup>21, 59</sup>. Metaanalysen bieten ihrer Meinung nach eine gute Möglichkeit, die Annahme zu prüfen, ob das Ausmaß der therapiebedingten Veränderung des klinisch relevanten Endpunkts proportional zur therapiebedingten Veränderung des Surrogatendpunkts verläuft. Während Daniels und Hughes zunächst damit beginnen, ausschließlich die Trial level-Validität zu modellieren<sup>24</sup>, erweitern Buyse, Molenberghs und Kollegen diese Überlegungen durch komplexere Modelle<sup>21, 59</sup>. Gemäß ihrer Überlegungen kann die Beziehung zwischen T und den Endpunkten S und E im Fall von  $i = 1, \dots, N$  Studien mit  $j = 1, \dots, n$  Personen folgendermaßen modelliert werden:

$$(8) \quad S_{ij} | T_{ij} = \mu_{Si} + \alpha_i T_{ij} + \varepsilon_{Sij}$$

$$(9) \quad E_{ij} | T_{ij} = \mu_{Ei} + \beta_i T_{ij} + \varepsilon_{Eii}$$

Dabei beschreibt der Suffix  $i$  den Trial und der Suffix  $j$  den Individual level. In der Regel wird bei normalverteilten Endpunkten ein hierarchisches Modell für die Parameterschätzung verwendet, in dem die y-Achsenabschnitte  $\mu_{Si}$  und  $\mu_{Ei}$  ebenso wie die Gradensteigungen  $\alpha_i$  und  $\beta_i$  jeweils in eine fixe ( $\mu_S$ ,  $\mu_E$ ,  $\alpha$ ,  $\beta$ ) und eine zufällige Komponente ( $m_{Si}$ ,  $m_{Ei}$ ,  $a_i$ ,  $b_i$ ) aufgeteilt werden:

$$(10) \quad S_{ij} | T_{ij} = (\mu_S + m_{Si}) + (\alpha + a_i) T_{ij} + \varepsilon_{Sij}$$

$$(11) \quad E_{ij} | T_{ij} = (\mu_E + m_{Ei}) + (\beta + b_i) T_{ij} + \varepsilon_{Eii}$$

Der Vorteil dieser Modelle ist, dass Fehler bei der Parameterschätzung berücksichtigt werden. Ein Nachteil ist allerdings, dass die Modelle komplex und mit hohem Rechenaufwand verbunden sind<sup>21, 59</sup>. Alternativ kann ein zweistufiger Ansatz verfolgt werden, indem zunächst jede Studie einzeln analysiert wird. Anschließend wird über die studienspezifischen Effekte ( $\alpha_i$  und  $\beta_i$ ) eine lineare Regression durchgeführt, wobei die Gradensteigung dem über alle Studien ermittelten RE ( $RE = \beta / \alpha$ ) entspricht. Das Bestimmtheitsmaß dieses Modells ( $R^2_{\text{trial}}$ ) ist dabei ein Maß für die Stärke des Zusammenhangs zwischen  $\alpha$  und  $\beta$ . Ein  $R^2_{\text{trial}}$  von eins würde bedeuten, dass der Effekt von T auf E fehlerfrei aus dem beobachteten Effekt von T auf S abgeschätzt werden kann, wobei ein Surrogatendpunkt als Trial level valide bezeichnet wird, wenn er ausreichend nahe bei eins liegt<sup>21, 59</sup>. Ebenso kann die *adjusted association* im Rahmen von Metaanalysen ermittelt werden, indem die Modellgleichung (4) stratifiziert für die einzelnen Studien gerechnet wird. Das Ausmaß der Beziehung zwischen den beiden Endpunkten nach Adjustierung für T wird durch das Bestimmtheitsmaß  $R^2_{\text{indiv}}$  angegeben. Ein Surrogatendpunkt wird als Individual level valide bezeichnet, wenn das  $R^2_{\text{indiv}}$  ausreichend nahe bei eins liegt. Ein perfekter Surrogatendpunkt ist sowohl Trial level valide als auch individual level valide<sup>21, 59</sup>.

Der Vorteil des Konzepts der Trial level- und Individual level-Validität besteht darin, dass es – anders als die Konzepte zuvor – nicht erfordert, dass der therapeutische Effekt auf einen der Endpunkte statistisch signifikant und darüber hinaus der Effekt auf den klinisch relevanten Endpunkt prognostizierbar ist<sup>90</sup>. Ein Nachteil ist hingegen, dass die Entscheidung, wie stark  $R^2_{\text{indiv}}$  und  $R^2_{\text{trial}}$  von eins abweichen dürfen, um einen Surrogatendpunkt noch als valide bezeichnen zu können, einer subjektiven Einschätzung bedarf<sup>21</sup>.

Während die Modelle von Buyse und Kollegen nur im Fall zweier normalverteilter Endpunkte Gültigkeit haben<sup>21</sup>, entwickeln Molenberghs et al. diese für die Situationen weiter, indem einer der beiden Endpunkte metrisch und der andere binär oder beide ordinal skaliert sind<sup>58</sup>. Renard und Kollegen ergänzen diese Arbeiten um Modelle mit zwei binären Endpunkten<sup>72</sup>. Handelt es sich hingegen bei beiden Endpunkten um Überlebenszeiten, so schlagen Burzykowski und Kollegen<sup>18</sup> eine bivariate Überlebenszeitanalyse vor, im Rahmen derer die Assoziation zwischen den beiden Überlebenszeiten auf Grundlage von Copula-Modellen modelliert wird. Die Individual level-Validität des Surrogatendpunkts wird dabei über Kendall's  $\tau$ , die Transformation der Assoziation zwischen den beiden Überlebenszeiten, angegeben. Die Trial level-Validität lässt sich hingegen aus der Korrelationsstärke zwischen  $\alpha_i$  und  $\beta_i$  ableiten<sup>18</sup>. Darüber hinaus entwickeln diese Autoren Copula-Modelle für den Fall binärer oder ordinaler Surrogatendpunkte sowie Überlebenszeit als klinisch relevantem Endpunkt<sup>17</sup>. Alonso et al.

machen zudem darauf aufmerksam, dass komplexere Modelle auch dann notwendig werden, wenn der Surrogatendpunkt und/oder der klinisch relevante Endpunkt längsschnittlich und wiederholt gemessen werden, da hier die Annahme einer über die Zeit konstanten Kovarianzstruktur der Fehlerterme nicht haltbar ist<sup>2</sup>. So ist es zum Beispiel möglich, dass sich die einzelnen Studien hinsichtlich der Anzahl an Messzeitpunkten und der Messfrequenz unterscheiden. Alonso und Kollegen machen deutlich, dass insbesondere das  $R^2_{\text{indiv}}$  bei wiederholt gemessenen Endpunkten nicht mehr anwendbar ist. Sie schlagen stattdessen die Verwendung des sogenannten *Variance reduction factor* ( $VRF_{\text{indiv}}$ ) vor, der auf der individuellen Ebene angibt, zu welchem Anteil die Gesamtvariabilität des wiederholt gemessenen klinisch relevanten Endpunkts durch die Adjustierung für den Therapieeffekt und die Adjustierung für die Variabilität des wiederholt gemessenen Surrogatendpunkts erklärt werden kann. Der Wertebereich des  $VRF_{\text{indiv}}$  reicht von 0 bis 1, wobei der  $VRF_{\text{indiv}}$  nur dann den Wert 0 annehmen kann, wenn die Fehlerterme für S und E innerhalb jeder betrachteten Studie voneinander unabhängig sind. Demgegenüber kann der  $VRF_{\text{indiv}}$  nur dann den Wert 1 erreichen, wenn es innerhalb jeder der Studien eine Abhängigkeit zwischen diesen Fehlertermen gibt.  $VRF_{\text{indiv}}$  entspricht dem  $R^2_{\text{indiv}}$ , wenn die Endpunkte nur einfach und nicht wiederholt gemessen werden.

Alonso und Kollegen erkennen darüber hinaus den Bedarf nach einheitlichen Ansätzen, in den Fällen, in denen weder der Surrogatendpunkt noch der klinisch relevante Endpunkt normalverteilt ist. Sie schlagen vor, statt der Bestimmtheitsmaße  $R^2_{\text{indiv}}$  und  $R^2_{\text{trial}}$  den Likelihood reduction factor (LRF) als Maß für die Individual-level-Validität eines Surrogatendpunkts zu verwenden<sup>7</sup>. Formuliert man die Modellgleichungen (2) und (4) als generalisiertes lineares Modell, so berechnet sich der LRF folgendermaßen:

$$(12) \quad \text{LRF} = 1 - 1 / N \sum_i \exp(-G_i^2 / n_i)$$

Dabei steht  $N$  für die Anzahl einbezogener Studien und  $n_i$  für die Anzahl an Personen in der  $i$ ten Studie.  $G_i^2$  ist die Likelihood-ratio-Teststatistik für den Vergleich der Modelle (2) und (4) in der  $i$ ten Studie. Im Fall normalverteilter Endpunkte entspricht der LRF dem  $R^2_{\text{indiv}}$ . Der LRF kann so skaliert werden, dass der Wertebereich zwischen 0 und 1 ist, wobei 1 einer hohen Validität entsprechen würde. Nachteilig ist, dass der LRF gleichen Einschränkungen unterliegt wie die PTE (siehe oben).

Hughes<sup>45</sup> weist auf die Bedeutsamkeit der Metaanalyseansätze im Kontext der Validierung von Surrogatendpunkten hin. Durch die Kombination von Informationen aus verschiedenen Studien eröffnen sie für neue Studien die Möglichkeit, den Therapieeffekt auf den klinisch relevanten Endpunkt abzuschätzen, wenn im Rahmen der Studie der Therapieeffekt auf den Surrogatendpunkt ermittelt wird<sup>45</sup>. Dennoch sind mit diesen Ansätzen auch einige Nachteile verbunden. So weist er auf Gails Kritik hin, dass Modelle, wie die von Buyse und Molenberghs<sup>19, 20, 21</sup> zu einer Überschätzung der Prädiktoren führen, da ihre Varianzen, insbesondere dann, wenn nur wenige Studien mit beiden Endpunkten vorliegen, unpräzise geschätzt werden. Darüber hinaus wird bei einer neuen Studie die Varianz des Schätzers für den Therapieeffekt auf den klinisch relevanten Endpunkt selbst bei einem  $R^2_{\text{trial}}$  von 1 (das in der Praxis nie erreicht wird) und einer fehlerfreien Messung des Therapieeffekts auf den Surrogatendpunkt nie vollständig zu eliminieren sein. Gail rät daher von Metaanalysen ab und schlägt stattdessen Bootstrap-Analysen vor<sup>35</sup>. Darüber hinaus stellt sich - wie bei allen Metaanalysen - die methodische Frage, welche Studien überhaupt vergleichbar sind und gepoolt werden können. Dazu gehört auch, dass die Verteilungen der Endpunkte in den Studien bekannt sein müssen und diese sich häufig von Studie zu Studie unterscheiden. Darüber hinaus besteht die Schwierigkeit, dass die Modelle häufig nicht invariant hinsichtlich der Kontrasteinstellung der Variablen sind (A versus B oder B versus A)<sup>57</sup>.

### **Surrogate threshold effect (STE)**

Burzykowski und Buyse greifen die Kritik von Gail<sup>35</sup> hinsichtlich der Ineffizienz bei der Effektschätzung im Kontext von Metaanalysen auf und stellen ein neues Validierungskonzept vor, das des  $STE$ <sup>16</sup>. Sie räumen ein, dass das ursprüngliche Konzept von Buyse et al. voraussetzt, dass bei der Bestimmung des  $R^2_{\text{trial}}$  (als Maß für die durch den Effekt von T auf S erreichte relative Verringerung der Variabilität des Prädiktors für den Effekt von T auf E) alle den Modellen (10) und (11) zugrunde liegenden fixen und zufälligen Parametern bekannt sind. In der Praxis müssen diese allerdings geschätzt werden. Aufgrund dieser Schätzung ist die tatsächlich erreichte Verringerung der Variabilität grundsätzlich geringer als durch  $R^2_{\text{trial}}$  angezeigt. Der STE ist definiert als der minimale therapeutische Effekt auf den Surrogatendpunkt, der notwendig ist, um einen von Null verschiedenen Schätzer für den therapeutischen Effekt auf den klinisch relevanten Endpunkt zu bekommen. Klinisch ausgedrückt ist der

STE der minimale therapeutische Effekt auf den Surrogatendpunkt, der notwendig ist, um überhaupt einen therapeutischen Effekt auf den klinisch relevanten Endpunkt nachweisen zu können. Burzykowski und Buyse erklären, dass der *STE* geeignet ist, um die Effizienzverluste bei der Schätzung des Effekts von T auf E im Rahmen von Metaanalysen zu quantifizieren<sup>16</sup>. Sie machen aber auch darauf aufmerksam, dass die Methode noch einer Weiterentwicklung bedarf<sup>16</sup>.

### **Multiple surrogates**

Xu und Zeger machen in ihrem Aufsatz von 2001 darauf aufmerksam, dass es häufig nicht darum geht einen einzelnen, sondern vielmehr eine Gruppe von Parametern hinsichtlich ihrer Eignung als Surrogatendpunkte zu überprüfen<sup>91</sup>. Für den Fall, dass die Surrogatendpunkte längsschnittlich gemessene Parameter sind und der klinisch relevante Endpunkt die Überlebenszeit darstellt, schlagen sie ein *latent variable model* vor. Um abschätzen zu können, welchen Wert die Verwendung von mehr als einem Surrogatparameter hat, wird zunächst der relative Präzisionszuwachs für das Konfidenzintervall des kumulativen Hazards ermittelt. Darüber hinaus wird ermittelt, welcher Anteil des Effekts von T auf E nicht durch die Surrogatendpunkte erklärt werden kann. Das Ziel der Methode besteht daher nicht darin, die Validität eines einzelnen Surrogatendpunkts zu überprüfen, sondern vielmehr darin, die bestmögliche Kombination von Surrogaten zu finden.

### **Longitudinale Surrogatendpunkte und Time to event-Endpunkte**

Wang und Taylor erklären, dass eines der wesentlichen Probleme der PTE (siehe oben) darin besteht, dass zu ihrer Ermittlung zwei verschiedene Modelle gerechnet werden, für die nicht sichergestellt werden kann, dass sie beide gleichzeitig die Daten fitten<sup>89</sup>. Sie schlagen stattdessen sogenannte *joint models* vor und demonstrieren ihre Überlegungen am Beispiel eines metrischen, longitudinal und wiederholt gemessenen Surrogatendpunkts und Überlebenszeit als klinisch relevantem Endpunkt. Der longitudinale Part des Modells ist dabei folgendermaßen formuliert:

$$(13) \quad S_{ij} = U_{ij} + e_{ij} \quad \text{mit } U_{ij} = a_i + b_i t_{ij} + \beta T_i$$

$a_i$  und  $b_i$  sind normalverteilte Zufallseffekte und  $t$  der Messzeitpunkt. Die *latent variable*  $U$  entspricht dem um den Messfehler reduzierten Wert von  $S$ . Das *failure time model* ist dabei folgendermaßen definiert:

$$(13) \quad \lambda(t, U, T) = \lambda_0(t) \exp(\gamma U + \omega T)$$

Dabei steht  $\omega$  für den Effekt von T auf E (hier Überlebenszeit) nach Adjustierung für S (hier longitudinal gemessener Surrogatendpunkt). Ersetzt man  $U$  durch die oben formulierte Modellgleichung und nimmt außerdem an, dass keine Messfehler und Zufallseffekte auftreten, so reduziert sich (12) folgendermaßen:

$$(14) \quad \lambda(t, U, T) = \lambda_{01}(t) \exp((\gamma\beta + \omega)T)$$

Somit ist der nicht adjustierte Therapieeffekt  $\omega^*$  gegeben durch  $\omega^* = (\gamma\beta + \omega)$ . Gemäß der der Formel zur Berechnung der PTE (siehe oben) ergibt sich somit folgender Zusammenhang:

$$(15) \quad p = 1 - \omega / \omega^* = \gamma\beta / (\gamma\beta + \omega)$$

In dem Modell ohne Zufallseffekte und Messfehler nimmt  $p$  Werte zwischen 0 und 1 an. Taylor und Wang erklären, dass durch die Tatsache, dass alle Parameter aus einem gemeinsamen Modell gewonnen werden, nicht mit Inkonsistenzen wie bei dem ursprünglichen Ansatz von Freedman zu rechnen ist<sup>83</sup>. Darüber hinaus liefern die Parameter  $\gamma$ ,  $\beta$  und  $\omega$  zusätzliche Information über die Therapieeffekte, die allein über den Wert für  $p$  nicht zu erkennen wären. So kann das Ergebnis  $p = 0$  bedeuten, dass entweder der longitudinal gemessene Surrogatendpunkt nicht durch die Therapie beeinflusst wird oder aber der Surrogatendpunkt nicht mit der Überlebenszeit assoziiert ist.  $p$  allein erlaubt diese Unterscheidung nicht. Weir und Walley erläutern in ihrer Übersichtsarbeit, dass die von Taylor und Wang formulierte Modellgleichung auch im Fall von Messfehlern Gültigkeit hat. Sie bezweifeln aber, dass die Gleichung verwendet werden kann, wenn zusätzlich Zufallseffekte berücksichtigt werden müssen, da in diesem Fall die dem Modell zugrunde liegende Annahme des *proportional hazard* verletzt wird<sup>90</sup>.

### **Average prediction error of the estimated predicted effect (APEP)**

Baker greift in seiner Arbeit die Methodik der Validierung binärer Surrogatendpunkte auf Basis von Metaanalysen auf<sup>7</sup>. In seinem Ansatz wird eine Gruppe von Studien, die sowohl Ergebnisse für den Surrogatendpunkt als auch für den klinisch relevanten Endpunkt erhoben haben, sukzessive in voran-

gegangene und neue Studien aufgeteilt. Liegen beispielsweise zehn Studien vor, so könnte die erste Aufteilung so erfolgen, dass die ersten neun Studien als Vorangegangene und die zehnte als Neue bezeichnet werden. Für die zweite Aufteilung könnten dann die ersten acht und die zehnte als vorangegangene sowie die neunte als neue Studie angesehen werden. Da jede der Studien einmal als neue Studie dienen kann, sind somit zehn verschiedene Aufteilungen möglich. Im nächsten Schritt wird dann auf Grundlage jeder einzelnen vorherigen Studie für eine neue prognostiziert, wie hoch der Interventionseffekt auf den klinisch relevanten Endpunkt sein wird, unter der Annahme, dass in der neuen Studie nur der Effekt auf den Surrogatendpunkt ermittelt wird. Die mittels der vorherigen Studien prognostizierten Werte werden anschließend zu einem einzelnen Prognosewert für die neue Studie zusammengeführt und dieses Verfahren für jede der neun anderen neuen Studien wiederholt. Als Validierungsmaß wird die über alle Aufteilungen gemittelte absolute Differenz zwischen den prognostizierten und den beobachteten Therapieeffekten, der sogenannte *APEP*, verwendet. Der *APEP* wird im Vergleich zur *average clinically meaningful difference (ACMD)* betrachtet. Ist der *APEP* wesentlich kleiner als der *ACMD*, dann ist der Fehler, der bei der Verwendung des Surrogatendpunkts zur Vorhersage des klinisch relevanten Effekts gemacht wird, relativ klein im Vergleich zu dem Effekt auf den klinisch relevanten Endpunkt, der mit einer Studie belegt werden soll, so dass dies ein wesentliches Qualitätsmerkmal für einen Surrogatendpunkt darstellt. Dennoch besteht die Möglichkeit, dass ein kleiner *APEP* allein dadurch zustande kommt, dass der Therapieeffekt auf den klinisch relevanten Endpunkt in allen zugrunde liegenden Studien gleich ist, dies aber nicht zwingend heißen muss, dass es einen Zusammenhang zwischen Surrogatendpunkt und klinisch relevantem Endpunkt geben muss. Aus diesem Grund schlägt Baker vor, zusätzlich im Rahmen von Metaanalysen den durchschnittlichen Fehler zu ermitteln, der bei der Vorhersage des Effekts auf andere klinisch relevante Endpunkte unter Verwendung des Surrogatendpunkts gemacht wird (*average prediction error of other true endpoints (APEO)*). Ist der *APEP* kleiner als der *APEO*, ist dies ein weiterer Hinweis auf einen geeigneten Surrogatendpunkt.

### **Zwischenschlussfolgerung**

Insgesamt kann festgehalten werden, dass es derzeit keinen Goldstandard für die Validierung von Surrogatendpunkten gibt. Aufgrund der Fehleranfälligkeit, die mit der Verallgemeinerung von Ergebnissen aus Einzelstudien verbunden ist, sind allerdings Ansätze auf Basis mehrerer Studien, wie dies z. B. bei Metaanalysen der Fall ist, zu bevorzugen. Dennoch gibt es bei keiner Validierungstechnik eine Garantie dafür, dass die ermittelten Zusammenhänge auch bei einer neuen Studie gleichermaßen gelten<sup>8</sup>. Molenberghs und Kollegen machen daher deutlich, dass die Ergebnisse statistischer Validierungsverfahren allein nicht ausreichen, um die Eignung eines Surrogatendpunkts zu beurteilen, sondern zudem Informationen über biologische und klinische Faktoren mit einzubeziehen sind<sup>57</sup>.

## **6.5.2 Aussagen zu Surrogaten in Methodenpapieren von HTA-Agenturen und Arzneimittelbewertungsinstitutionen**

### **6.5.2.1 Quantitative Ergebnisse**

Die INAHTA besteht zum Recherchezeitpunkt (Juni 2007) aus 45 Mitgliedern (öffentlich finanzierte HTA-Agenturen), von denen 15 Agenturen (aus insgesamt zwölf verschiedenen Ländern) eines oder mehrere Methodenpapiere auf ihrer Homepage veröffentlicht haben (n = 24 Methodenpapiere). Diese sind in der folgenden Tabelle 6 aufgelistet. Die Methodenpapiere sind zwischen 1995 und 2008 veröffentlicht worden.

Tabelle 6: INAHTA-Mitglieder mit veröffentlichtem Methodenpapier

Land	Institution	Dokument	Jahr der Veröffentlichung
Australien	ASERNIP-S	General Guidelines for Assessing, Approving & Introducing New Procedures into a Hospital or Health Service	Keine Angaben
		ASERNIP-S Systematic Review Process	2003
	MSAC	Funding for new medical technologies and procedures: application and assessment guidelines	2005
		Guidelines for the assessment of diagnostic technologies	2005
Belgien	KCE	Recommandations provisoires pour les évaluations pharmaco économiques en Belgique	2006
Brasilien	DECEIT-CGATS	Diretrizes metodologicas paea elaboraçao de estudos em avaliaçao de tenologias para o Ministerio de Saude	2007
Dänemark	DACEHTA	Health Technology Assessment Handbook	2001
Deutschland	IQWiG	Methoden Version 3.0	2008
Frankreich	HAS	Guide d'analyse de la litterature et gradations des recommandations	2000
		Bases méthodologiques pour l'élaboration de recommandations professionnelles par consensus formalisé	2006
Kanada	CADTH	Guideline for the economic evaluation of health technologies: Canada	2006
		Guidelines for clinical reviewers common drug review process	2006
		General guideline for reviewers common drug review process	2006
		Procedure for Common Drug Review	2006
Norwegen	NOKC	Slik oppsummerer vi forskning Håndbok for Nasjonalt kunnskapssenter for helsetjenesten	2006
Österreich	LBI@HTA	Externes Manual – Selbstverständnis und Arbeitsweise	2007
Spanien	OSTEBA	La identificación de tecnologías sanitarias emergentes	2000
		Guía para le evaluación del impacto en la salud y en el bienestar de proyectos, programas, o politicas extrasanitarias	2005
		Guía de evaluación económica en el sector sanitario	1999
	AETSA	Framework for the assessment of genetic testing in the andalusian public health system	2006
	AETS	Guía para la elaboración de informes de evaluación de tecnologías sanitarias	1999
		Evaluación epidemiologica de tecnologías de salud	1995
UK	CRD	Undertaking Systematic Reviews of research on effectiveness	2000
USA	VATAP	Evaluating Diagnostic Tests: A Guide to the Literature	2000

AETS = Agencia de Evaluación de Tecnologías Sanitarias. AETSA = Agencia de Evaluación de Tecnologías Sanitarias de Andalucía. ASERNIP-S = Australian Safety and Efficacy Register of New Interventional Procedures – Surgical. CADTH = Canadian Agency for Drugs and Technologies in Health. CRD = Centre for Reviews and Dissemination. DACEHTA = Danish Centre for Evaluation and HTA. DECIT-CGATS = Department of Science and Technology, Brazilian Health Technology Assessment General Coordination. HTA = Health Technology Assessment. INAHTA= International Network of Agencies for Health Technology Assessment. KCE = Belgian Health Care Knowledge Centre. LBI@HTA = Ludwig Boltzmann Institute of Health Technology Assessment. MSAC = Medical Services Advisory Committee. NOKC = Norwegian Knowledge Centre for the Health Services. OSTEBA = Basque Office for Health Technology Assessment. UK = Großbritannien. USA = Vereinigte Staaten von Amerika. VATAP = Veterans Affairs Technology Assessment Program.

Das Methodenpapier aus der norwegischen Agentur kann nicht näher betrachtet werden, da es auf Norwegisch verfasst ist. Es bleiben 23 Methodenpapiere von INAHTA-Mitgliedern, die auf ihre Aussagen bzgl. der in der Bewertung zu verwendenden Ergebnisparameter im Allgemeinen und zur Anwendung von Surrogatendpunkten im Besonderen analysiert werden.

Die mit Hilfe der jeweiligen Webseiten-Suchmotoren durchgeführte Suche in den Webauftritten der INAHTA-Mitglieder liefern keine weiteren relevanten Methodenpapiere bzw. keine relevanten Stellungnahmen zum Stellenwert von Surrogatparametern in der Arbeit der jeweiligen Agenturen.

Darüber hinaus werden elf weitere Institutionen identifiziert, die mit der Evaluation von Arzneimitteln im Rahmen der Preisfestsetzung und von Kostenerstattungsentscheidungen beauftragt sind. Die Durchsicht der Webauftritte dieser Institutionen ermöglicht die Identifikation von zwölf weiteren methodologischen Leitlinien, die für diesen Teil des Berichts analysiert werden, wobei eine in finnischer Sprache verfasst ist und nicht weiter berücksichtigt werden kann (siehe Tabelle 7). Die Leitlinien sind zwischen 1999 und 2006 veröffentlicht worden.

**Tabelle 7: Berücksichtigte Institutionen zur Arzneimittelnutzenbewertung und zu Methodenpapieren**

Land	Institution	Dokument	Jahr der Veröffentlichung
Australien	Pharmaceutical Benefits Advisory Committee	Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee	2006
Kanada	Patented Medicine Prices Review Board	Compendium of Guidelines Policies and Procedures	2003
Niederlande	Health Care Insurance Board	Dutch guidelines for pharmacoeconomic research	1999
Norwegen	Norwegian Medicines Agency	Norwegian guidelines for pharmacoeconomic analysis in connection with applications for reimbursement	2005
Neuseeland	Pharmaceutical Management Agency Ltd (PHARMAC)	Recommended methods to derive clinical inputs for proposals to PHARMAC	2005
		Prescription for Pharmacoeconomic Analysis Methods for cost-utility analysis	2006
Schweden	Pharmaceutical Benefits Board	General guidelines for economic evaluations from the Pharmaceutical Benefits Board	2004
UK	National Institute for Health and Clinical Excellence	A guide for manufacturers and sponsors	2001
		Guide to the Technology Appraisal Process	2004
		Guide to the Methods of Technology Appraisal	2004
		Guidance for manufacturers and sponsors	2004

PHARMAC = Pharmaceutical Management Agency. UK = Großbritannien.

Insgesamt werden 34 Methodenpapiere durch diese Informationsgewinnungsstrategie für die Bearbeitung der zweiten Frage (siehe 6.3 „Forschungsfragen“) identifiziert und analysiert.

### 6.5.2.2 Qualitative Ergebnisse – Aussagen zu Surrogaten in Methodenpapieren

13 der 23 analysierten Methodenpapiere der INAHTA-Mitglieder und sieben der elf analysierten Leitlinien von Arzneimittelbewertungsinstitutionen geben Empfehlungen zur Auswahl der in der Bewertung der Technologie zu verwendenden Ergebnisparameter (siehe Tabelle 22 und Tabelle 23 im Anhang 8.3 „Dokumentation der Methodenpapiere aus HTA-Institutionen“ für die ausführliche Dokumentation der extrahierten Aussagen). Dabei sind sich alle Institutionen darüber einig, dass in HTA-Berichten und Nutzenbewertungen patientenrelevante Ergebnisparameter vorrangig berücksichtigt werden müssen. Als patientenrelevante Endpunkte werden Mortalität, Morbidität (einschließlich Symptomatik), Lebensqualität sowie die durch die Anwendung der Technologie hervorgerufenen unerwünschten Wirkungen bzw. Komplikationen erwähnt (siehe Tabelle 8). Diese werden in mehreren Methodenpapieren

auch finale Ergebnisparameter bzw. definitive Endpunkte genannt, was die Abgrenzung zu dem intermediären, indirekten und kurzfristigen Charakter von Surrogatparametern Nachdruck verleiht. Mit dieser Art von Outcomes können sowohl der langfristige Nutzen als auch der potentielle Schaden der Behandlung erfasst werden. So erläutern beispielsweise die kanadischen und australischen Guidelines, dass bei Erkrankungen, die zu Behinderung oder Tod führen, die Reduktion der Inzidenz von allgemeiner und krankheits-/zustandsspezifischer Mortalität und Morbidität (mit ggf. mehreren krankheitsspezifischen Folgen) als primäre Ergebnisparameter gelten (z. B. tödlicher bzw. nichttödlicher Myokardinfarkt bzw. zerebraler Insult bei kardiovaskulären Erkrankungen), da diese Outcomes sowohl den langfristigen Nutzen als auch den Schaden der Behandlung spiegeln, je nach dem ob sich die Inzidenz von Mortalität/Morbidität erhöht oder verringert<sup>6, 22</sup>.

**Tabelle 8: Von HTA-Agenturen und Arzneimittelbewertungsinstitutionen bevorzugte Ergebnisparameter**

<b>Mortalität</b>	Gesamt Krankheitsspezifisch
<b>Morbidität</b>	Gesamt Krankheitsspezifische Beschwerden und Komplikationen Krankenhausaufenthalt Pflegebedarf
<b>Lebenszeit</b>	Überlebenszeit Krankheitsfreie Überlebenszeit Gewonnene Lebensjahre
<b>Lebensqualität</b>	Gesundheitsbezogene Lebensqualität Krankheitsspezifische Lebensqualitätsmessungen
<b>Nebenwirkungen</b>	Unerwünschte Arzneimittelwirkungen Komplikationen
<b>Andere</b>	Patientenzufriedenheit

HTA = Health Technology Assessment.

Nur drei der betrachteten Organisationen geben eine Definition des Begriffs Surrogatendpunkt (der australische MSAC<sup>56</sup>, die kanadische CADTH<sup>22</sup> und die neuseeländische PHARMAC<sup>69</sup>) wobei die Definitionen des MSAC und der PHARMAC Wort für Wort identisch sind. Surrogatendpunkte werden als physiologische Parameter oder Laborparameter bezeichnet, für die ein valider Zusammenhang zwischen Surrogatendpunkt und korrespondierendem klinisch relevanten Endpunkt nachgewiesen ist.

Die Validität von Surrogatendpunkten wird darüber hinaus von mehreren Organisationen thematisiert (d. h. auch von solchen, die keine begriffliche Definition vorgenommen haben). In diesem Zusammenhang wird die Validität eines Surrogatendpunkts nur dann als akzeptabel gesehen wenn

- a) es eine statistisch nachgewiesene Assoziation zwischen Surrogat und hartem Endpunkt gibt und dieser Zusammenhang darüber hinaus biologisch sowie pathophysiologisch plausibel ist<sup>56, 69</sup>.
- b) es eine starke, unabhängige sowie konsistente Assoziation zwischen dem Surrogatendpunkt und dem klinisch relevanten Endpunkt gibt und darüber hinaus Evidenz aus RCT vorliegt, die zeigt, dass eine Verbesserung/Verschlechterung des Surrogatendpunkts auch zu einer Verbesserung/Verschlechterung des Zielparameters führt<sup>22, 51, 62</sup>.

Das IQWiG macht darüber hinaus deutlich, dass ein Surrogatendpunkt als „nicht valide“ bezeichnet wird, wenn in wissenschaftlichen Untersuchungen gezeigt wird, dass eine Intervention

- a) einen Einfluss auf den Surrogatendpunkt hat, nicht jedoch auf den patientenrelevanten Endpunkt oder
- b) einen Einfluss auf den patientenrelevanten Endpunkt hat, nicht jedoch auf den Surrogatendpunkt oder
- c) inkonsistente Effekte beim Surrogat- und patientenrelevanten Endpunkt erzeugt<sup>51</sup>.

Alle Agenturen betonen in ihren Methodenpapieren, dass grundsätzlich die Verwendung von harten klinischen Endpunkten gegenüber Surrogatendpunkten zu bevorzugen ist, wobei die Mehrzahl der Institutionen beschreibt, dass bei der Erfüllung bestimmter Voraussetzungen die Verwendung von Surrogatendpunkten als Ausnahme akzeptabel sein kann<sup>6, 9, 22, 39, 51, 56, 62, 65, 68, 69</sup>. Dazu ist es im We-

sentlichen erforderlich, dass die oben beschriebene Validität des verwendeten Endpunkts nachgewiesen werden kann. Dies soll i. d. R. durch die Vorlage zusätzlicher Evidenz erfolgen, die den kausalen Zusammenhang zwischen Surrogat und klinischem Endpunkt belegt. Am weitestreichenden sind in diesem Zusammenhang die Vorgaben des australischen Pharmaceutical Benefits Advisory Committee (PBAC). Bei der Verwendung von Surrogatendpunkten wird erwartet, dass zusätzlich zu der eigentlichen Nutzenbewertung folgende Unterlagen eingereicht werden<sup>6</sup>:

- a) Eine systematische Übersichtsarbeit, in der der Zusammenhang zwischen verwendetem und klinisch relevantem Endpunkt unter epidemiologischen Gesichtspunkten sowie hinsichtlich biologischer Plausibilität untersucht wird. Dabei werden einige Zusammenhänge als allgemein anerkannt angesehen. Als Beispiele werden die linksventrikuläre Ejektionsfraktion als Surrogat für Überleben nach Myokardinfarkt sowie die Viruslast als Surrogat für die Heilung viraler Hepatitis-erkrankungen genannt. Aus den Ausführungen in dem Methodenpapier kann entnommen werden, dass für diese beiden Parameter keine weitere Evidenz vorgelegt werden muss. Ob sich der Umfang an akzeptierten Surrogatendpunkten auf diese beiden Beispiele beschränkt, wird aus dem Methodenpapier allerdings nicht deutlich. Aus einer schriftlichen Auskunft der PBAC ist zu entnehmen, dass bisher keine weiteren Surrogatendpunkte als allgemein akzeptiert gelten<sup>67</sup>.
- b) Eine systematische Übersichtsarbeit von RCT, die zeigt, dass die Verwendung ähnlicher Arzneistoffe zu einer Veränderung des Surrogatendpunkts und einer gleichgerichteten Veränderung des klinisch relevanten Endpunkts geführt hat. Zudem soll die Übertragbarkeit der Ergebnisse auf den zu untersuchenden Arzneistoff unter pharmakologischen Gesichtspunkten erläutert werden (z. B. Klasseneffekt).

Ein in den Methodenpapieren der Agenturen häufig genanntes Einsatzgebiet von Surrogatendpunkten ist die pharmakoökonomische Modellierung, in der Kosten-Effektivitäts-Verhältnisse üblicherweise über einen längeren Zeitraum simuliert werden. Zu diesem Zweck werden die Ergebnisse aus Studien mit Surrogatendpunkten um die aus anderen Studien gewonnen Erkenntnisse über das Ausmaß des Zusammenhangs zwischen Surrogat und klinischem Endpunkt erweitert<sup>9, 22, 39, 68, 69</sup>. Auch hier gilt, dass die Validität dieses Zusammenhangs belegt sein muss.

Zusammenfassend zeigen die Empfehlungen der hier berücksichtigten Institutionen dahingehend Übereinstimmung, dass die Bewertung von Technologien nur in sehr gut begründeten Ausnahmefällen ausschließlich auf Basis von Surrogatendpunkten erfolgen sollte.

### **6.5.3 Anwendung von Surrogatendpunkten in HTA-Berichten**

#### **6.5.3.1 Internationale HTA-Berichte**

Über die am 10.07.2007 durchgeführte Datenbankabfrage werden insgesamt N = 6.081 zu diesem Zeitpunkt abgeschlossene HTA-Berichte in der HTA-Datenbank identifiziert. Davon erfüllten N = 5.372 das Einschlusskriterium in deutscher, englischer, französischer, spanischer oder portugiesischer Sprache publiziert worden zu sein. Die 5 %-Zufallsstichprobe umfasst insgesamt n = 269 HTA-Berichte. Insgesamt 128 Berichte der Zufallsstichprobe werden nachträglich aus verschiedenen Gründen ausgeschlossen (siehe Abbildung 4). Die Anzahl der für die Untersuchung verbleibenden Berichte verringert sich auf insgesamt n = 140. Die Liste der untersuchten Berichte ist im Anhang 8.5 „HTA-Berichte aus der HTA-Datenbank“ dokumentiert.

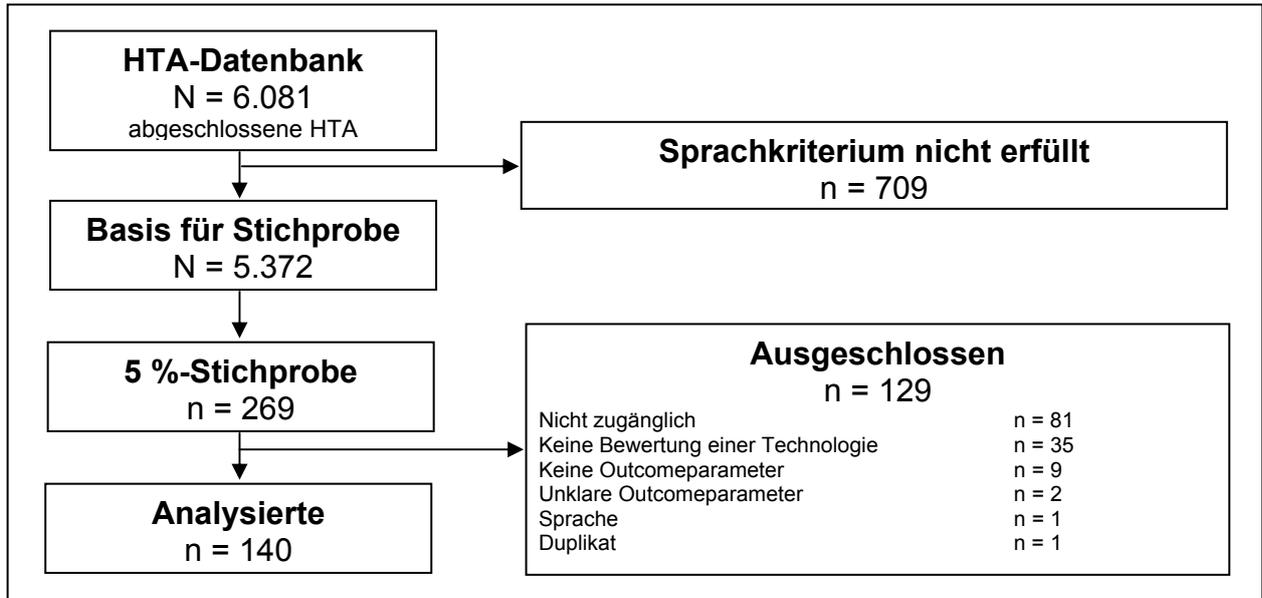


Abbildung 4: Study-Flow internationaler HTA-Berichte

HTA = Health Technology Assessment.

Wie Tabelle 9 zu entnehmen ist, weisen die Berichte der Zufallsstichprobe (n = 269) zwar eine ähnliche Länder- und Jahresverteilung auf wie die Berichte der Grundgesamtheit (n = 5.372). Die Verteilungen in der *endgültigen* Stichprobe (n = 140, nach Entfernung eines Duplikats) weichen aber etwas von denen in der Gesamtstichprobe ab. So stammen 56 % der Berichte in der endgültigen Stichprobe von europäischen Agenturen, wohingegen der Anteil in der Gesamtstichprobe nur 45 % beträgt. Auch der Prozentsatz kanadischer und australischer bzw. neuseeländischer Berichte ist mit insgesamt 32 % um sechs Prozentpunkte höher als in der Gesamtstichprobe. Diese Abweichungen sind vor allem dem Sachverhalt geschuldet, dass viele US-amerikanische Berichte kostenpflichtig bzw. nicht über das Internet erhältlich sind. Der Anteil US-amerikanischer Berichte in der endgültigen Stichprobe ist daher mit 9 % wesentlich geringer als in der Gesamtstichprobe mit insgesamt 24 %. Darüber hinaus sind viele ältere Berichte (vor 2000) nicht über das Internet verfügbar, was dazu führt, dass der Prozentsatz neuerer Studien (die nach 2005 veröffentlicht werden) in der letztendlich untersuchten Stichprobe deutlich höher ist als in der Gesamtstichprobe (34 % versus 25 %). Der Anteil älterer Studien (vor 2000 veröffentlicht) fällt hingegen wesentlich geringer aus (9 % versus 17 %) als in der ursprünglichen Stichprobe. Die berücksichtigten Berichte werden zwischen 1999 und 2007 veröffentlicht.

Tabelle 9: Länder- und Jahresverteilung in der Grundgesamtheit, in der 5 %-Zufalls- und in der endgültigen Stichprobe

Merkmal	Alle die Einschlusskriterien erfüllende Berichte (Grundgesamtheit, n = 5.372)	5 %-Zufallsstichprobe (n = 269)	Endgültige Stichprobe (n = 140)
Land			
Europa	2.442 (45 %)	135 (50 %)	79 (56 %)
USA	1.310 (24 %)	59 (22 %)	13 ( 9 %)
Südamerika	139 ( 3 %)	3 ( 1 %)	2 ( 1 %)
Kanada	934 (17 %)	50 (19 %)	30 (21 %)
Australien/Neuseeland	509 ( 9 %)	20 ( 7 %)	15 (11 %)
Asien	38 ( 1 %)	2 ( 1 %)	1 ( 1 %)

**Fortsetzung Tabelle 9: Länder- und Jahresverteilung in der Grundgesamtheit, in der 5 %-Zufalls- und in der endgültigen Stichprobe**

Publikationsjahr			
< 2000	912 (17 %)	55 (20 %)	13 ( 9 %)
2000-2004	3.001 (56 %)	148 (55 %)	80 (57 %)
≥ 2005	1.459 (27 %)	66 (25 %)	47 (34 %)

USA = Vereinigte Staaten von Amerika.

Die internationalen Berichte decken verschiedene Technologien ab, wobei das Assessment von medizinischen und chirurgischen Prozeduren die Mehrheit der Berichte darstellt (siehe Tabelle 10). Das Spektrum der Indikationen ist sehr breit, wobei verschiedene Gebiete der klinischen Praxis vertreten sind z. B. verschiedene Spezialisierungen der Inneren Medizin, Chirurgie, Frauenheilkunde, Augenheilkunde oder Neurologie (siehe Tabelle 15).

**Tabelle 10: Internationale HTA-Berichte - Bewertete Technologien (n = 140)**

Technologie	n (%)
Arzneimittel	46 (32,9 %)
Prozeduren	72 (51,4 %)
Diagnostik/Screening	19 (13,6 %)
Andere	3 ( 2,1 %)

HTA = Health Technology Assessment.

In weniger als der Hälfte der internationalen HTA-Berichte werden die im Assessment zu extrahierenden Ergebnisparameter bei der Formulierung der Fragestellung bzw. im Methodenteil explizit und detailliert festgelegt. In etwas weniger als die Hälfte der analysierten internationalen HTA-Berichte erfolgt eine eindeutige prospektive Festlegung der zu verwendenden Parameter. Bei der knappen Mehrheit werden höchstens Aussagen wie „*all outcomes reported in included studies were considered*“ oder „*information relevant to the safety and/or efficacy*“ diesbezüglich getroffen, ohne genauer darzulegen ob und wenn ja, welche Ergebnisparameter prospektiv festgelegt werden (siehe Tabelle 11).

**Tabelle 11: Prospektive Festlegung der für die Bewertung relevanten Ergebnisparameter in internationalen HTA-Berichten (n = 140)**

	n (%)
<b>Ja (gesamt)</b>	<b>67 (47,9 %)</b>
<b>Nach Technologie</b>	
Arzneimittel	21/46 (46 %)
Nicht-medikamentöse Verfahren	34/75 (45 %)
Diagnostik/Screening	12/19 (64 %)
<b>Nein (gesamt)</b>	<b>73 (52,1 %)</b>
<b>Nach Technologie</b>	
Arzneimittel	25/46 (54 %)
Nicht-medikamentöse Verfahren	41/75 (55 %)
Diagnostik/Screening	6/19 (36 %)

HTA = Health Technology Assessment.

Insgesamt werden in 87 (62 %) der internationalen HTA-Berichte die Verwendung von Surrogatparametern geplant bzw. Surrogatparameter aus den für das Assessment vorliegenden Studien extrahiert und berichtet. Fast alle Berichte betrachten einen klinisch relevanten, patientennahen Endpunkt. Nur in fünf Berichten werden ausschließlich Surrogatendpunkte geprüft (siehe Tabelle 12, Definitionen siehe 6.4.3.5 „Definition verwendeter Endpunkte“).

**Tabelle 12: Art der verwendeten Endpunkte in internationalen HTA-Berichten (n = 140)**

Art der Endpunkte*	Betrachtet/berichtet in**
Klinisch relevante Endpunkte	134 (96 %)
Harte Endpunkte	131 (93 %)
Patient Reported Outcomes	78 (56 %)
Sicherheitsendpunkte	122 (87 %)
Surrogatendpunkte	87 (62 %)
<i>Ausschließlich Surrogatendpunkte</i>	5 (3,6 %)

\* Definitionen siehe 6.4.3.5 „Definition verwendeter Endpunkte“.

\*\* Mehrfachnennungen möglich.

HTA = Health Technology Assessment.

Eine getrennte Betrachtung der diagnostischen Technologien einerseits und der anderen Interventionen andererseits ergibt ein genaueres Bild über die Verwendung von Surrogatendpunkten in HTA-Berichten (Tabelle 13): Bei der Bewertung von diagnostischen Technologien kommen Surrogatparameter häufiger in Betracht als bei der Bewertung von therapeutischen Technologien. HTA-Berichte bei denen Arzneimittel oder nicht-medikamentöse Verfahren evaluiert werden, betrachten klinisch relevante, patientennahe Endpunkte häufiger als HTA-Berichte zu diagnostischen Technologien. In vier der fünf Berichte, die ausschließlich Surrogatparameter berücksichtigen, werden Testcharakteristika (Sensitivität, Spezifität, Likelihood ratios etc.) extrahiert und berichtet. Ein anderer Bericht, der ausschließlich Surrogatparameter berücksichtigt, evaluiert Maßnahmen zur Steigerung der der Teilnahme am Mammographiescreening. Der verwendete Parameter war die Teilnehmerate<sup>76</sup>. Somit wird in keinem Bericht zu nicht-diagnostischen Technologien die Bewertung ausschließlich anhand von Surrogatendpunkten vorgenommen.

**Tabelle 13: Verwendete Endpunkten in internationalen HTA-Berichten – Vergleich diagnostische versus therapeutische Technologien**

Art der Endpunkte*	Technologie (n = 140)	
	Diagnostische** (n = 19)	Nicht diagnostische** (n = 121)
Klinisch relevante Endpunkte	14 (74 %)	120 (99 %)
Harte Endpunkte	13 (68 %)	118 (97 %)
Patient Reported Outcomes	6 (32 %)	72 (60 %)
Sicherheitsendpunkte	11 (58 %)	111 (92 %)
Surrogatendpunkte	17 (89 %)	70 (58 %)
<i>Ausschließlich Surrogatendpunkte</i>	5 (26 %)	0 ( 0 %)

\* Definitionen siehe 6.4.3.5 „Definition verwendeter Endpunkte“.

\*\* Mehrfachnennungen möglich.

HTA = Health Technology Assessment.

Ob die Bewertungsparameter prospektiv festgelegt sind oder nicht, hat keinen wesentlichen Einfluss auf die Art der verwendeten Endpunkte (Tabelle 14).

**Tabelle 14: Verwendete Endpunkte in internationalen HTA-Berichten – prospektiv versus nicht-prospektiv festgelegt**

Art der Endpunkte*	Festlegung der Parameter für die Bewertung (n = 140)	
	Prospektiv (n = 67)**	Ad hoc (n = 73)**
Klinisch relevante Endpunkte	65 (97 %)	69 (94 %)
Harte Endpunkte	62 (92 %)	69 (94 %)
Patient Reported Outcomes	41 (61 %)	37 (51 %)
Sicherheitsendpunkte	56 (84 %)	66 (90 %)
Surrogatendpunkte	42 (63 %)	45 (62 %)
<i>Ausschließlich Surrogatendpunkte</i>	2 ( 3 %)	3 ( 4 %)

\* Definitionen siehe 6.4.3.5 „Definition verwendeter Endpunkte“.

\*\* Mehrfachnennungen möglich.

Die folgende Tabelle 15 gibt einen Überblick zu den in internationalen HTA-Berichten verwendete Surrogatparameter bei nicht-diagnostischen Technologien. In keinem dieser Berichte basierte die Bewertung ausschließlich auf Surrogatparametern. Mit einer einzigen Ausnahme, wird in allen Berichten zu therapeutischen Interventionen die Wirksamkeit auch anhand klinisch relevanter und/oder patientennaher Ergebnisparameter bewertet. In einem Bericht wird zwar die Wirksamkeit nur anhand des Surrogatparameters evaluiert, für die Gesamtbewertung der Technologie werden jedoch auch die unerwünschten Effekte betrachtet. Es handelt sich um einen Bericht des englischen National Institute for Health and Clinical Excellence (NICE) zu antiviraler Behandlung von Hepatitis C. In diesem Dokument wird auf den Forschungsbedarf bzgl. Parameter wie Lebensqualität und Progression der Erkrankung hingewiesen<sup>63</sup>. Das Dokument beruht auf einem ausführlichen HTA-Bericht, in dem neben Surrogatparametern auch Lebensqualität, unerwünschte Effekte und Überleben analysiert werden<sup>77</sup>.

**Tabelle 15: Überblick über Surrogatendpunkte in internationalen HTA-Berichten – nicht-diagnostische Technologien**

Zielkondition	Surrogatendpunkte	Klinisch-/patientenrelevante
Augenheilkunde	Visuell evozierten Potentiale	Sehschärfe Sehfeld Augensymptome
(Auto)immun- erkrankungen	Autoantikörper Titer Nasaler Luftflusswiderstand Eosinophile im nasalen Sekret Volumen von nasalem Sekret Radiologischer Befund (z. B. Sharp-Kriterien)	Krankheitsaktivitätsscores/Responderkriterien die Allgemeinzustand, Symptome, Beschwerden und/oder Laborbefunde kombinieren (z. B. ACR-Response) Beschwerden (z. B. Niesen-Score, Rhinorrhea-Score, multisymptome Scores) Lebensqualität Medikamentenverbrauch (z. B. Antihistaminika, Kortison) Asthmainzidenz Schlafstörungen Einschränkungen im Aktivitäten (u. a. Fahren)
Diabetes	HbA1c Nüchtern-Blutzuckerspiegel Lipidspiegel % Körperfett Gewicht Body-Mass-Index	Diabeteskomplikationen (langfristige) Akute Hyper- und Hypoglykämien Mortalität Kardiovaskuläre Morbidität und Mortalität Lebensqualität
Gastrointestinale Erkrankungen	Magenentleerung Nahrungsmittelvolumen Gewichtszunahme Albuminspiegel	Beschwerden (z. B. Schmerz, Übelkeit etc.) Lebensqualität Mortalität
Geburtsmedizin	Gewicht bei Geburt Gestationsalter Placenta-Blutfluss Placenta-Morphologie Implantationsrate Schwangerschaftsrate Entbindungsrate	Überleben Komplikationen Lebendgeburt
HNO	Elektro-Cochleographie Apnoe/hypoapnoe Index Schlafaufzeichnungen	Anzahl von symptomatischen Vertigo-Anfällen Soziale Funktionsfähigkeit Audition Lebensqualität Morbidität (einschließlich Unfälle)

**Fortsetzung Tabelle 15: Überblick über Surrogatendpunkte in internationalen HTA-Berichten – nicht-diagnostische Technologien**

<p>Infektions- erkrankungen</p>	<p>Virale Last Virales Ansprechen (sustained virological response) Biochemisches Ansprechen Histologisches Ansprechen Besiedelung mit Erreger, Nachweis von Erreger Immunogenität (Immunglobulin/Antikörper Titer)</p>	<p>Manifeste, behandlungsbedürftige Infektionen (Inzidenz, Dauer, Schweregrad) Dauer und Intensität der Symptome Infektionsfreie Intervalle Mortalität (gesamt, krankheitsspezifisch) Nicht-infektiöse Morbidität Krankenhausaufnahmen und Dauer Arbeits-/Schulabwesenheit Lebensqualität</p>
<p>Kardiovaskuläre Erkrankungen</p>	<p>LVEF Oxidativer Stress Messungen/antioxidative Parameter Angiografische Endpunkte (Referenz-Durchmesserstenose, minimaler Gefäßdurchmesser, akuter Lumen-Gewinn, später Lumen-Verlust, Netto-Lumen-Gewinn) Blutdruck Herzrhythmus Cholesterinspiegel Muskelperfusion Angiogenesis Rheologische Parameter Aneurysma-Durchmesser Endoleaks Adhärenz zur antihypertensive Medikation Adhärenz zu diätätischen und Gewichtsreduktionsmaßnahmen Zentraler Venendruck Druckgradient Hämodynamische Funktion</p>	<p>Mortalität (gesamt, krankheitsspezifisch) Inzidenz kardiovaskulärer Ereignisse (Schlaganfall, Herzinfarkt) Anginasymptomatik Dauer des Krankenhausaufenthalts Schmerzfreie Gehstrecke Lebensqualität Konversion von einer Prozedur zu einer anderen Notwendigkeit einer 2. Intervention Revaskularisation</p>
<p>Krebs- erkrankungen</p>	<p>Tumorreduktion Remissionen, Teilremissionen Tumorresponse (z. B. radiologisch) Zytogenetische Response Hämatologische Response PSA Immunologische Parameter</p>	<p>Überlebenszeit (z. B. 5-J, 2-J) Überlebensrate Komplikationen Krankheits- bzw. progressionsfreie Überlebenszeit Reduktion von schweren Symptomen (z. B. Dysphagie, Schmerzen) Zeit frei von Symptomen Lebensqualität Hematemesis Mortalität assoziiert mit Behandlung Psychosoziale Outcomes</p>
<p>Lebererkrankungen</p>	<p>Virale Response</p>	<p>Überlebenszeit Hepatoenzephalopathie Grad/Dekompensierung Leberkarzinominzidenz Leberzirrhoseinzidenz Dekompensierung</p>
<p>Lungen- erkrankungen</p>	<p>Lungenfunktionstests (FEV1, PEFR) Knochendichte, Knochenmetabolismus (Bewertung von Kortisoneffekt) Patienten mit korrekter Applikation von Sprays</p>	<p>Häufigkeit der Anfälle, Exazerbationen Mortalität Mediakionsverbrauch (z. B. kurzwirksame Beta2-Agonisten, Steroide) Lebensqualität Asthmasymptomatik Dyspnoe-Scores Hospitalisierung</p>

**Fortsetzung Tabelle 15: Überblick über Surrogatendpunkte in internationalen HTA-Berichten – nicht-diagnostische Technologien**

Neurologische Erkrankungen	Muskeltonus Ganganalysen Passive Motilität	Aktive Motilität Lebensqualität
Orthopädie, Knochen-erkrankungen	Radiologische Befunde (z. B. Wirbelhöhe, Prothesenposition) Kyphose-Winkel Acetabulare-Winkel Dauer der Operation Prothesen-Funktionalität Knochendichte	Schmerzen Funktionsfähigkeit, Bewegungseinschränkung (z. B. Oswestry Disability Index) Länge des Krankenhausaufenthalts Schmerzmittelkonsum Frakturen Lebensqualität Überlebenszeit der Prothesen
Transplantations- medizin	Länge der Gefäße Funktion des Transplantats (anhand z. B. biochemischer Parameter, z. B. Glomeruläre Filtrationsrate, Serum-Kreatinin)	Mortalität (auch Spendermortalität bei Lebendspenden) Morbidität (z. B. immunsuppressionsbedingte Erkrankungen wie Infektionen, Krebserkrankungen) Lebensqualität Zeit bis zur Wiederherstellung der sozialen Funktionsfähigkeit
Wunden	Epithelialisierung Aspekt des Transplantats Volumen Fläche Tiefe	Zeit bis zur vollständigen Heilung Länge des Krankenhausaufenthalts Länge der Behandlung Behandlungsversagen Transplantatwiederholung

ACR = American college of rheumatology. FEV1 = Forced Expiratory Volumen in 1 second. HbA1C = Glykolisiertes Hämoglobin. HNO = Hals-Nase-Ohren. HTA = Health Technology Assessment. J = Jahre. LVEF = Linksventrikuläre Ejektionsfraktion. PEFR = Peak Expiratory Flow Rate. PSA = Prostataspezifisches Antigen. SCORAD = Scoring index of atopic dermatitis.

### 6.5.3.2 Deutsche Berichte

Die HTA-Datenbank der DAHTA beim DIMDI wird im Mai 2008 abgefragt. Insgesamt werden 117 Berichte identifiziert, die in Deutschland produziert worden sind und kein methodisches Thema haben. Alle Berichte sind durch das Internet zugänglich. Es werden sechs Berichte ausgeschlossen, die nur ökonomische Aspekte berücksichtigen sowie ein weiterer bei dem die verwendeten Outcomeparameter unklar sind. Somit bleiben für die Extraktion der Outcomeparameter 110 HTA-Berichte aus dieser Quelle.

Die Website des IQWiG wird im August 2008 abgefragt, wobei die Suche auf „Abschlussbericht erstellt“ beschränkt wird. Dabei werden 34 Berichte identifiziert, von denen 21 eine Nutzenbewertung vornehmen. Diese werden in die Analyse eingeschlossen.

Insgesamt werden 131 Berichte von deutschen Institutionen in diesem Teil der Arbeit berücksichtigt, die zwischen 1999 und 2007 veröffentlicht werden (siehe Tabelle 16). Die Produktion von HTA-Berichten in Deutschland hat seit 2004 stark zugenommen. Die Liste der untersuchten Berichte ist im Anhang 8.6 „Deutsche HTA-Berichte“ dokumentiert.

Die deutschen Berichte decken verschiedene Technologien ab, wobei das Assessment von medizinischen und chirurgischen Prozeduren die Mehrheit der Berichte darstellt (siehe Tabelle 16). Das Spektrum der abgedeckten Indikationen ist sehr breit, wobei verschiedene Gebiete der klinischen Praxis vertreten sind z. B. Onkologie, Kardiologie, Pneumologie, Chirurgie, Orthopädie, Frauenheilkunde, Dermatologie oder Zahnmedizin.

Tabelle 16: Berücksichtigte deutsche HTA-Berichte

Institution	Anzahl (N = 131)	Prozent
KBV	3	2,5 %
BUA	23	17,5 %
HTA-Projekt (DIMDI)	25	19,0 %
DAHTA@DIMDI	55	42,0 %
Private Unternehmen	4	3,0 %
IQWiG	21	16,0 %
<b>Publikationsjahr</b>		
< 2001	15	11,5 %
2001-2004	43	32,8 %
≥ 2005	73	55,7 %
<b>Technologie</b>		
Prozeduren	66	50,4 %
Arzneimittel	21	16,0 %
Diagnostik	18	13,7 %
Screening	15	11,5 %
Andere	11	8,4 %

BUA = Bundesausschuss der Ärzte und Krankenkassen. DAHTA = Deutsche Agentur für Health Technology Assessment. DIMDI = Deutsches Institut für Medizinische Dokumentation und Information. HTA = Health Technology Assessment. IQWiG = Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. KBV = Kassenärztliche Bundesvereinigung.

Nicht in allen deutschen HTA-Berichten werden die im Assessment zu extrahierenden Ergebnisparameter bei der Formulierung der Fragestellung bzw. im Methodenteil explizit und detailliert festgelegt. In ca. einem Drittel der analysierten deutschen HTA-Berichte wird lediglich erwähnt, dass der Bericht beispielweise sich mit „Nutzen und Risiken“, „Wirksamkeit und Sicherheit“ beschäftigt, ohne genauer darzulegen ob und, wenn ja, welche Ergebnisparameter prospektiv festgelegt werden (siehe Tabelle 17). Die Berichte des IQWiG sind hier eine Ausnahme, da sie immer in einer speziellen Sektion im Kapitel Methodik die berücksichtigten Ergebnisparameter detailliert auflisten. Das Gegenteil sind die Berichte des Bundesausschuss der Ärzte und Krankenkassen (BUA), die Beratungen und Entscheidungsprozesse der Selbstverwaltung dokumentieren. In diesen Dokumenten wird nicht darüber berichtet, dass die Parameter für die Bewertung der Wirksamkeit a priori festgelegt worden sind. Ein Teil des Beratungsprozesses besteht darin, die Ziele der Behandlungen sowie die Zielgrößen zu identifizieren, mit denen die Technologie bewertet wird. So, enthalten alle BUA-Berichte im Anhang ein Fragenkatalog mit u. a. Fragen wie: „Welche prioritären Ziele gelten für die Behandlung der o. g. Erkrankung?, Welche dieser Ziele sind (ggf. teilweise), in welchem Stadium der o. g. Erkrankung, mit der [Technologie] zu erreichen?, Mit welcher Diagnostik (z. B. Untersuchungstechniken, Apparaten, Gesundheitsskalen) werden die therapeutischen Ergebnisse (Zielgrößen) gemessen?“. Aus diesem Grund, werden in den BUA-Berichten nicht die Zielgrößen prospektiv festgelegt sondern alle in den berücksichtigten Studien berichteten Parameter extrahiert. Eine Ausnahme hierfür ist jedoch in dem Bericht zu den Beratungen über die Positronen-Emissions-Tomografie zu sehen, bei dem in den Ausführungen zur Bewertung der Literatur die präferierten Endpunkte festgelegt werden<sup>4</sup>.

Tabelle 17: Prospektive Festlegung der für die Bewertung relevanten Ergebnisparameter in deutschen HTA-Berichten (n = 131)

	n (%)
<b>Ja (gesamt)</b>	<b>84 (64 %)</b>
<b>Nach Institutionen</b>	
BUA	1/23 (4 %)
Deutsches HTA-Projekt (DIMDI)	22/25 (88 %)
DAHTA beim DIMDI	36/55 (66 %)
IQWiG	21/21 (100 %)
Andere	4/7 (57 %)
<b>Nach Technologie</b>	
Arzneimittel	17/21 (81 %)
Nicht-medikamentöse Verfahren	40/77 (51 %)
Diagnostik	14/18 (77 %)
Screening	14/15 (93 %)

**Fortsetzung Tabelle 17: Prospektive Festlegung der für die Bewertung relevanten Ergebnisparameter in deutschen HTA-Berichten (n = 131)**

<b>Nein (gesamt)</b>	<b>47 (36 %)</b>
<b>Nach Institutionen</b>	
BUA	22/23 (96 %)
Deutsches HTA-Projekt (DIMDI)	3/25 (12 %)
DAHTA beim DIMDI	19/55 (34 %)
IQWiG	0/21 (0 %)
Andere	3/7 (43 %)
<b>Nach Technologie</b>	
Arzneimittel	4/21 (19 %)
Nicht-medikamentöse Verfahren	36/77 (47 %)
Diagnostik	4/18 (22 %)
Screening	1/15 (7 %)

BUA =Bundesausschuss der Ärzte und Krankenkassen. DAHTA = Deutsche Agentur für Health Technology Assessment. DIMDI = Deutsches Institut für Medizinische Dokumentation und Information. HTA = Health Technology Assessment. IQWiG = Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen.

Insgesamt werden in 74 (56 %) der deutschen HTA-Berichte die Verwendung von Surrogatparametern geplant bzw. Surrogatparameter aus den für das Assessment vorliegenden Studien extrahiert und berichtet. Fast alle Berichte betrachten einen klinisch relevanten, patientennahen Endpunkt (siehe Tabelle 18, Definitionen siehe 6.4.3.5 „Definition verwendeter Endpunkte“).

**Tabelle 18: Art der verwendeten Endpunkte in deutschen HTA-Berichten (n = 131)**

Art der Endpunkte*	Festlegung der Parameter für die Bewertung (n = 140)	
	Prospektiv (n = 67)**	Ad hoc (n = 73)**
Klinisch relevante Endpunkte	65 (97 %)	69 (94 %)
Harte Endpunkte	62 (67 %)	69 (94 %)
Patient Reported Outcomes	41 (61 %)	37 (51 %)
Sicherheitsendpunkte	57 (85 %)	41 (87 %)
Surrogatendpunkte	42 (63 %)	45 (62 %)
<i>Ausschließlich Surrogatendpunkte</i>	<i>2 (3%)</i>	<i>3 (4 %)</i>

\* Definitionen siehe 6.4.3.5 „Definition verwendeter Endpunkte“.

\*\* Mehrfachnennungen möglich.

HTA = Health Technology Assessment.

Eine getrennte Betrachtung der diagnostischen Technologien einerseits und der anderen Interventionen andererseits ergibt folgendes Bild (Tabelle 19): Bei der Bewertung von diagnostischen Technologien kommen Surrogatparameter häufiger in Betracht als bei der Bewertung anderer Technologien. HTA-Berichte, in denen Arzneimittel oder nicht-medikamentöse Verfahren evaluiert werden, betrachten klinisch relevante, patientennahe Endpunkte häufiger als HTA-Berichte zu diagnostischen Technologien.

**Tabelle 19: Verwendete Endpunkte in deutschen HTA-Berichten – Vergleich diagnostische versus therapeutische Technologien**

Art der Endpunkte*	Technologie (n = 130)	
	Diagnostische** (n = 33)	Nicht diagnostische** (n = 97)
Klinisch relevante Endpunkte	27 (82 %)	97 (100 %)
Harte Endpunkte	23 (70 %)	96 (99 %)
Patient Reported Outcomes	7 (21 %)	60 (62 %)
Sicherheitsendpunkte	24 (73 %)	90 (93 %)
Surrogatendpunkte	29 (88 %)	44 (45 %)
<i>Ausschließlich Surrogatendpunkte</i>	<i>6 (19 %)</i>	<i>0 (0 %)</i>

\* Definitionen siehe 6.4.3.5 „Definition verwendeter Endpunkte“.

\*\* Mehrfachnennungen möglich.

HTA = Health Technology Assessment.

Ob die Bewertungsparameter prospektiv festgelegt sind oder nicht, hat keinen wesentlichen Einfluss auf die Art der verwendeten Endpunkte, wobei die wenigen Berichte, die ausschließlich Surrogatparameter betrachten, dies prospektiv festgelegt haben (Tabelle 20).

**Tabelle 20: Verwendete Endpunkte in deutschen HTA-Berichten – prospektiv versus nicht-prospektiv festgelegt**

Art der Endpunkte*	Festlegung der Parameter für die Bewertung (n = 131)	
	Prospektiv (n = 84)**	Ad hoc (n = 47)**
Klinisch relevante Endpunkte	78 (93 %)	47 (100 %)
Harte Endpunkte	74 (88 %)	46 (98 %)
Patient Reported Outcomes	42 (50 %)	26 (55 %)
Sicherheitsendpunkte	74 (88 %)	41 (87 %)
Surrogatendpunkte	48 (57 %)	26 (55 %)
<i>Ausschließlich Surrogatendpunkte</i>	6 (7 %)	0 (0 %)

\* Definitionen siehe 6.4.3.5 „Definition verwendeter Endpunkte“.

\*\* Mehrfachnennungen möglich.

HTA = Health Technology Assessment.

Über ein Drittel (n = 29, 39 %) der Berichte mit Surrogatparametern bewerten nicht-therapeutische Interventionen (d. h. diagnostische Technologien oder Screeningverfahren). Nur in sechs der 74 Berichte, die Surrogatparameter verwenden, werden keine weiteren Parameter berücksichtigt. Das heißt, dass von den 131 deutschen HTA-Berichten nur 4,5 % ausschließlich auf Surrogatendpunkten beruhen. Davon evaluieren fünf eine diagnostische Technologie unter ausschließlicher Betrachtung der Testcharakteristika (Sensitivität, Spezifität, prädiktive Werte, diagnostische Genauigkeit und/oder Likelihood ratios); ein Bericht bewertet verschiedene Maßnahmen zur Steigerung der Teilnehmeraten an Screeningprogramme, wobei nur die Zielgröße „Teilnehmerate“ berücksichtigt wird.

Die folgende Tabelle 21 gibt einen Überblick über die verwendeten Surrogatparameter in deutschen HTA-Berichten von nicht-diagnostischen Technologien. In keinem dieser Berichte basiert die Bewertung ausschließlich auf Surrogatparametern. Mit einer einzigen Ausnahme, wird in allen Berichten zu therapeutischen Interventionen, die Wirksamkeit auch anhand klinisch relevanter und/oder patienten-naher Ergebnisparameter bewertet. In einen Bericht wird zwar die Wirksamkeit nur anhand des Surrogatparameters evaluiert, für die Gesamtbewertung der Technologie werden jedoch auch die unerwünschten Effekte betrachtet. Die Autoren dieses Berichts weisen zudem darauf hin, dass die Verwendung des Surrogatendpunkts (*sustained viral response* – SVR) „auf der Annahme [beruht], dass die SVR als valider und relevanter Surrogatendpunkt gelten kann. Nach Ansicht der Autoren sind jedoch weitere Studien erforderlich, um dies zu bestätigen.“<sup>80</sup>.

In den HTA-Berichten mit einer prospektiven Festlegung der zu verwendenden Parameter, wird deutlich auf den nachrangigen Charakter der Surrogatparameter hingewiesen. Die Mehrheit der in Tabelle 21 dokumentierten Surrogatendpunkte stammt aus Berichten des BUA, bei denen verfahrenskonform alle der in der eingeschlossenen Studien berichteten Ergebnisparameter für das Assessment extrahiert werden. Allerdings wird von den Auswertern der in den BUA-Berichten eingeschlossenen Studien regelmäßig auf den Mangel an Evidenz bzgl. der Validität der Surrogatparameter hingewiesen.

**Tabelle 21: Überblick über Surrogatendpunkte in deutschen HTA-Berichten – nicht-diagnostische Technologien**

Zielkondition	Surrogatendpunkte	Klinisch-/patientenrelevant
Autoimmun- erkrankungen	Autoantikörper Titer	Krankheitsaktivitätsscores die Allgemeinzustand, Symptome, Beschwerden und Laborbefunde kombinieren (z. B. ACR-Kriterien) Beschwerden
Diabetes	HbA1c Nüchtern-Blutzuckerspiegel Lipidspiegel	Erblindungsrate Rate terminaler Niereninsuffizienzen mit Dialysenotwendigkeit Amputationsrate Rate ketoazidotischer Komata Rate an schwere Hypoglykämien Mortalität Kardiovaskuläre Morbidität und Mortalität Hospitalisierung Durch chronische Hyperglykämie bedingten Symptomatik Lebensqualität Therapiezufriedenheit
Frakturen, Knochen- erkrankungen	Radiologische Konsolidierung Radiologische Befunde (z. B. Pseudarthrose, Fehlstellung)	Schmerzen Funktionsfähigkeit, Bewegungseinschränkung Zeit bis zur Wiederherstellung der Funktionsfähigkeit
Gelenk- erkrankungen	Knorpelqualität (histologische immunhistologische Untersuchungen, MRT) Knorpelbelastbarkeit (arthroskopisch)	Funktionsfähigkeit und Aktivität Schmerzen Lebensqualität Nebenwirkungen
Hauterkrankungen	Minimale UVB-Erythem-Dosis IgE-Spiegel Kutane Elastometrie	Hauterscheinungsscores (z. B. PASI, SCORAD) Lebensqualität Pruritus
Infektions- erkrankungen	Virale Last (sustained virological response) Besiedelung mit Erregern, Nachweis von Erregern Immunogenität Mukosaler Antikörper Kreuzimmunität	Manifeste, behandlungsbedürftige Infektionen Dauer und Intensität der Symptome Mortalität Morbidität Lebensqualität
Kardiovaskuläre Erkrankungen	LVEF Gefäßmotilität Oxidative Stressmessungen Angiografische Endpunkte (Referenz-Durchmesserstenose, minimaler Gefäßdurchmesser, akuter Lumen-Gewinn, später Lumen-Verlust, Netto-Lumen-Gewinn) Blutdruck Herzfrequenz Cholesterinspiegel Blutgase Muskelperfusion Rheologische Parameter	Mortalität Inzidenz kardiovaskulärer Ereignisse Krankenhausaufenthalte Schmerzfreie Gehstrecke Verbrauch von Nitratmedikation Lebensqualität
Krebs- erkrankungen	Tumorreduktion (MRT-Aufnahme, CT-Aufnahme) CEA Remissionen, Teilremissionen	Überlebenszeit Schmerzen Komplikationen Krankheits- bzw. progressionsfreie Überlebenszeit
Lebererkrankungen	Fischer Index Biochemische Laborparameter (z. B. Phospholipidfettsäuren)	Überlebenszeit Hepatoenzephalopathie-Grad

**Fortsetzung Tabelle 21: Überblick über Surrogatendpunkte in deutschen HTA-Berichten – nicht-diagnostische Technologien**

Zielkondition	Surrogatendpunkte	Klinisch-/patientenrelevant
Lungen- erkrankungen	Lungenfunktionstests Blutgase	Häufigkeit der Anfälle Anfall-Medikationsverbrauch
Neurologische Erkrankungen	Evoziert potential MRT-Aufnahmen Elektrophysiologische Messungen (z. B. Elektronystagmografie Transkranielle Doppler-Untersuchung der intra- und extrakraniellen Arterien, Messung der Flussgeschwindigkeit der Basalarterien	Anzahl der Anfälle (z. B. Migräne, epileptische) Anzahl der Schübe Behinderungsgrad (z. B. EDSS) Symptomatik
Urologische Erkrankungen	Zystotonometrie Restharn Harnflussrate	Fähigkeit Wasserzulassen Miktionsfrequenz Lebensqualität Katheterisierung Inkontinenz
Wunden	Granulationsdynamik	Infektionsraten Dauer der Heilung Dauer Hospitalisierung
Zahnmedizin	Osteointegration des Implantats Implantatfestigkeit Sondierungstiefe Knochengewinn Winkelstellung/Normbisslage	Implantatverlust Überlebenszeit des Implantats Erhaltung der natürlichen Bezahnung

ACR = American college of rheumatology. CEA = Carcino-embryonales Antigen. CT = Computertomogramm. EDSS = Expanded disability status scale. HbA1C = Glykolisiertes Hämoglobin. HTA = Health Technology Assessment. IgE = Immunoglobulin E. LVEF = Linksventrikuläre Ejektionsfraktion. MRT = Magnetresonanztomografie. PASI = Psoriasis area and severity index. SCORAD = Scoring index of atopic dermatitis. UVB = Ultraviolett-B.

## 6.6 Diskussion

### 6.6.1 Diskussion der Methodik

Die Beantwortung der diesen Bericht leitenden Fragestellungen (siehe 6.3 „Forschungsfragen“) bedarf verschiedener methodischer Ansätze. Den ersten zwei Fragen wird im Rahmen eines Literaturreviews nachgegangen. Die Beantwortung der dritten Frage beruht auf der Analyse von methodischen Leitlinien von internationaler Institutionen, die im Bereich HTA tätig sind. Für die Beantwortung der vierten Fragestellung wird eine Stichprobe von HTA-Berichten aus internationalen HTA-Agenturen sowie aus Deutschland analysiert.

Die Durchführung von systematischen Übersichten zur Methodologie unterscheidet sich von der Durchführung eines anderen systematischen Reviews (z. B. zur Wirksamkeit von Interventionen) in mehreren Punkten. So stellt bereits die Konstruktion einer Recherchestrategie, die bei einer hohen Sensitivität eine noch handhabbare Trefferzahl liefern und eine akzeptable Spezifität aufweisen soll, eine Herausforderung dar. Zudem gibt es für methodische Fragestellungen keine standardisierten Recherchestrategien, wie sie für andere Reviewtypen zur Verfügung stehen (z. B. für therapeutische, diagnostische oder prognostische Studien), in die relevante Suchwörter einfach eingesetzt werden können. Weiterhin ist auch bei einer großen Anzahl von Treffern in der Recherche, die Sensitivität nicht gewährleistet (d. h. viele relevante Publikationen können übersehen werden), so dass die Durchsicht der Referenzlisten von potenziell relevanten Artikeln einmal mehr an Bedeutung gewinnt, und somit relevante Artikel identifiziert werden können, die durch die elektronische Recherche unentdeckt bleiben.<sup>82</sup> Ein weiteres Problem ist die Darstellung der Selektion der Literatur. Eine große Anzahl von Artikeln kann die Selektionskriterien erfüllen, da diese eher inhaltlicher Natur sind. Eine Begrenzung auf eindeutig zu definierende Studiendesigns (wie z. B. RCT bei systematischen Übersichten zu therapeutischen Interventionen) ist aufgrund der Fragestellung nicht möglich. Auch der Ausschluss bestimmter Publikationstypen, z. B. Editorials, Diskussionsbeiträge, Leserbriefe, ist riskant, da in diesen Artikeln relevante theoretische Beiträge oder Überlegungen zu finden sein können. Gleichzeitig ist die tatsächliche Relevanz von vielen Artikeln, die die inhaltlichen Auswahlkriterien

erfüllen, fraglich. Viele Artikel über Methoden, stellen keine methodischen Weiterentwicklungen dar<sup>82</sup>, sondern arbeiten z. B. die Thematik für eine andere Leserschaft auf, wobei die Argumente anderer Autoren wiederholt werden, die Beispiele ähnlich sind etc. Der Ausschluss dieser Arbeiten kann aufgrund der gewählten Ein- und Ausschlusskriterien für den Reviewer schwer zu begründen sowie für den Leser unter Umständen schwer nachzuvollziehen sein. Jedoch ist der Ausschluss von Arbeiten, die keine Weiterentwicklung darstellen, bzw. die frühere Überlegungen wiederholen, nicht mit einem relevanten Informationsverlust verbunden. Bei einem methodischen Review geht es darum, die verschiedenen Ansätze darzustellen, so dass redundante Arbeiten nichts Zusätzliches beitragen können. Das ist ein wichtiger Unterschied zu Übersichten von Studien, bei denen die Ergebnisse jeder Arbeit eine relevante Information darstellen. In diesem Teil des Berichts werden drei aktuelle Übersichtsarbeiten identifiziert, die neben den durch Recherche gefundenen Arbeiten auch weitere relevante Artikel zusammenfassen. Die Darstellungen in 6.5.1.2 „Qualitative Ergebnisse“ ermöglichen einen umfassenden Überblick zu Anforderungen an Surrogatparameter und zu Validierungsmethoden. Die Identifizierung von methodischen Leitlinien aus HTA-Agenturen bzw. Agenturen, die in Kostenerstattungsentscheidungen für Arzneimittel involviert sind, basiert auf Recherchen in den jeweiligen Internetauftritten. Die wichtigste Limitation dieses Ansatzes ist, dass Dokumente, die nicht durch das Internet öffentlich zugänglich gemacht werden, nicht zu erfassen sind. Diese hätten möglicherweise durch direkte Nachfragen bei den Agenturen, die in ihrem Webauftritt keine methodischen Papiere aufweisen, identifiziert werden können. Allerdings stellt sich angesichts der Homogenität der Aussagen in den öffentlichen Dokumenten, die Frage, ob die Identifizierung von weiteren methodischen Leitlinien, die Ergebnisse unserer Dokumentenanalyse verändert hätten. Weiterhin ist die Auswahl der Dokumente auf fünf Sprachen beschränkt, so dass nur ein Dokument aufgrund der Sprache ausgeschlossen wird (Methoden der norwegischen HTA-Agentur). Diese Ausschlüsse stellen die Validität der Aussagen in 6.5.2.2 „Qualitative Ergebnisse – Aussagen zu Surrogaten in Methodenpapieren“ jedoch nicht in Frage. Wie schon erwähnt, sind die Aussagen der Agenturen so homogen, dass eine Veränderung der Einschätzung durch den möglichen Einschluss eines weiteren Methodenpapiers nicht zu erwarten ist. Auch wenn man annehmen würde, dass das ausgeschlossene Methodenpapier aus Norwegen sich für den Einsatz von Surrogatendpunkten in der Bewertung von medizinischen Technologien ausgesprochen hätte, würde dies die Gesamteinschätzung nicht ändern. Hier wird also ein repräsentativer Überblick über den Stellenwert, den internationale HTA-Organisationen Surrogatendpunkten einräumen, gegeben.

Um der Frage nachzugehen, ob Surrogatendpunkte bei internationalen HTA-Berichten verwendet werden, wird eine Stichprobe aus der HTA-Datenbank gezogen. Die Größe der Stichprobe wird aus pragmatischen Gründen auf 5 % der Grundgesamtheit festgelegt, um die Anzahl der zu untersuchenden Berichte im Rahmen des innerhalb dieses Projekts Machbaren zu halten. Insbesondere aufgrund der nicht-möglichen Zugänglichkeit einiger Berichte konnte am Ende 140 HTA-Berichte analysiert werden (effektiver Stichprobenumfang von 2,6 % der Grundgesamtheit), was die Präzision der Ergebnisse reduziert (Breite des 95 %-Konfidenzintervalls liegt bei ca. +/- 8 %-Punkten). Andere Surveys über HTA-Berichte haben zwischen 3,5 % und 7 % der in der INAHTA-Datenbank enthaltenden HTA-Berichte untersucht, wobei sie sich nur auf wenige HTA-Agenturen beschränken<sup>25, 26, 52</sup>. Unsere Stichprobe hingegen wird zufällig gezogen und hinsichtlich der Herkunft der Berichte und des Publikationsjahres stratifiziert. Bezüglich des Technologietypen war es jedoch nicht möglich die Stichprobenziehung zu stratifizieren, dies hätte die Durchsicht der Gesamtheit der Berichte erfordert. Die thematische Verteilung der hier untersuchten Stichprobe ist jedoch vergleichbar zu der des Berichtepools von anderen empirischen Untersuchungen zum Inhalt von HTA-Berichten<sup>25, 26</sup>. Während für die Untersuchung der Methodenpapiere sowohl HTA-Agenturen (Mitglied der INAHTA), als auch Agenturen, die in Kostenerstattungsentscheidungen involviert sind (sogenannte Fourth hurdle-Institutionen<sup>47</sup>), sind in die Analyse von HTA-Berichten nur Assessments der INAHTA-Datenbank eingegangen. Die Assessments aus Fourth hurdle-Institutionen sind meistens nicht öffentlich und werden auch nicht in einer gemeinsamen Datenbank gesammelt. Eine Ausnahme sind die Berichte des englischen NICE, die in die INAHTA-Datenbank eingeschlossen sind. Angesichts der methodischen Empfehlungen dieser Institutionen zum Umgang mit Surrogaten, ist es jedoch nicht zu erwarten, dass in diesen von uns nicht-berücksichtigten Assessments Surrogatendpunkte stärker herange-

zogen werden. Der Einschluss der Assessments von Fourth hurdle-Institutionen hätte deshalb die Gesamtaussagen sehr wahrscheinlich nicht verändern können.

Im Bezug auf die Analyse der deutschen HTA-Berichte ist zu erwähnen, dass die Datenbank der DAHTA beim DIMDI nicht alle vom G-BA herausgegebenen Berichte erfasst. Neuere G-BA-Berichte, also solche, die nach 2003 veröffentlicht werden, sind nicht in dieser Datenbank. Wieder stellt sich die Frage, ob der Einschluss dieser Berichte (ca. 30) zu anderen Ergebnissen der Analyse geführt hätte. Angesichts der eingeschlossenen Berichte dieser Institution, ist dies auch hier nicht zu erwarten. Nach 2003 sind keine relevanten Änderungen in dem Umgang mit Surrogatendpunkten in dieser Institution zu erwarten, da es hinsichtlich der Verfahren zur Sammlung und Bewertung von Evidenz keine relevanten Änderungen gegeben hat.

Die Verwendung von Surrogatendpunkten in HTA-Berichten wird in zwei Schritten untersucht. Zunächst wird in den Fragestellungen bzw. Methodenteilen der Berichte die Festlegung der in den Berichten verwendeten Zielparameter überprüft. Wenn dort keine detaillierte Information zu den verwendeten Endpunkten gefunden wird, werden die Ergebnis-Extraktionstabellen durchsucht, und die dort berichteten Zielgrößen extrahiert. Überraschenderweise werden die verwendeten Zielgrößen nur in ca. der Hälfte der Berichte in den methodischen Sektionen benannt. Das heißt jedoch nicht unbedingt, dass in der anderen Hälfte keine prospektive Festlegung der Assessmentparameter stattgefunden hätte. Es ist nicht auszuschließen, dass dies z. B. in nicht-veröffentlichten Protokollen bzw. Projektplänen geschehen ist, ohne dass diese wichtige Information in die Berichte eingeflossen ist. Auch nicht auszuschließen ist, dass in diesen Fällen die Verwendung von Surrogatendpunkten festgelegt wird, diese jedoch nicht in den Ergebnistabellen extrahiert werden, weil sie bereits nicht in den eingeschlossenen Primärstudien berichtet werden. In diesen Fällen wären diese Berichte fälschlicherweise in der Kategorie „keine Verwendung von Surrogatendpunkten“ klassifiziert. Wir gehen davon aus, dass diese Situation sehr selten auftritt. Es würde außerdem zu keiner bedeutenden Verzerrung der Einschätzung des Stellenwerts von Surrogatendpunkten führen. Wenn bereits die Primärstudien keine Surrogatendpunkte verwenden, bedeutet das, dass sie für die Untersuchung der spezifischen Technologie bzw. Indikation keine große Bedeutung haben.

Unser Ansatz ermöglicht uns deshalb eine Abschätzung des Stellenwerts von Surrogatendpunkten in HTA-Berichten auf Basis einer quantitativen Analyse vorzunehmen.

## 6.6.2 Diskussion und Beantwortung der Forschungsfragen

### 6.6.2.1 Welche Kriterien muss ein Surrogatparameter erfüllen, um als valider Endpunkt angesehen werden zu können?

Damit ein Surrogatendpunkt als valider Endpunkt für die Bewertung der Nutzen einer Technologie berücksichtigt werden kann, müsste er mehrere Kriterien erfüllen.

- Die Beziehung zwischen Surrogatendpunkt und dem klinisch relevanten Nutzen muss **biologisch plausibel** sein. Es muss Evidenz aus tierexperimentellen oder epidemiologischen Studien vorliegen, die auf einen kausalen Zusammenhang zwischen dem Surrogatparameter und dem klinischen Endpunkt hinweisen. Der Surrogatparameter liegt demnach auf dem pathophysiologischen Kausalpfad zum klinisch relevanten Endpunkt.
- Die Beziehung zwischen Surrogatendpunkt und klinischem Endpunkt muss konsistent sein. Epidemiologische Studien müssen wiederholt zeigen, dass durch eine Veränderung des Surrogatendpunkts eine qualitativ und quantitativ gleichgerichtete Veränderung des klinisch relevanten Endpunkts erreicht werden kann.
- Es muss ein **Wirkungsnachweis aus RCT vorliegen**. Es gibt Ergebnisse aus RCT, die zeigen, dass durch eine therapiebedingte Veränderung des Surrogatendpunkts eine gleichgerichtete Veränderung des klinisch relevanten Endpunkts erreicht werden kann. Der Surrogatendpunkt muss den Effekt der Intervention vollständig abbilden. Dabei ist zu beachten, dass es nicht nur bei unterschiedlichen Technologietypen und pharmakologisch unterschiedlichen Wirkstoffen, sondern auch bei Wirkstoffen der gleichen Wirkstoffgruppe zu abweichenden Wirkmechanismen kommen kann und daher die Übertragbarkeit der Studienergebnisse von einer Substanz auf eine andere zusätzlich geprüft werden muss.

Darüber hinaus sind unabhängig von der Validität auch rein pragmatische Anforderungen an einen Surrogatendpunkt zu stellen, damit ihre Anwendung anstatt der endgültigen Endpunkte gerechtfertigt werden kann. Der Surrogatendpunkt sollte eine schnellere Bewertung der Technologie erlauben. Dafür ist es notwendig, dass Veränderungen in der Surrogatzielgröße früher und häufiger auftreten als bei dem klinisch relevanten Endpunkt. Zudem soll die Erhebung des Surrogatendpunkts einfach, reliabel und reproduzierbar<sup>12, 5</sup>, und somit auch für Therapieerfolgskontrollen im klinischen Alltag nützlich sein<sup>10</sup>.

### **6.6.2.2 Welche Methoden werden zur Validierung von Surrogatendpunkten in der Literatur diskutiert?**

In der Literatur werden verschiedene statistische Methoden zur Validierung von Surrogatendpunkten beschrieben, die in 6.5 „Ergebnisse“ ausführlich dargestellt sind.

Insgesamt kann festgehalten werden, dass es derzeit keinen Goldstandard für die Validierung von Surrogatendpunkten gibt. Aufgrund der Fehleranfälligkeit, die mit der Verallgemeinerung von Ergebnissen aus Einzelstudien verbunden ist, sind allerdings Ansätze auf Basis mehrerer Studien, wie dies z. B. bei Metaanalysen der Fall ist, zu bevorzugen. Dennoch gibt es bei keiner Validierungstechnik eine Garantie dafür, dass die ermittelten Zusammenhänge auch bei einer neuen Studie gleichermaßen gelten<sup>8</sup>. Aus diesem Grund reichen die Ergebnisse statistischer Validierungsverfahren allein nicht aus, um die Eignung eines Surrogatendpunkts zu beurteilen. Die Einbeziehung von Informationen über biologische und klinische Faktoren ist unverzichtbar. Die Validierung eines Surrogatendpunkts für die Bewertung des Nutzens ist in diesem Sinn technologiespezifisch. Es hängt von dem Wirkmechanismus der Technologie ab, ob der Surrogatendpunkt ihre Effekte auf den klinisch relevanten Endpunkt (z. B. Mortalität) vollständig abbilden kann oder nicht. Dies ist unabhängig davon, ob ein starker und konsistenter Zusammenhang zwischen den Veränderungen des Surrogatendpunkts und denen des klinisch relevanten Endpunkts statistisch nachgewiesen wird.

### **6.6.2.3 Welche methodischen Vorgaben machen internationale Agenturen aus dem Bereich HTA oder Arzneimittelnutzenbewertung hinsichtlich des Einsatzes von Surrogatendpunkten?**

Zusammenfassend zeigen die Empfehlungen der hier berücksichtigten Institutionen eine kritische Einstellung zur Verwendung von Surrogatendpunkten in HTA-Berichten. Klinisch relevante Endpunkte wie Mortalität oder Morbidität haben Vorrang, weil sie eine valide Beurteilung der Wirksamkeit und der Sicherheit ermöglichen, die die Grundlage jeglicher Nutzenbewertung darstellt. Es herrscht Übereinstimmung darüber, dass die Bewertung von Technologien nur in sehr gut begründeten Ausnahmefällen ausschließlich auf Basis von Surrogatendpunkten erfolgen darf. Die HTA-Institutionen zitieren wenige Beispiele von Surrogatendpunkten, die als validiert gelten können, wobei keine einen eindeutigen Katalog bzw. eine Liste von akzeptierten Surrogatendpunkten vorlegt. Die von den HTA-Institutionen gestellten Anforderungen sowie ihre Überlegungen zur Akzeptanz eines Surrogatendpunkts im Rahmen von HTA-Berichten werden durch Verweise auf die theoretische und methodische Literatur begründet und sind entsprechend mit dieser im Einklang.

### **6.6.2.4 Welchen Stellenwert haben Surrogatendpunkte in HTA-Berichten internationaler HTA-Agenturen und in Berichten, die in Deutschland produziert werden?**

Aus der quantitativen Analyse lässt sich feststellen, dass Surrogatendpunkten einen geringer Stellenwert in HTA-Berichten eingeräumt wird. In keinem der untersuchten HTA-Berichte wird die Bewertung von therapeutischen (bzw. nicht-diagnostischen) Technologien ausschließlich auf die Ergebnisse von Surrogatendpunkten gestützt. Alle Berichte legen in ihrer Methodik fest und/oder extrahieren immer klinisch relevante Parameter, die die Wirksamkeit und/oder die Sicherheit der Technologie erfassen. In ca. der Hälfte der Berichte werden zusätzlich Surrogatendpunkte betrachtet. Die Extraktion von

Surrogatendpunkten in HTA-Berichten ermöglicht ein vollständigeres Bild der Technologie und ihrer Wirkmechanismen zu konstruieren.

Die Bewertung von diagnostischen Technologien im Rahmen von HTA erfolgt nur in Ausnahmefällen ausschließlich auf der Basis von Testparametern. Dies kann valide sein in solchen Situationen, in denen die diagnostische Technologie einen klaren substitutiven Charakter hat (bei gleicher Sicherheit für Patienten und Anwender), so dass der Wirkmechanismus (d. h. ihre Platzierung in der Krankheits-Managementkette) der substituierten und der zu substituierenden Technologie die selben sind. Ansonsten wird gefordert, dass die Bewertung der klinischen Relevanz (z. B. Veränderungen bzgl. Überleben und Morbidität) Bestandteil des Assessments diagnostischer Technologien sein muss.

Die Ergebnisse der Analyse der deutschen HTA-Berichte sind annähernd identisch mit denen der Stichprobe internationaler HTA-Berichte. Es liegt kein Hinweis vor, dass in Deutschland ein strengerer Standard als in den Ländern des internationalen Umfelds beim Umgang mit Surrogatendpunkten praktiziert wird.

## **6.7 Schlussfolgerungen**

Die Validierung eines Surrogatendpunkts erfordert ausführliche Forschung, einschließlich der Durchführung von RCT mit klinisch relevanten Endpunkten. Die Validität eines Surrogatendpunkts ist eher technologie- als krankheitsspezifisch, so dass die Ergebnisse der Validierung eines Surrogatendpunkts nicht auf anderen Technologien übertragen werden können (auch nicht bei angeblich ähnlichem Wirkmechanismus). Um ein Höchstmaß an Sicherheit zu erreichen, muss die Validität eines Surrogatendpunkts bei jeder Technologie bzw. jedem Wirkstoff einzeln geprüft werden.

Nach wie vor ist der Einsatz von Surrogatendpunkten bei der Nutzenbewertung sehr kritisch zu betrachten.

## 7 Literatur

1. Alonso A, Molenberghs G, Burzykowski T, Renard D, Geys H, Shkedy Z et al. Prentice's approach and the meta-analytic paradigm: a reflection on the role of statistics in the evaluation of surrogate endpoints. *Biometrics* 2004; 60(3):724-728.
2. Alonso A, Molenberghs G, Geys H, Buyse M, Vangeneugden T. A unifying approach for surrogate marker validation based on Prentice's criteria. *Statistics in medicine* 2006; 25(2):205-221.
3. Antes G, Jöckel KH, Kohlmann T, Raspe H, Wasem J. Kommentierende Synopse der Fachpositionen zur Kosten-Nutzenbewertung für Arzneimittel. Erstellt im Auftrag des Bundesministeriums für Gesundheit. Freiburg, Essen, Greifswald, Kiel 2007
4. Arbeitsausschuss „Ärztliche Behandlung“ des Bundesausschusses der Ärzte und Krankenkassen. Positronen-Emissions-Tomographie (PET); Positron emission tomography (PET). Siegburg 2002.
5. ASCO. Outcomes of cancer treatment for technology assessment and cancer treatment guidelines. American Society of Clinical Oncology. *Journal of Clinical Oncology* 1996; 14(2):671-679.
6. Australian Government – Department of Health and Aging. Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (Version 4.1). Canberra 2006
7. Baker SG. A simple meta-analytic approach for using a binary surrogate endpoint to predict the effect of intervention on true endpoint. *Biostatistics* 2006; 7(1): 58-70.
8. Baker SG. Surrogate endpoints: Wishful thinking or reality? *Journal of the National Cancer Institute* 2006; 98(8): 502-503.
9. Belgian Health Care Knowledge Centre. Recommandations provisoires pour les évaluations pharmaco-économiques en Belgique. Brussels 2006.
10. Biomarkers Definition Working Group. Biomarkers and surrogate endpoints: preferred definitions and conceptual framework. *Clinical Pharmacology and Therapeutics* 2001;69(3):89-95.
11. Böger RH. Endpunkte klinischer Studien: Surrogatparameter oder „harte klinische Endpunkte“? *Internist* 2002; 43(4): 493-497.
12. Boissel JP, Collet JP, Moleur P, Haugh M. Surrogate endpoints: a basis for a rational approach. *European journal of clinical pharmacology* 1992; 43(3): 235-244.
13. Bruzzi P, del Mastro L, Sormani MP, Bastholt L, Danova M, Focan C, Fountzilias G, Paul J, Rosso R, Venturini M. Objective response to chemotherapy as a potential surrogate end point of survival in metastatic breast cancer patients. *Journal of Clinical Oncology* 2005; 23(22): 5117-5125.
14. Bucher HC, Guyatt GH, Cook DJ, Holbrook A, McAlister FA. Users' guides to the medical literature: XIX. Applying clinical trial results. A. How to use an article measuring the effect of an intervention on surrogate end points. Evidence-Based Medicine Working Group. *Journal of the American Medical Association* 1999; 282(8): 771-778.
15. Bucher HC, Griffith LE, Guyatt GH. Systematic review of risk and benefit of different cholesterol lowering interventions. *Arteriosclerosis, Thrombosis and Vascular Biology* 1999; 19: 187-195.
16. Burzykowski T, Buyse M. Surrogate threshold effect: an alternative measure for meta-analytic surrogate endpoint validation. *Pharmaceutical Statistics* 2006; 5(3):173-186.
17. Burzykowski T, Molenberghs G, Buyse M. The validation of surrogate end points by using data from randomized clinical trials: a case-study in advanced colorectal cancer. *Journal of the Royal Statistical Society Series A-Statistic in Society* 2004; 167(P1): 103-124.

18. Burzykowski T, Molenberghs G. Validation of surrogate end points in multiple randomized clinical trials with failure time end points. *Journal of the Royal Statistical Society Series C-Applied Statistics* 2001; 50(P4): 405-422.
19. Buyse M, Molenberghs G, Burzykowski T. Erratum: Criteria for the validation of surrogate endpoints in randomized experiments (*Biometrics* (September 1998) 54 (1014-1029)). *Biometrics* 2000; 56(1):324.
20. Buyse M, Molenberghs G. Criteria for the validation of surrogate endpoints in randomized experiments. *Biometrics* 1998; 54(3): 1014-1029.
21. Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. Statistical validation of surrogate endpoints: Problems and proposals. *Drug Information Journal* 2000; 34(2): 447-454.
22. Canadian Agency for Drugs and Technologies in Health. Guidelines for the economic evaluation of health technologies: Canada. Ottawa 2006.
23. Collette L, Burzykowski T, Carroll KJ, Newling D, Morris T, Schröder FH. Is prostate-specific antigen a valid surrogate end point for survival in hormonally treated patients with metastatic prostate cancer? Joint research of the European Organisation for Research and Treatment of Cancer, the Limburgs Universitair Centrum, and AstraZeneca Pharmaceuticals. *Journal of Clinical Oncology* 2005; 23(25): 6139-6148.
24. Daniels MJ, Hughes MD. Meta-analysis for the evaluation of potential surrogate markers. *Statistics in Medicine* 1997; 16: 1965-1982.
25. Draborg E, Gyrd-Hansen D, Poulsen PB, Horder M. International comparison of the definition and the practical application of health technology assessment. *International Journal of Technology Assessment in Health Care* 2005; 21(1): 89-95.
26. Draborg E, Kronborg-Andersen C. What influences the choice of assessment methods in health technology assessments? Statistical analysis of international health technology assessments from 1989 to 2002. *International Journal of Technology Assessment in Health Care* 2006; 22(1): 19-25.
27. Echt DS, Liebson PR, Mitchell LB, Peters RW, Obias-Manno D, Barker AH et al. Mortality and morbidity in patients receiving encainide, flecainide, or placebo. The Cardiac Arrhythmia Suppression Trial. *N Engl J Med* 1991; 324(12):781-788.
28. Ellenberg SS. Surrogate endpoints: the debate goes on. *Pharmacoepidemiology and Drug Safety* 2001; 10: 493-496.
29. Ellenberg SS, Hamilton JM. Surrogate endpoints in clinical trials: cancer. *Statistics in Medicine* 1989;8: 405-413.
30. Fleming TR, de Mets DL. Surrogate end points in clinical trials: are we being misled? *Annals of Internal Medicine* 1996; 125(7):605-613.
31. Food and Drug Administration. New drug, antibiotic and biological drug product regulations: accelerated approval. Proposal Rule. 57 Federal Register 13234-13242. Washington: 1992.
32. Food and Drug Administration. Guidance for Industry: Providing Clinical Evidence of Effectiveness for Human Drugs and Biological Products. Rockville 1998.
33. Freedman LS, Graubard BI, Schatzkin A. Statistical validation of intermediate endpoints for chronic diseases. *Statistics in medicine* 1992; 11(2): 167-178.
34. Fryback DG, Thornbury JR. The efficacy of diagnostic imaging. *Medical Decision Making* 1991; 11: 88-94.
35. Gail MH, Pfeiffer R, van Houweliagen HC, Carroll RJ. On meta-analytic assessment of surrogate outcomes. *Biostatistics* 2000; 1: 231-246.
36. Gemeinsamer Bundesausschuss. Verfahrensordnung des Gemeinsamen Bundesausschusses vom 20. September 2005, zuletzt geändert am 18. April 2006. Siegburg 2006.

37. Gotzsche PC, Liberati A, Torri V, Rossetti L. Beware of surrogate outcome measures. *International Journal Technology Assessment in Health Care* 1996; 12(2):238-246.
38. Grimes DA, Schulz KF. Surrogate end points in clinical research: hazardous to your health. *Obstetrics and Gynecology* 2005; 105(5 Pt 1):1114-1118.
39. Health Care Insurance Board. Dutch Guidelines for pharmacoeconomic research. Amstelveen 1999.
40. Heintze C, Velasco Garrido M, Kröger A. What do community-based dengue control programmes achieve? A systematic review of published evaluations. *Transactions of the Royal Society of Tropical Medicine and Hygiene* 2007; 101(4): 317-325.
41. Hillis A, Siegel D. Surrogate endpoints in clinical trials: ophthalmologic disorders. *Statistics in Medicine* 1989; 8: 427-430.
42. Hjalmarson A, Goldstein S, Fagerberg B, Wedel H, Waagstein F, Kjekshus J et al. Effects of controlled-release metoprolol on total mortality, hospitalizations, and well-being in patients with heart failure: the Metoprolol CR/XL Randomized Intervention Trial in congestive heart failure (MERIT-HF). MERIT-HF Study Group. *Journal of the American Medical Association* 2000; 283(10): 1295-1302.
43. Hofmann B. Is there a technological imperative in health care? *International Journal Technology Assessment in Health Care* 2002; 18(3): 675-689.
44. HTAi Policy Forum Working Group on Surrogate Outcomes (2008). Call for expressions of interest and comments, September 2008  
[www.htai.org/index.php?id=265](http://www.htai.org/index.php?id=265); 15.01.2009
45. Hughes MD. Evaluating surrogate endpoints. *Controlled Clinical Trials* 2002; 23(6): 703-707.
46. Hughes MD, Daniels MJ, Fischl MA, Kim S, Schooley RT. CD4 cell count as a surrogate endpoint in HIV clinical trials: a meta-analysis of studies of the AIDS Clinical Trials Group. *AIDS* 1998; 12: 1823-1832.
47. Hutton J, McGrath C, Frybourg JM et al. Framework for describing and classifying decision-making systems using technology assessment to determine the reimbursement of health technologies (fourth hurdle systems). *International Journal Technology Assessment in Health Care* 2006; 22:10-18.
48. Hutton J, Trueman P, Henshall C. Coverage with Evidence Development: an examination of conceptual and policy issues. *International Journal Technology Assessment in Health Care* 2007; 23: 425-435.
49. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonized Tripartite Guideline. General Considerations for Clinical Trials E8. 1997.
50. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonized Tripartite Guideline. Statistical Principles for Clinical Trials E9. 1998.
51. Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. Allgemeine Methoden. Version 3.0 vom 27.05.2008. Köln 2008.
52. Lavis JN, Wilson MG, Grimshaw J, Haynes RB, Ouimet M, Raina P, Gruen R, Graham I. Towards optimally packaged and relevance assessed health technology assessments. Report submitted to the Canadian Agency for Drugs and Therapeutics in Healthcare. Hamilton 2007.
53. Li Z, Meredith MP, Hoseyni MS. A method to assess the proportion of treatment effect explained by a surrogate endpoint. *Statistics in Medicine* 2001; 20(21): 3175-3188.
54. Lin DY, Fleming TR, de Gruttola V. Estimating the proportion of treatment effect explained by a surrogate marker. *Statistics in Medicine* 1997; 16: 1515-1527.
55. Lonn E. The use of surrogate endpoints in clinical trials: focus on clinical trials in cardiovascular diseases. *Pharmacoeconomics and Drug Safety* 2001; 10(6): 497-508.

56. Medical Services Advisory Committee. Funding for new medical technologies and procedures: application and assessment guidelines. Canberra 2005.
57. Molenberghs G, Burzykowski T, Alonso A, Buyse M. A perspective on surrogate endpoints in controlled clinical trials. *Statistical Methods in Medical Research* 2004; 13(3): 177-206.
58. Molenberghs G, Geys H, Buyse M. Evaluation of surrogate endpoints in randomized experiments with mixed discrete and continuous outcomes. *Statistics in medicine* 2001; 20(20): 3023-3038.
59. Molenberghs G, Buyse M, Geys H, Renard D, Burzykowski T, Alonso A. Statistical challenges in the evaluation of surrogate endpoints in randomized trials. *Controlled Clinical Trials* 2002; 23(6): 607-625.
60. Mühlhauser I, Berger M. Surrogat-Marker: Trugschlüsse. *Deutsches Ärzteblatt* 1996; 93(49): A3280-A3283.
61. Mühlhauser I, Meyer G. Evidence Based Medicine: Widersprüche zwischen Surrogat-ergebnissen und klinischen Endpunkten. *Psychotherapie Psychosomatik Medizinische Psychologie* 2006; 56: 193-201.
62. National Institute of Health and Clinical Excellence. Guidance for manufacturers and sponsors. London 2001.
63. National Institute for Health and Clinical Excellence. Peginterferon alfa and ribavirin for the treatment of mild chronic hepatitis C. London 2006.
64. Norris SL, Nichols PJ, Caspersen CJ, Glasgow RE, Engelgau MM, Jack L, Isham G, Snyder SR, Carande-Kulis VG, Garfield S, Briss P, McCulloch D, and the Task Force on Community Preventive Services. The effectiveness of disease and case management for people with diabetes. A systematic review. *American Journal of Preventive Medicine* 2002; 22(4S): 15-38.
65. Norwegian Medicines Agency. Norwegian guidelines for pharmacoeconomic analysis in connection with applications for reimbursement. Oslo 2005.
66. Packer M, Rouleau J, Swedberg K, Pitt B, Fisher L, Klepper M, et al. Effect of Flosequinan on survival in chronic heart failure: preliminary results of the PROFILE study [Abstract]. *Circulation*. 1993; 88(Suppl 1): 301.
67. Persönliche Kommunikation von Pharmaceutical Benefits Advisory Committee.
68. Pharmaceutical Benefits Board. General guidelines for economic evaluations from the pharmaceutical benefits board. Stockholm 2003.
69. Pharmaceutical Management Agency. Recommended methods to derive clinical inputs for proposals to PHARMAC. Christchurch 2005.
70. Prentice RL. Surrogate endpoints in clinical trials: definition and operational criteria. *Statistics in Medicine* 1989; 8(4):431-440.
71. Psaty BM, Weiss NS, Furberg CD, Koepsell TD, Siscovick DS, Rosendaal FR, Smith NL, Heckbert SR, Kaplan RC, Lin D, Fleming TR, Wagner EH. Surrogate end points , health outcomes and the drug-approval process for the treatment of risk factors for cardiovascular disease. *Journal of the American Medical Association* 1999; 282(8): 786-790.
72. Renard D, Geys H, Molenberghs G, Burzykowski T, Buyse M. Validation of surrogate endpoints in multiple randomized clinical trials with discrete outcomes. *Biometrical Journal* 2002; 44(N8): 921-935.
73. Riggs BL, Hodgson SF, O'Fallon WM, Chao EY, Wahner HW, Muhs JM, Cedel SL, Melton LJ. Effect of fluoride treatment on the fracture rate in postmenopausal women with osteoporosis. *New England Journal of Medicine* 1990; 322(12): 802-809.
74. Rosenson RS, Tangney CC. Antiatherothrombotic properties of statins: implications for cardiovascular event reduction. *Journal of the American Medical Association* 1998; 279: 1643-1650.

75. Royal ST, Kendrick D, Coleman T. Non-legislative interventions for the promotion of cycle helmet wearing by children. *Cochrane Database of Systematic Reviews* 2005; 2: CD003985.
76. Sarria-Santamera A, Timoner-Aguilera J. Assessment of the effectiveness of health promotion and disease prevention technologies. Madrid 2002.
77. Schatzkin A, Freedman LS, Dorgan J, MacShane LM, Schiffman MH, Dawswey SM. Surrogate end points in cancer research: a critique. *Cancer Epidemiology, Biomarkers & Prevention* 1996; 5(12): 947-953.
78. Schmidt J. Risiko und Nutzen einer Behandlung: Das Beispiel Clofibrat. *Schweizerisch medizinische Wochenschrift* 1987; 117(20): 773-776.
79. Shepherd J, Jones J, Hartwell D, et al. Interferon alfa (pegylated and nonpegylated) and ribavirin for the treatment of mild chronic hepatitis C – a systematic review and economic evaluation, October 2005.
80. Siebert U, Sroczynski G und die German Hepatitis Model Group und der HTA-Expertenpanesl für Hepatitis C. Antivirale Therapie bei Patienten mit chronischer Hepatitis C in Deutschland. Medizinische und ökonomische Evaluation der initialen Kombinations-therapie mit Interferon/Peginterferon und Ribavirin. Köln 2003.
81. Strippoli GFM, Navaneethan SD, Craig JC. Haemoglobin and haematocrit targets for the anaemia of chronic kidney disease. *Cochrane Database of Systematic Reviews* 2006, Issue 4. Art. No.: CD003967. DOI: 10.1002/14651858.CD003967. pub2.
82. Sutton AJ, Abrams KR, Jones DR, Sheldon TA, Song F. Systematic Review of trials and other studies. *Health Technology Assessment* 1998; 2(19): 3-5.
83. Taylor JM, Wang Y. Surrogate markers and joint models for longitudinal and survival data. *Controlled Clinical Trials* 2002; 23(6): 626-634.
84. Temple R. Are surrogate markers adequate to assess cardiovascular disease drugs. *Journal of the American Medical Association* 1999; 282: 790-795.
85. The Albumin Reviewers (Alderson P, Bunn F, Li Wan Po A, Li L, Roberts I, Schierhout G). Human albumin solution for resuscitation and volume expansion in critically ill patients *Cochrane Database of Systematic Reviews* 2004, Issue 4. Art. No.: CD001208. DOI:10.1002/14651858.CD001208.pub2.
86. The BEST Steering Committee. Design of the Beta-Blocker Evaluation Survival Trial (BEST). *American Journal of Cardiology* 1995; 75(17): 1220-1223.
87. The Cardiac Insufficiency Bisoprolol Study II (CIBIS-II): a randomised trial. *The Lancet* 1999; 353(9146): 9-13.
88. Velasco Garrido M, Perleth M, Drummond M, Gürtner F, Jorgensen T, Jovell A, Malone J, Rütther A, Wild C. Best practice in undertaking and reporting health technology assessments. *International Journal of Health Technology Assessmen in Health Care* 2002; 18(2): 361-422.
89. Wang Y, Taylor JM. A measure of the proportion of treatment effect explained by a surrogate marker. *Biometrics* 2002; 58(4): 803-812.
90. Weir CJ, Walley RJ. Statistical evaluation of biomarkers as surrogate endpoints: a literature review. *Statistics in Medicine* 2006; 25(2): 183-203.
91. Xu J, Zeger SL. The evaluation of multiple surrogate endpoints. *Biometrics* 2001; 57(1): 81-87.
92. Zaza S, Sleet DA, Thompson RS, Sosin DM, Bolen JC, and the Task Force on Community Preventive Services. Reviews of Evidence Regarding Interventions to Increase Use of Child Safety is Seats. *American Journal of Preventive Medicine* 2001; 21(4S): 31-47.
93. Zentner A, Velasco Garrido M, Busse R. Methoden zu vergleichenden Bewertung pharmazeutischer Produkte. Eine internationale Bestehsaufnahme zur Arzneimittel-evaluation. Köln 2005.

## 8 Anhang

### 8.1 Dokumentation der Literaturrecherche

Die Recherche wird von der DAHTA beim DIMDI am 07.03.2007 durchgeführt und ist hier dokumentiert.

Suchschritt	Angabe	Treffer
1	INAHTA; DAHTA; NHSEED; CDAR94; CDSR93; ME90; EM90; CB85; BA90; MK77; CCTR93; GA03; SM78; CV72; AZ72	33683776
2	CT=SURROGATE END#POINT#	69555
3	CT=ENDPOINT#, SURROGATE	69542
4	CT=END POINT#, SURROGATE	69542
5	CT=BIOLOGICAL MARKER#	91941
6	2 TO 5	91952
7	CT=VALIDATION	975
8	CT=STATISTICS	43623
9	CT=BIOMETRY	8319
10	CT=DECISION SUPPORT TECHNIQUES	7637
11	CT=ENDPOINT DETERMINATION	15869
12	CT=CAUSALITY	17605
13	7 TO 12	93213
14	6 AND 13	1218
15	14 AND PY>=1990	1217
16	15 AND LA=(GERM OR ENGL)	1193
17	check duplicates: unique in s=16	1109

### 8.2 Aus der Methodenübersicht ausgeschlossene Literatur

1. Aitio A, Anderson D, Blain P, Bond J, Buratti M, Calder I et al. Biomarkers and risk assessment: Concepts and principles. *Environmental Health Criteria* 1993; 99: 3-82.
2. Albert JM, Ioannidis JP, Reichelderfer P, Conway B, Coombs RW, Crane L et al. Statistical issues for HIV surrogate endpoints: point/counterpoint. An NIAID workshop. *Statistics in Medicine* 1998; 17(21): 2435-2462.
3. Allison MA, Wright CM. A comparison of HDL and LDL cholesterol for prevalent coronary calcification. *International journal of cardiology* 2004; 95(1): 55-60.
4. Alonso A, Geys H, Molenberghs G, Kenward MG, Vangeneugden T. Validation of surrogate markers in multiple randomized clinical trials with repeated measurements: canonical correlation approach. *Biometrics* 2004; 60(4): 845-853.
5. Biglan KM, Holloway RG. Surrogate endpoints in Parkinson's disease research. *Current neurology and neuroscience reports* 2003; 3(4): 314-320.
6. Boone CW, Kelloff GJ. Intraepithelial neoplasia, surrogate endpoint biomarkers, and cancer chemoprevention. *Journal of Cellular Biochemistry Supplement* 1993; 17F:37-48.
7. Brown ER, Ibrahim JG, DeGruttola V. A flexible B-spline model for multiple longitudinal biomarkers and survival. *Biometrics* 2005; 61(1): 64-73.
8. Bruzzi P, del Mastro L, Sormani MP, Bastholt L, Danova M, Focan C, Fountzilias G, Paul J, Rosso R, Venturini M. Objective response to chemotherapy as a potential surrogate end point of survival in metastatic breast cancer patients. *Journal of Clinical Oncology* 2005; 23(22): 5117-5125.
9. Carter BL. Blood pressure as a surrogate end point for hypertension. *The Annals of Pharmacotherapy* 2002; 36(1): 87-92.

10. Chen SX, Leung DHY, Qin J. Information recovery in a study with surrogate endpoints. *Journal of the American Statistical Association* 2003; 98(N464): 1052-1062.
11. Chuang-Stein C. Surrogate endpoints in AIDS drug development: Current status. *Drug Information Journal* 1998; 32(N2): 439-448.
12. Collette L, Burzykowski T, Carroll KJ, Newling D, Morris T, Schröder FH. Is prostate-specific antigen a valid surrogate end point for survival in hormonally treated patients with metastatic prostate cancer? Joint research of the European Organisation for Research and Treatment of Cancer, the Limburgs Universitair Centrum, and AstraZeneca Pharmaceuticals. *Journal of Clinical Oncology* 2005; 23(25): 6139-6148.
13. Cooper R, Kaanders JH. Biological surrogate end-points in cancer trials: potential uses, benefits and pitfalls. *European Journal of Cancer* 2005; 41(9): 1261-1266.
14. Daskalakis C, Shenassa E. Inference about mediators or surrogate endpoints. *American journal of epidemiology* 2006; 163(11, Suppl. S): S241. Ref ID: 3093
15. D'Amico AV, Moul JW, Carroll PR, Sun L, Lubeck D, Chen MH. Surrogate end point for prostate cancer-specific mortality after radical prostatectomy or radiation therapy. *Journal of the National Cancer Institute* 2003; 95(N18): 1376-1383.
16. D'Agostino RB, Campbell MJ, Greenhouse JB. Surrogate markers: Back to the future (special papers for the 25th Anniversary of Statistics in Medicine). *Statistics in Medicine* 2006; 25(2): 181-182.
17. De Milito A, Titanji K, Zazzi M. Surrogate markers as a guide to evaluate response to antiretroviral therapy. *Current Medicinal Chemistry* 2003; 10(5): 349-365.
18. Einspahr JG, Alberts DS, Gapstur SM, Bostick RM, Emerson SS, Gerner EW. Surrogate end-point biomarkers as measures of colon cancer risk and their use in cancer chemoprevention trials. *Cancer Epidemiology, Biomarkers & Prevention* 1997; 6(1):37-48.
19. Ellenberg SS. Surrogate endpoints: the debate goes on. *Pharmacoepidemiology and drug safety* 2001; 10(6): 493-496.
20. Flandre P, O'Quigley J. A two-stage procedure for survival studies with surrogate endpoints. *Biometrics* 1995; 51(3): 969-976.
21. Fleming TR. Surrogate end points in cardiovascular disease trials. *American Heart Journal* 2000; 139(4): S193-S196.
22. Fleming TR. Surrogate endpoints in clinical trials. *Drug Information Journal* 1996; 30(2):545-551.
23. Fleming TR. Objective response rate as a surrogate end point: a commentary. *Journal of Clinical Oncology* 2005; 23(22): 4845-4846.
24. Gelfand JM, Hoffstad O, Margolis DJ. Surrogate endpoints for the treatment of venous leg ulcers. *The Journal of investigative dermatology* 2002; 119(6): 1420-1425.
25. Gensini GF, Conti AA. The evaluation of the results of clinical trials: surrogate end points and composite end points. *Minerva medica* 2004; 95(1):71-75.
26. Gill S, Sargent D. End points for adjuvant therapy trials: has the time come to accept disease-free survival as a surrogate end point for overall survival? *The Oncologist* 2006; 11(6): 624-629.
27. Grimes DA, Schulz KF. Surrogate end points in clinical research: hazardous to your health. *Obstetrics and Gynecology* 2005; 105(5 Pt 1): 1114-1118.
28. Hilsenbeck SG, Clark GM. Surrogate endpoints in chemoprevention of breast cancer: guidelines for evaluation of new biomarkers. *Journal of Cellular Biochemistry* 1993; 17G: 205-211.
29. Hohlfeld R, Montalban X. Surrogate markers of clinical disease in MS. *International MS Journal* 2002; 9(3): 79-80.
30. Hulka BS, Margolin BH. Methodological issues in epidemiologic studies using biologic markers. *American journal of epidemiology* 1992; 135(2): 200-209.

31. Johnston K. What are surrogate outcome measures and why do they fail in clinical research? *Neuroepidemiology* 1999; 18(4): 167-173.
32. Jones PW, Agusti AG. Outcomes and markers in the assessment of chronic obstructive pulmonary disease. *The European Respiratory Journal* 2006; 27(4): 822-832.
33. Kawut SM, Palevsky HI. Surrogate end points for pulmonary arterial hypertension. *American heart journal* 2004; 148(4): 559-565.
34. Kelloff GJ, O'Shaughnessy JA, Gordon GB, Hawk ET, Sigman CC. Counterpoint: Because some surrogate end point biomarkers measure the neoplastic process they will have high utility in the development of cancer chemopreventive agents against sporadic cancers. *Cancer Epidemiology Biomarkers and Prevention* 2003; 12(7): 593-596.
35. Kleist P. The impact of biomarkers and surrogate end points on regulatory approval of new drugs. *Heart Drug* 2002; 2(5): 235-245.
36. Kluff C. Principles of use of surrogate markers and endpoints. *Maturitas* 2004; 47(4): 293-298.
37. Kramer BS, Wilentz J, Alexander D, Burklow J, Friedman LM, Hodes R et al. Getting it right: being smarter about clinical trials. *PLoS medicine* 2006; 3(6):e144.
38. Lachenbruch PA, Rosenberg AS, Bonvini E, Cavaillé-Coll MW, Colvin RB. Biomarkers and surrogate endpoints in renal transplantation: present status and considerations for clinical trial design. *American Journal of Transplantation* 2004; 4(4): 451-457.
39. Lagakos SW. Surrogate markers in AIDS clinical trials: conceptual basis, validation, and uncertainties. *Clinical infectious diseases: an official publication of the Infectious Diseases Society of America* 1993; 16 Suppl 1:S22-S25.
40. Lee CR, Adams KF, Jr., Patterson JH. Surrogate end points in heart failure. *The Annals of Pharmacotherapy* 2002; 36(3): 479-488.
41. Leung DHY. Statistical methods for clinical studies in the presence of surrogate end points. *Journal of the Royal Statistical Society Series A-Statistics in Society* 2001; 164(P3): 485-503.
42. Levenson VV. Biomarkers: diagnostic highlights and surrogate end points. Cambridge Healthtech Institute's biomarker series: biomarker validation: bringing discovery to the clinic & cancer biomarkers: from discovery to clinical practice. May 3-5, 2004, Philadelphia, Pennsylvania, USA. *Pharmacogenomics* 2004; 5(5):459-461.
43. Li Z, Chines AA, Meredith MP. Statistical validation of surrogate endpoints: is bone density a valid surrogate for fracture? *Journal of Musculoskeletal & Neuronal Interactions* 2004; 4(1): 64-74.
44. Lin DY, Fischl MA, Schoenfeld DA. Evaluating the role of CD4-lymphocyte counts as surrogate endpoints in human immunodeficiency virus clinical trials. *Statistics in medicine* 1993; 12(9): 835-842.
45. Lohmander LS, Eyre DR. From biomarker to surrogate outcome to osteoarthritis--what are the challenges? *The Journal of Rheumatology* 2005; 32(6): 1142-1143.
46. Lotze MT, Wang E, Marincola FM, Hanna N, Bugelski PJ, Burns CA et al. Workshop on cancer biometrics: Identifying biomarkers and surrogates of cancer in patients - A meeting held at the masur auditorium, national institutes of health. *Journal of Immunotherapy* 2005; 28(2): 79-119.
47. Manns B, Owen WF, Jr., Winkelmayr WC, Devereaux PJ, Tonelli M. Surrogate markers in clinical studies: problems solved or created? *American Journal of Kidney Diseases* 2006; 48(1):159-166.
48. Marcovina SM, Crea F, Pieri PL, Peracino AP. Integrated biochemical and bioimaging markers as surrogate end-points in cardiocerebrovascular, disease drug evaluation: Validation proposals. *Atherosclerosis Supplements* 2006; 7(N3): 480-481.
49. Mildvan D, Landay A, De G, V, Machado SG, Kagan J. An approach to the validation of markers for use in AIDS clinical trials. *Clinical Infectious Diseases* 1997; 24(5): 764-774.

50. Mühlhauser I, Meyer G. Evidence Based Medicine: Widersprüche zwischen Surrogatergebnissen und klinischen Endpunkten. *Psychotherapie, Psychosomatik, medizinische Psychologie* 2006; 56(5): 193-201.
51. Mutti A. Use of intermediate end-points to prevent long-term outcomes. *Toxicology Letters* 1995; 77(1-3): 121-125.
52. Orloff DG. Use of surrogate endpoints: a practical necessity in lipid-altering and antiatherosclerosis drug development. *The American Journal of Cardiology* 2001; 87(4A): 35A-41A.
53. Orloff DG, Stein E, Nissen S, Davidson MH, Bolton R. Use of surrogate endpoints: A practical necessity in lipid-altering and antiatherosclerosis drug development. *American Journal of Cardiology* 2001; 87(4): 35A-41A.
54. Pérez-Soler R. Can rash associated with HER1/EGFR inhibition be used as a marker of treatment outcome? *Oncology* 2003; 17(11 Suppl 12): 23-28.
55. Powers JH. Microbiologic surrogate end points in clinical trials of infectious diseases: example of acute otitis media trials. *Pharmacotherapy* 2005; 25(12 Pt 2): 109S-123S.
56. Richter WS. Imaging biomarkers as surrogate endpoints for drug development. *European Journal of Nuclear Medicine and Molecular Imaging* 2006; 33(Suppl 13): S6-S10.
57. Sacks SL, Griffiths PD, Corey L, Cohen C, Cunningham A, Dusheiko GM et al. Introduction: Is viral shedding a surrogate marker for transmission of genital herpes? *Antiviral research* 2004; 63(Suppl 1): S3-S9.
58. Sargent D. General and statistical hierarchy of appropriate biologic endpoints. *Oncology* 2006; 20(6 Suppl 5): 5-9.
59. Schatzkin A, Freedman LS, Dorgan J, Mcshane LM, Schiffman MH, Dawsey SM. Surrogate end points in cancer research: a critique. *Cancer Epidemiology, Biomarkers & Prevention* 1996; 5(12): 947-953.
60. Schatzkin A. Problems with using biomarkers as surrogate end points for cancer: a cautionary tale. *Recent results in cancer research. Fortschritte der Krebsforschung* 2005; 166: 89-98.
61. Schatzkin A, Freedman LS, Dorgan J, McShane L, Schiffman MH, Dawsey SM. Using and interpreting surrogate end-points in cancer research. *IARC Scientific Publications* 1997; 142: 265-271.
62. Schatzkin A. Intermediate markers as surrogate endpoints in cancer research. *Hematology/Oncology Clinics of North America* 2000; 14(4): 887-905.
63. Schulte PA. Biomarkers in epidemiology: Scientific issues and ethical implications. *Environmental health perspectives* 1992; 98: 143-147.
64. Sen PK. Incomplete multiresponse designs and surrogate end-points in clinical trials. *Journal of Statistical Planning and Inference* 1994; 42(N1-2): 161-186.
65. Somberg J. Surrogate endpoints and drug approval. *American Journal of Therapeutics* 2006; 13(5):388.
66. Somberg J. Biomarker and surrogate endpoints. *American journal of therapeutics* 2003; 10(4):239-240.
67. Staner L. Surrogate outcomes in neurology, psychiatry, and psychopharmacology. *Dialogues in Clinical Neuroscience* 2006; 8(3): 345-352.
68. Tao Y, Klause A, Vickers A, Bae K, Ellis M. Clinical and biomarker endpoint analysis in neoadjuvant endocrine therapy trials. *Journal of Steroid Biochemistry and Molecular Biology* 2005; 95(1-5): 91-95.
69. Thompson DF. Surrogate end points, skepticism, and the CAST study. *The Annals of pharmacotherapy* 2002; 36(1): 170-171.
70. Trock BJ. Validation of surrogate endpoint biomarkers in prostate cancer chemoprevention trials. *Urology* 2001; 57(4 Suppl 1): 241-247.

71. Valicenti R, DeSilvio M, Hanks G, Porter A, Brereton H, Shipley W et al. Surrogate endpoint for prostate cancer-specific survival: Validation from an analysis of Radiation Therapy Oncology Group Protocol 92-02. *Journal of Clinical Oncology* 2005; 23(N16,S,P1): 390S.
72. Vellas B, Andrieu S, Sampaio C, Wilcock G. Disease-modifying trials in Alzheimer's disease: a European task force consensus. *Lancet Neurology* 2007; 6(1): 56-62.
73. Wagner JA, Williams SA, Webster CJ. Biomarkers and surrogate end points for fit-for-purpose development and regulatory evaluation of new drugs. *Clinical pharmacology and therapeutics* 2007; 81(1): 104-107.
74. Wagner JA. Overview of biomarkers and surrogate endpoints in drug development. *Disease Markers* 2002; 18(2): 41-46.
75. Zirrgiebel U, Drevs J. Surrogate marker in clinical studies with anti-angiogenic drugs. *International Journal of Clinical Pharmacology and Therapeutics* 2003; 41(12): 600-602.

### 8.3 Dokumentation der Methodenpapiere aus HTA-Institutionen

Tabelle 22: Aussagen zu Surrogaten in Methodenpapieren der Mitgliedinsitutionen von INAHTA

Land	Agentur	„Surrogate“ im Suchmotor	Methodenpapier	Aussage zur Auswahl der Endpunkten im Allgemeinen und zu Surrogaten im Besonderen
AR	IECS	Nein	Nein	-
AT	LBI@HTA	Nein	Externes Manual – Selbstverständnis und Arbeitsweise (2007)	<b>S. 25</b> Nutzen definiert sich durch die Relevanz für den Patienten Die Perspektive des Nutzens orientiert sich an der Relevanz für den Patienten. Dazu gehören Mortalität Morbidität (Beschwerden und Komplikationen) Gesundheitsbezogene Lebensqualität Interventions- und erkrankungsbezogener Aufwand Patientenzufriedenheit
AU	AHTA	Nein	Nein	-
	ASERNIP-S	Nein	General Guidelines for Assessing, Approving & Introducing New Procedures into a Hospital or Health Service (Jahr: keine Angaben)	Keine Angaben
			ASERNIP-S Systematic Review Process (2003)	Keine Angaben
	MSAC	Nein	Funding for new medical technologies and procedures: application and assessment guidelines (2005)	<b>S. 43</b> The types of outcomes that are measured and reported in a clinical trial determine whether all of the benefits and harms associated with a treatment can be assessed. It may be useful to consider outcomes as being one of three possibly overlapping types. These are 'surrogate' outcomes, 'clinical' outcomes, and 'patient-relevant' outcomes. In general, clinical trials measure and report either surrogate or clinical outcomes, and these may or may not be of importance to the patient. Patient-relevant outcomes One of the difficulties in measuring patient-relevant outcomes is that they may be composite outcomes, and include items that are not readily amenable to objective measurement or quantification. They are likely to include a measure of a clinical effect, adverse effects, tolerability and change in quality of life, and these may need to be individualised for each treatment. The following types of outcomes could be considered: <ul style="list-style-type: none"> <li>• all-cause mortality;</li> <li>• cause-specific mortality;</li> <li>• changes in morbidity, subdivided according to type (eg hospital admission, nursing home requirements);</li> <li>• side effects of treatment, including adverse reactions to drug therapies; and</li> <li>• disease specific outcomes, including disease specific quality of life measures.</li> </ul>

				<p><b>S. 44</b> [...] The term 'surrogate outcome' has a variety of definitions, and other terms are often used in its place (eg intermediate outcome or biological marker). The common features of the definitions are:</p> <ul style="list-style-type: none"> <li>• that the surrogate outcome is commonly a physiological variable (eg serum cholesterol concentration, blood pressure);</li> <li>• there is a statistical association between the surrogate outcome and the clinical outcome of interest (eg bone mineral density and fracture, CD4 cell concentrations and progression of HIV); or</li> <li>• there is a biological and pathophysiological basis for believing that the surrogate outcome is a major determinant of the clinical outcome in the disease being studied (eg glycosylated haemoglobin measurements and diabetic complications).</li> </ul> <p>A surrogate outcome should possess all of these features, but experience suggests that few do. If you are using a study with surrogate outcomes as the basis for your application, it is important to present some evidence that it is an appropriate surrogate —that is, that there is an established relationship between the surrogate outcome and the clinical outcome of interest. Studies relying on surrogate outcomes alone are not likely to provide a sufficient basis on which to make a judgment about the true benefits of a technology.</p> <p><b>S. 48</b> [...] Identify and justify the outcome that best reflects the best comparative clinical performance of the alternatives, based on the results of the clinical trials. If the most robust data available relates to a surrogate outcome, ensure that it is adequately validated as a surrogate in relation to the final clinical or patient relevant outcome.</p>
			Guidelines for the assessment of diagnostic technologies (2005)	<p><b>S. 18</b> Consideration of the following sequence of questions may be helpful when developing the review protocol:</p> <p>[...] C. The impact of the test on health outcomes What is the causal pathway between the intended use of the index test and improved health outcomes?</p> <p><b>S. 63</b> (Economic evaluation): Surrogate outcomes eg improved hypertension control that can be clearly associated with patient relevant consequences may also be used.</p>
BE	KCE	Nein	Recommandations provisoires pour les évaluations pharmaco économiques en Belgique KCE (2006)	<p><b>S.16</b> [...] Models are used for different reasons: extension of time horizons, extrapolation of intermediate outcome parameters to final outcome parameters, consideration of externalities associated with a treatment, translation of foreign data to the Belgian context, pooling data from multiple trials./Another reason for modelling is the simulation of final outcomes based on observed data on intermediate outcomes. Often in clinical trials, only intermediate outcome measures are included (e. g. blood pressure reduction). Other studies may provide information on the relationship between the intermediate outcome measure and a final outcome measure (e. g. blood pressure and mortality).</p> <p><b>S.20</b> [...].Outcomes in pharmaco-economic evaluations should be expressed in terms of endpoints instead of intermediary outcomes. Clearly defined outcome measures, there is little debate about the measurement methods, are recommended.</p>

BR	DECIT-CGATS	Nein	Diretrizes metodologicas paea elaboração de estudos em avaliação de tenologias para o Ministerio de Saude (2007)	<p><b>S. 11</b> Alguns critérios de seleção que poderão ser utilizados são:                  [...] Comparadores (resultados em saúde): mortalidade, morbidade, incidência de complicações, qualidade de vida, etc.                  [Mögliche Selektionskriterien für die Literatur können sein: Zielparameter (Gesundheitsoutcomes): Mortalität, Morbidität, Komplikationsraten, Lebensqualität etc.]</p>
CA	AETMIS	Ja	Nein	Verweis auf Programm der HTAi-Tagung
	CADTH	Ja		Anzeige eines Projekts zu Evaluierung von Surrogatparameter in Rheumathoid Arthritis
			Guideline for the economic evaluation of health technologies: Canada (2006)	<p><b>S. 4</b> [...] Use a final outcome (e. g., life-years gained), or if that is impossible, an important patient outcome. Only use a surrogate outcome if it has a well established link (i. e., validated) with one of those outcomes.</p> <p><b>S. 6</b> [...] Where feasible and scientifically credible, translate efficacy data into the best quantitative estimate of effectiveness in the Reference Case, using the best available evidence and appropriate modelling techniques. This may involve linking surrogate outcomes to important patient outcomes or extrapolating data beyond the duration of the trial.</p> <p><b>S. 13</b> [...] A surrogate outcome is “a laboratory measurement or a physical sign used as a substitute for a clinically meaningful endpoint that measures directly how a patient feels, functions, or survives.” Examples for cardiovascular disease include blood pressure or cholesterol level.</p> <ul style="list-style-type: none"> <li>Validated surrogate outcomes are proven to be predictive of an important patient outcome. A surrogate outcome is valid only if there is a “strong, independent, consistent association” with an important patient outcome, and there is “evidence from randomized trials that...improvement in the surrogate end point has consistently lead to improvement in the target outcome. Unvalidated (unproven) surrogate outcomes have not been proven to be predictive of an important patient outcome.</li> </ul> <p>[...] In determining effectiveness, the evidence on final outcomes (e. g., life-years gained) is preferred to that of validated surrogate outcomes. Outcomes with less clear validation or relevance for patients should also be described, if relevant.</p> <p><b>S 14.</b> [...] It is preferred that the outcome measure be a final outcome (e. g., life-years), or if that is impossible, an important clinical outcome. In general, a surrogate outcome should only be used as an outcome measure if it has a validated, well established link with an important patient outcome. Such an analysis should be appropriately tested through a sensitivity analysis.</p> <p><b>S. 21</b>....Analysts are encouraged to use modelling to link surrogate outcomes to more important patient outcomes. For studies using surrogate outcomes, the surrogate should be highly predictive of an important patient outcome.</p>
			Procedure for Common Drug Review (2006)	Keine Angaben
			Guidelines for clinical reviewers common drug review process (2006)	Keine Angaben

			General guideline for reviewers common drug review process (2006)	Keine Angaben
	IHE	Nein	Nein	-
	MAS	Nein	Nein	-
CH	MTU-SFOPH	Nein	Nein	-
DE	DAHTA@DIMDI	Nein	Nein	-
	IQWiG	Nein	Methoden Version 3.0 (2008)	<p><b>S. 32-33</b> Es werden in erster Linie Endpunkte berücksichtigt, die zuverlässig und direkt konkrete Änderungen des Gesundheitszustands abbilden. Dabei werden insbesondere einzelne Betroffene, Patientenvertretungs- und/oder Verbraucherorganisationen in die themenbezogene Definition patientenrelevanter Endpunkte einbezogen. Zur Erfassung der Lebensqualität und der Patientenzufriedenheit sollen nur Instrumente eingesetzt werden, die für den Einsatz in klinischen Studien geeignet und entsprechend evaluiert sind. Darüber hinaus können für die Nutzenbewertung valide Surrogatendpunkte berücksichtigt werden.</p> <p><b>S. 33-34 3.1.2 Surrogate des patientenrelevanten medizinischen Nutzens</b> Surrogatendpunkte werden in der medizinischen Forschung häufig als Ersatz für patientenrelevante Endpunkte verwendet, meist um Aussagen über den patientenrelevanten Nutzen früher und einfacher zu erhalten. Die meisten Surrogatendpunkte sind jedoch in dieser Hinsicht nicht verlässlich und können bei der Nutzenbewertung irreführend sein. Surrogatendpunkte werden im Rahmen der Nutzenbewertung des Institutes in der Regel nur dann als Beleg des (Zusatz-)Nutzens einer Intervention in Betracht gezogen, wenn zuvor anhand geeigneter statistischer Methoden gezeigt wird, dass der Effekt von Interventionen mit vergleichbarem Wirkmechanismus auf den zu ersetzenden patientenrelevanten Endpunkt durch den Effekt auf den Surrogatendpunkt in einem ausreichenden Ausmaß erklärt wird. In der Regel sind dafür eindeutige Belege aus Interventionsstudien eines plausiblen, starken, konsistenten und gleichgerichteten Zusammenhangs zwischen der Änderung des Surrogats und der Änderung des patientenrelevanten Endpunktes erforderlich. Ein gleichgerichteter Zusammenhang bedeutet, dass eine positive bzw. negative Veränderung des Surrogates mit einer positiven bzw. negativen Veränderung des patientenrelevanten Endpunktes einhergeht. Die Validität eines Surrogates wird als nicht belegt betrachtet, wenn keine aussagekräftigen Studien zur Modifikation dieses Surrogates in Verbindung mit einer Beschreibung des entsprechenden patientenrelevanten Endpunktes vorliegen. Darüber hinaus wird ein Surrogat als nicht valide betrachtet, wenn in wissenschaftlichen Untersuchungen gezeigt wird, dass eine Intervention</p> <ul style="list-style-type: none"> <li>• einen Einfluss auf den Surrogatendpunkt hatte, nicht jedoch auf den patientenrelevanten Endpunkt, oder</li> <li>• einen Einfluss auf den patientenrelevanten Endpunkt hatte, nicht jedoch auf den Surrogatendpunkt, oder</li> <li>• inkonsistente Effekte beim Surrogat- und patientenrelevanten Endpunkt erzeugt.</li> </ul>

				<p>Beobachtete Zusammenhänge zwischen einem Surrogat- und dem entsprechenden patientenrelevanten Endpunkt für eine Intervention mit einem bestimmten Wirkmechanismus sind nicht zwangsläufig auf Interventionen zur Behandlung derselben Erkrankung, aber mit einem anderen Wirkmechanismus, anwendbar.</p> <p>Surrogatendpunkte von unklarer oder umstrittener Validität können zwar in den Berichten des Institutes dargestellt werden. Derartige Endpunkte sind aber nicht als Belege für den Nachweis des (Zusatz-) Nutzens einer Intervention geeignet. Bei besonders gravierenden Erkrankungen im Sinne von Morbidität und Mortalität ohne Therapiealternative werden Surrogatendpunkte unklarer Validität ggf. als Hinweis auf einen Nutzen akzeptiert werden müssen.</p> <p>Je nach Nähe zu einem entsprechenden patientenrelevanten Endpunkt werden in der Literatur verschiedene andere Bezeichnungen für Surrogatendpunkte verwendet (z. B. intermediärer Endpunkt). Auf eine diesbezügliche Differenzierung wird an dieser Stelle allerdings verzichtet, da die Frage nach der notwendigen Validität davon unberührt bleibt.</p> <p><b>S.42</b> (Nutzenbewertung in kleinen Populationen) [...]Ebenso kann bei kleineren Stichproben eher die Notwendigkeit bestehen, einen zu selten auftretenden patientenrelevanten Endpunkt durch Surrogate zu ersetzen. Solche Surrogate müssen allerdings auch bei kleinen Stichproben valide sein.</p>
DK	DACEHTA	Nein	Health Technology Assessment Handbook (2001)	<p><b>S. 109</b> The measure of effectiveness can be divided into two main groups – intermediate measures and final measures. <i>Intermediate measures of effectiveness</i> concerns changes in a health-related variable from the use of a health technology. Intermediate measures of effectiveness can be further divided into surrogate measures, e. g. reduction in mmHg blood pressure, per cent serum cholesterol; or into (avoided) events, e. g. number of correct diagnosed patients, number of heart attacks. The intermediate measures of effectiveness are close to or equal to the output measures chosen in the clinical studies, which is why they are often relevant for clinical decision-making. On the other hand are the <i>final measures of effectiveness</i> indicators of survival that reflect the likelihood or frequency of survival a in defined time interval. Examples of final measures of effectiveness are gained life years or lives saved. Final measures of effectiveness are related to the end result of using the technology and not just a clinical output, which is why they are more relevant for the patient and in priority setting in general. Using final measures of effectiveness makes comparison of different types of health technologies possible. Finally, the effectiveness can be covered by a health status measure or health status profile, if only health status, and not survival, is affected by the technology [...]</p>
	DSI	Nein	Nein	-

ES	AETS	Nein	Guía para la elaboración de informes de evaluación de tecnologías sanitarias (1999)	<p><b>S.19</b> The selection of the outcome parameter most appropriate for answering the research question is a difficult task which requires time and research.</p> <p>An HTA Report based on outcomes parameters describing health effects or utilities allows solid and less questionable conclusions. However when the clinical results are described using measures which are indirect of intermediate, such as changes in biochemical parameters, the real effectiveness of the technology can be questioned. It is important to be aware of the common utilisation of indirect evidence or surrogate end-points, which are usually physiological or biochemical parameters part of causal or association chain with clinical effects (see examples in Table 1) [Hier wird das Bsp. Antiarrhythmika benannt]</p> <p><b>S.42</b> Conclusions and recommendations are the most frequently read parts of a report – together with the summary.</p> <p>For the elaboration of conclusions concerning the consistency of a cause-effect relationship between an intervention and a clinical outcome indirect evidence is commonly used: surrogate end-points, relationships established with similar technologies, or in similar populations. Acknowledging that in many situations indirect evidence provides useful information they need to be carefully analysed and their limitations accounted for.</p>
			Evaluación epidemiologica de tecnologías de salud (1995)	Keine Angaben
	AETSA	Nein	Framework for the assessment of genetic testing in the Andalusian public health system (2006)	<p><b>S. 5</b> At the same time, and following the trend of evidence-based medicine and outcome research, the comparison should be performed by means of well-designed studies and on the basis of relevant and final outcomes (as opposed to interim outcomes).</p> <p><b>S. 15</b> The decision to incorporate new genetic tests into clinical practice should be based on an assessment of the evidence concerning how it runs and the health outcomes resulting from the interventions / decisions taken on the basis of the information provided by the test, in addition to an estimate of the social, ethical, organisational and economic implications of its inclusion in the services portfolio.</p> <p><b>S. 24</b> In order to assess the evidence for the clinical utility of a genetic test, the ideal type of study is a clinical trial, which evaluates the complete intervention strategy (diagnostic tests and subsequent intervention), with a sufficiently long follow-up period to enable assessment of the final outcome.</p> <p><b>S 26</b> Nevertheless, in many cases these conditions are not met and other options could be assessed with regard to deciding on the incorporation of a new genetic test into clinical practice. In this case, the decision would be taken on the basis of different studies which separately measure different parts of the causal chain (a chain which starts with conducting a genetic test and ends with a major health outcome). Nevertheless, this approach has major limitations and must therefore be considered with caution.</p>
	AVALIA-T	Nein	Nein	-
	CAHTA	Nein	Nein	-
ES	OSTEBA	Nein	La identificación de tecnologías sanitarias emergentes (2002)	Keine Angaben

			Guia para le evaluaci3n del impacto en la salud y en el bienestar de proyectos, programas, o politicas extrasanitarias (2005)	Es handelt sich um Health Impact Assessment und nicht um Health Technology Assessment
			Guía de evaluaci3n econ3mica en el sector sanitario (1999)	<b>S. 53</b> Una evaluaci3n econ3mica normalmente requiere una gran cantidad de datos sobre los efectos de las opciones que se comparan, las cuales se pueden clasificar en: [...] Datos sobre los efectos en salud (efectividad o resultados) de las opciones. [Eine gesundheits3konomische Evaluation bedarf eine groÙe Menge an Daten 3ber die Effekte der zu vergleichenden Alternativen, die wie folgt klassifiziert werden k3nnen: [...]] Daten 3ber die Gesundheitseffekte (effectiveness oder endpunkte) der Alternativen.
ES	UETS	Nein	Nein	-
FI	FinOHTA	Nein	Nein	-
FR	CEDIT	Nein	Nein	-
	HAS	Nein	Guide d'analyse de la litterature et gradations des recommandations (2000)	Keine Angaben
			Bases m3thodologiques pour l'3laboration de recommandations professionnelles par consensus formalis3 (2006)	Keine Angaben
HU	HunHTA	Nein	Nein	-
IL	ICHTACH	-	Suchmaske nur auf h3br3isch anwendbar	-
LV	VMSMTVA	Nein	Ja	-
MX	CENETEC	Nein	Nein	-
	IMSS	Nein	Nein	-
NL	CVZ	Nein	Nein	-
	GR	Nein	Nein	-
	ZonMW	Nein	Nein	-
NO	NOKC	-	Slik oppsummerer vi forskning H3ndbok for Nasjonalt kunnskapssenter for helsetjenesten (2006)	Web-Site nur auf norwegisch Methodenpapier nur auf norwegisch
NZ	NZHTA	Nein	Nein	-
PL	AHTA-Pol	-	Web-Site nur auf polnisch	-
SE	CMT	Nein	Nein	-
	SBU	Nein	Nein	-

UK	CRD	Nein	Undertaking Systematic Reviews of research on effectiveness (2000)	<b>S. 4</b> [...] When formulating review questions, all clinically relevant and important outcomes should be identified. Use of intermediate, surrogate or proxy outcomes (e. g. intraocular pressure as a surrogate for visual field damage in ocular glaucoma or loss of bone mineral content as a surrogate for fractures in hormone replacement therapy) can be misleading and conclusions from reviews based on these are likely to be less robust.
	NCCHTA	Nein	Nein	-
	NHS-QIS	Nein	Nein	-
	NHSC	Ja	Nein	Literaturhinweis: An epistemology of patient safety research: A framework for study design and interpretation Report to patient safety alliance (World Health Organisation) (2007)
	IAHS	-	Nein	-
USA	AHRQ	Nein	Nein	-
	VATAP	Nein	Evaluating Diagnostic Tests: A Guide to the Literature (1997)	<p><b>S. 4.</b> führt das hierarchische Model der Evaluation von diagnostischen Technologien nach Fryback and Thronbury:</p> <p><i>Level Typical Measures of Analysis</i></p> <p>1. Technical Efficacy</p> <ul style="list-style-type: none"> <li>• Resolution of line pairs</li> <li>• Modulation transfer function change</li> <li>• Gray-scale range</li> <li>• Amount of mottle</li> <li>• Sharpness</li> </ul> <p>2. Diagnostic Accuracy Efficacy</p> <ul style="list-style-type: none"> <li>• Yield of abnormal or normal diagnoses in a case series</li> <li>• Diagnostic accuracy (% correct diagnoses in case series)</li> <li>• Sensitivity and specificity in a defined clinical problem setting</li> <li>• Measures of area under the ROC curve</li> </ul> <p>3. Diagnostic Thinking Efficacy</p> <ul style="list-style-type: none"> <li>• Number (%) of cases in a series in which image judged “helpful” to making the diagnosis</li> <li>• Entropy change in differential diagnosis probability distribution</li> <li>• Difference in clinicians’ subjectively estimated diagnosis probabilities pre- to post-test information</li> <li>• Empirical subjective log-likelihood ratio for test positive and negative in a case series</li> </ul> <p>4. Therapeutic efficacy</p> <ul style="list-style-type: none"> <li>• Number (%) of times image judged helpful in planning management of the patient in a</li> </ul>

				<p>case series</p> <ul style="list-style-type: none"> <li>• % of times medical procedure avoided due to image information</li> <li>• % of times therapy planned pretest changed after the image information was obtained (retrospectively inferred from clinical records)</li> <li>• % of times clinicians' prospectively stated therapeutic choices changed after test Information</li> </ul> <p>5. Patient Outcome Efficacy</p> <ul style="list-style-type: none"> <li>• % of patients improved with test compared with no test</li> <li>• Morbidity or procedure avoided after having image information</li> <li>• Change in quality-adjusted life years (QALYs)</li> <li>• Cost per QALY saved with image information</li> </ul> <p>6. Societal Efficacy</p> <ul style="list-style-type: none"> <li>• Cost-benefit analysis from societal perspective</li> <li>• Cost-effectiveness analysis from societal perspective</li> <li>• Cost-utility analysis from societal perspective</li> </ul>
--	--	--	--	--

AETMIS = Agence d'Evaluation des Technologies. AETS = Agencia de Evaluación de Tecnologías Sanitarias. AETSA = Agencia de Evaluación de Tecnologías Sanitarias de Andalucía. AHTA = Adelaide Health Technology Assessment. AHTA-Pol = Agency for Health Technology Assessment in Poland. AHRQ = Agency for Healthcare Research and Quality. AR = Argentinien. ASERNIP-S = Australian Safety and Efficacy Register of New Interventional Procedures – Surgical. AT = Österreich. AU = Australien. AVALIA-T = Galician Agency for Health Technology Assessment. BE = Belgien. BR = Brasilien. CA = Kanada. CADTH = Canadian Agency for Drugs and Technologies in Health. CAHTA = Catalan Agency for Health Technology Assessment and Research. CEDIT = Comité d'Evaluation et de Diffusion des Innovations Technologiques Assistance Publique Hôpitaux de Paris. CENETEC = Centro Nacional de Excelencia Tecnológica en Salud. CH = Schweiz. CMT = Centre for Medical Technology Assessment. CRD = Centre for Reviews and Dissemination. CVZ = College voor Zorgverzekeringen. DACEHTA = Danish Centre for Evaluation and Health Technology Assessment. DAHTA@DIMDI = Deutsche Agentur für Health Technology Assessment im Deutschen Institut für Medizinische Dokumentation und Information. DE = Deutschland. DECIT-CGATS = Department of Science and Technology, Brazilian Health Technology Assessment General Coordination. DK = Dänemark. DSI = Danish Institute for Health Services Research and Development. ES = Spanien. FI = Finnland. FINoHTA = Finnish Office for Health Care Technology Assessment. FR = Frankreich. GR = Health Council of the Netherlands. HAS = Haute autorité de santé. HTA = Health Technology Assessment. HTAi = Health Technology Assessment international. HU = Ungarn. HunHTA = Unit of Health Economics and Health Technology Assessment. IAHS = Institute of Applied Health Sciences. IECS = Institute for Clinical Effectiveness and Health Policy. IHE = Institute of Health Economics. IL = Israel. INAHTA = International Network of Agencies for Health Technology Assessment. ICTAHC = Israel Center for Technology Assessment in Health Care. IQWiG = Institut für Qualität und Wirtschaftlichkeit im Gesundheitswesen. KCE = Belgian Health Care Knowledge Centre. LBI@HTA = Ludwig Boltzmann Institute of Health Technology Assessment. LV = Lettland. MAS = Medical Advisory Secretariat. MSAC = Medical Services Advisory Committee. MTU-SFOPH = Medical Technology Unit, Swiss Federal Office of Public Health. MX = Mexiko. NCCHTA = National Coordinating Centre for Health Technology Assessment. NHS-QIS = National Health Service – Quality Improvement Scotland. NHSC = National Horizon Scanning Centre. NL = Niederlande. NO = Norwegen. NOKC = Norwegian Knowledge Centre for the Health Services. NZ = Neuseeland. NZHTA = New Zealand Health Technology Assessment. OSTEBA = Basque Office for Health Technology Assessment. PL = Polen. QALY = Quality adjusted life years. ROC = Receiver-Operator Characteristic Curve. SBU = Swedish Council on Technology Assessment in Health Care. SE = Schweden. UETS = Unidad de Evaluación de Tecnologías Sanitarias. UK = Großbritannien. USA = Vereinigte Staaten von Amerika. VATAP = Veterans Affairs Technology Assessment Program. ZonMW = The Netherlands Organisation for Health Research and Development.  
[Hinweise auf Literatur werden entfernt]

**Tabelle 23: Aussagen zu Surrogaten in Methodenpapieren der Agenturen zur Arzneimittelnutzenbewertung im Rahmen von Kostenerstattungsentscheidungen**

Land	Institution	Methodenpapiere	Aussage zur Auswahl der Endpunkten im Allgemeinen und zu Surrogaten im Besonderen
AT	Hauptverband der Österreichischen Sozialversicherungsträger, Heilmittelvaluierungskommission	Nein (Verfahrensordnung)	-
AU	Australian Government- Commonwealth Department of Health and Ageing Pharmaceutical Benefits Advisory Committee	Guidelines for preparing submissions to the Pharmaceutical Benefits Advisory Committee (2006)	<p><b>S.91</b> [...] The outcomes measured in the trial might not be the patient-relevant final outcomes of treatment. In this case, the clinical evaluation would need to be transformed to take account of the patient-relevant final outcomes (in terms of quality-adjusted life-years gained). Examples of premodelling studies of transformation include transforming comparative treatment effects measured on surrogate outcomes to final outcomes and scenario-based studies to value health outcomes using utilities.</p> <p><b>S.95</b> [...] Define any issues with outcomes that indicate a need to transform the nature of the outcome(s) measured in the direct randomised trials to those relied on in the economic evaluation. For example, the direct randomised trials might only report outcomes that are of less patient relevance than intended final outcomes of treatment. These less relevant outcomes are known as surrogate outcomes. Arguably, the closer a surrogate outcome is to the final outcome, the more useful it is, but generally the more difficult it is to measure accurately. To transform the surrogate outcomes measured in the trials to final outcomes and to extend the range of outcomes (for instance, the number of patients with unhealed peptic ulcers who eventually need surgery), the trial results might need to be supplemented by estimates obtained from other sources. For most drugs the ultimate outcome of therapy is to improve quality of life and/or survival, and in theory all outcomes could be expressed as quality-adjusted life-years (QALYs) gained. In practice, few randomised trials have measured the impact of drug therapy on QALYs, because few are large enough or long enough to measure changes in final outcomes directly. For instance, the ultimate aim of lowering moderately elevated blood pressure with a new antihypertensive medication is to reduce the risk of death and impaired quality of life from a stroke or possibly a myocardial infarction. The ultimate aim of treating a patient with severe asthma is to prevent death, to prevent hospitalisation and to return the patient to a normal level of functioning. The response measures used in many trials will usually be readily measured physiological variables (surrogate outcomes). For the two examples given above, this would be blood pressure and spirometry.</p> <p><b>S. 107</b> (Use of surrogate outcomes to estimate final outcomes) The claim that an incremental treatment effect on a surrogate outcome measured with the proposed drug quantitatively predicts a subsequent incremental treatment effect on a final outcome is more persuasively shown if attention is given to the following issues.</p> <ul style="list-style-type: none"> <li>• Step 1 — Present a systematic review of the literature to examine whether epidemiological evidence and biological reasoning has established that there is a relationship between the surrogate outcome and the final outcome independent of any intervention. In a few instances, relationships have been established, or have been proposed, between surrogate outcomes and final outcomes. Examples include blood</li> </ul>

			<p>left ventricular ejection fraction and survival after myocardial infarction, or viral load and cure of viral hepatitis.</p> <ul style="list-style-type: none"> <li>• Step 2 — Present a systematic review of the literature to examine whether randomised trial evidence using other drugs has shown that there is a basis to conclude that a treatment effect on the surrogate outcome has satisfactorily predicted a treatment effect on the final outcome. (If there is evidence of this type for the proposed drug, this might help support a biological argument for the treatment.) Based on this evidence, quantify the relationship between these treatment effects with an assessment of the uncertainty of the relationship. Discuss the reproducibility of these findings (eg whether they have been consistently shown across more than one trial and for more than one alternative drug and mechanism of action).</li> <li>• Step 3 — Explain why this relationship between the treatment effects on these outcomes with these other drugs is likely to apply to the proposed drug. Refer in this explanation to the mechanism of action of the proposed drug compared with the mechanism(s) of action of the drugs contributing evidence to Step 2 (a so-called 'class effects' argument). At present, it is difficult to give categorical advice. Consider which outcomes are most appropriate and most feasible, given the data available. The clinical importance and patient relevance of the outcomes should be established and, where possible, supported with data. Having addressed the three steps above in transforming a treatment effect on a surrogate outcome to a treatment effect on a final outcome, explain in response to Subsection D.4 how this is included in the economic evaluation, including by specifying and referencing the sources of the longer term natural history (eg longitudinal population studies) as well as the transformed treatment effects.</li> </ul>
CA	Patented Medicine Prices Review Board	Compendium of Guidelines Policies and Procedures (2003)	Keine Angaben
CH	Bundesamt für Gesundheit Eidgenössische Arzneimittelkommission/Eidgenössische Rekurskommission für die Spezialitätenliste in der Krankenversicherung	Nein (Verordnungen und Bundesgesetze)	-
FI	Pharmaceuticals Pricing Board (Lääkkeiden hintalautakunta)	Ja (finnisch)	-
FR	Ministère de la Santé et de la Protection Sociale Commission de Transparence National Transparency Commission Comité économique des Produits de Santé	Nein	-
NL	Health Care Insurance Board (College voor zorgverzekeringen)	Dutch guidelines for pharmacoeconomic research	<b>S. 13</b> [...] In a CEA, the so-called intermediate outcomes, such as the millimetres of mercury in blood pressure, should preferably be translated into final outcomes, such as life-years saved.

NO	Norwegian Medicines Agency (Statens Legemiddelverk)	Norwegian guidelines for pharmacoeconomic analysis in connection with applications for reimbursement (2005)	<p>[...] 4.4 Health outcome The drug's documented primary effects with direct clinical relevance for the indication shall be given. Testing methods and parameters used must be described. If intermediary/surrogate endpoints are used to describe the effects of the drug, the relation of these to the final end points relevant to the medically approved indication must be clarified. Should the documented outcome include only parts of the patient group, outcome must be presented through sub-group analyses.</p> <p>Appendix: 2. Cost-effectiveness analysis With calculation of intermediary outcome measure, it is important to describe/document the connection between the intermediary and final health outcome, such as cholesterol level and development of heart/vascular disease or death.</p>
NZ	PHARMAC/Pharmacology and Therapeutic Advisory Committee	Recommended methods to derive clinical inputs for proposals to PHARMAC (2005)	<p><b>S. 17</b> Outcome measures...If relevant (particularly when clinically relevant/valid outcome data are not available), use intermediate or surrogate outcomes. When using intermediate/surrogate outcomes, describe the strength of evidence showing that extrapolating the specific intermediate outcomes to clinically relevant patient outcomes is valid.</p> <p><b>S. 23</b> Relevance of the evidence....For each study, assess its relevance according to the following relevance grades: Grade I Evidence of effect of long term patient focused outcomes, taking into account benefits and harms, especially quality of life and survival Grade II Evidence of effect on short to medium term clinical outcomes, taking into account benefits and harms Grade III Evidence of effect confined to intermediate or surrogate outcomes</p> <p><b>S. 24</b> Surrogate/intermediate outcomes 'Surrogate/intermediate' outcomes are essentially biological markers. Commonly a physiological variable (e. g. serum LDL-cholesterol concentration, blood pressure), a surrogate/intermediate outcome has a statistical association with clinical outcome of interest (e. g. bone mineral density with fracture, CD4 cell concentrations with progression of HIV). There will also be a biological and pathophysiological basis for believing that the surrogate/intermediate outcome is a major determinant of the clinical outcome in the disease being studied (e. g. glycosylated haemoglobin (HbA1c) and diabetes complications).</p> <p>A surrogate/intermediate outcome should possess all of the above features, but few do. If using a study with surrogate/intermediate outcomes as the basis for an application, then present evidence linking the surrogate/intermediate outcome with the clinical outcome of interest. Evidence relying on surrogate/intermediate outcomes alone will be unlikely to be sufficient base judgement about the true benefits of a technology.</p>
		Prescription for Pharmacoeconomic Analysis Methods for cost-utility analysis (2006)	<p><b>S.27</b> 5.4.1 Extrapolation to Longer Terms or Final Outcomes Many trials have endpoints that may be too early. It may therefore be necessary to use intermediate outcomes to obtain final endpoints by extrapolating data beyond the period observed in the clinical trials (for example, when estimating life-expectancy). This requires explicit assumptions regarding the continuation of treatment effect once treatment has ceased.</p>

SE	Pharmaceutical Benefits Board (Läkemedelsförmånsämnden)	General guidelines for economic evaluations from the Pharmaceutical Benefits Board (LFNAR 2003:2) (2003)	<p><b>S.1</b> (5. Analytical method)                      Cost-effectiveness analysis is recommended, with quality-adjusted life years (QALY's) as the measure of effect. In treatments that mostly affect survival, both QALY's and gained life years should be shown. If so-called surrogate end-points are used, the account should also include modelling from these end-points to illustrate the effects on mortality and morbidity, i. e. QALY's gained. The same applies to other types of events studied which are usual in clinical studies (e. g. expected number of heart attacks 2 or migraine attacks). If it is difficult to use QALY's (e. g. with heavy pain over a short time in connection with treatment), then a cost-benefit analysis with the willingness to pay may be used as a measure of effect. If there is supporting evidence that the drug to which the application refers has the same health effect as the best comparable treatment, a cost comparison may suffice.</p>
UK	National Institute of Health and Clinical Excellence	Guide to the Technology Appraisal Process (2004)	<p><b>S. 30</b> [Eintrag im Glossar] Outcome: Measure of the possible results that may stem from exposure to a preventive or therapeutic intervention. Outcome measures may be intermediate endpoints or they can be final endpoints.</p>
		Guide to the Methods of Technology Appraisal (2004)	<p><b>S. 17</b> The written submissions provide a unique contribution outlining the professional view of the place of the technology in current clinical practice. This includes evidence that relates to some or all of the following:                      [...] the identification of appropriate outcome measures and the appropriate use of surrogate outcome measures  <b>S. 22</b> the perspective on outcomes should be all direct health effects whether for patients or, where relevant, other individuals (principally carers).  <b>S.23</b> The analysis of clinical effectiveness should consider the range of typical patients, normal clinical circumstances, clinically relevant outcomes and comparison with relevant comparators.</p>
		Guidance for manufactors and sponsors (2001)	<p><b>S. 3</b> The revised Guidance is based on the following principles:                      (i) the best available data relevant to the problem are required;                      (ii) data on final outcomes are preferred to intermediate (surrogate) outcomes;  <b>S. 12</b> [...]the design of efficacy trials can be made more relevant, for example, by the inclusion of usual care in comparisons, the use of final outcome measures as well as intermediate clinical endpoints and the collection of resource use data.                      [...] The measures of clinical effectiveness should be appropriate for the condition and include final outcomes such as reduced mortality and improved quality of life. In studies relying on intermediate (surrogate) outcomes one of the following should be demonstrated: (i) a strong and consistent association between the intermediate outcome and final outcome; (ii) where the intermediate results are large, precise and lasting enough to be considered a possible basis for making a treatment decision.                      [...] For some technologies, for example in diagnostics, it has been customary to design trials to measure only diagnostic accuracy, rather than impact on final patient outcomes. The latter carries more weight. The points made above about the use of surrogate outcomes are particularly relevant for such technologies.</p>
		A guide for manufacturers and sponsors (2004)	Keine Angaben

†T= Österreich. AU = Australien. CA = Kanada. CEA: = Cost-effectiveness analysis. CH = Schweiz. FI = Finnland. FR = Frankreich. NL = Niederlande. NO = Norwegen. NZ = Neuseeland.

‡HARMAC = Pharmaceutical Management Agency. SE = Schweden. UK = Großbritannien.

## 8.4 HTA-Agenturen, deren Berichte in der HTA-Datenbank enthalten sind

### INAHTA-Mitglieder sowie ihr Herkunftsland

AETMIS	Agence d'Evaluation des Technologies, Kanada
AETS	Agencia de Evaluación de Tecnologías Sanitarias, Spanien
AETSA	Agencia de Evaluación de Tecnologías Sanitarias de Andalucía, Spanien
AHFMR	Alberta Heritage Foundation for Medical Research, Kanada
AHRQ	Agency for Healthcare Research and Quality, USA
AHTA	Adelaide Health Technology Assessment, Australien
AHTA-Pol	Agency for Health Technology Assessment in Poland, Polen
ASERNIP-S	Australian Safety and Efficacy Register of New Interventional Procedures - Surgical, Australien
CADTH	Canadian Agency for Drugs and Technologies in Health, Kanada
CAHTA	Catalan Agency for Health Technology Assessment and Research, Spanien
CEDIT	Comité d'Evaluation et de Diffusion des Innovations Technologiques Assistance Publique Hôpitaux de Paris, Frankreich
CMT	Center for Medical Technology Assessment, Schweden
CRD	Centre for Reviews and Dissemination, UK
CVZ	College voor Zorgverzekeringen, Niederlande
DACEHTA	Danish Centre for Evaluation and HTA, Dänemark
DAHTA@DIMDI	Deutsche Agentur für Health Technology Assessment im Deutschen Institut für Medizinische Dokumentation und Information, Deutschland
DECIT-CGATS	Department of Science and Technology, Brazilian Health Technology Assessment General Coordination, Brasilien
DSI	Danish Institute for Health Services Research and Development, Dänemark
FinOHTA	Finnish Office for Health Care Technology Assessment, Finnland
GR	Health Council of the Netherlands, Niederlande
HAS	Haute Autorité de Santé/French National Authority for Health, Frankreich
HunHTA	Unit of Health Economics and Health Technology Assessment, Ungarn
IAHS	Institute of Applied Health Sciences, UK
ICTAHC	Israel Center for Technology Assessment in Health Care, Israel
IECS	Institute for Clinical Effectiveness and Health Policy, Argentinien
IHE	Institute of Health Economics, Kanada
KCE	Belgian Health Care Knowledge Centre, Belgien
LBI@ITA	Ludwig Boltzmann Institute of Health Technology Assessment, Österreich (früher: ITA)
MAS	Medical Advisory Secretariat, Ontario Ministry of Health and Long-Term Care, Kanada
MSAC	Medical Services Advisory Committee, Australien
MTU-SFOPH	Medical Technology Unit, Swiss Federal Office of Public Health, Schweiz
NCCHTA	National Coordinating Centre for Health Technology Assessment, UK
NHSC	National Horizon Scanning Center, UK
NHS-QIS	National Health Service - Quality Improvement Scotland, UK
NOKC	Norwegian Knowledge Centre for the Health Services, Norwegen (früher: SMM)

**INAHTA-Mitglieder sowie ihr Herkunftsland**

NZHTA	New Zealand Health Technology Assessment, Neuseeland
OSTEBA	Basque Office for Health Technology Assessment, Spanien
SBU	Swedish Council on Technology Assessment in Health Care, Schweden
UETS	Unidad de Evaluacion de Tecnologias Sanitarias, Spanien
VATAP	VA Technology Assessment Program, USA
ZonMW	The Netherlands Organisation for Health Research and Development, Niederlande

**Andere internationale Institutionen (nicht INAHTA-Mitglied)**

BCBS	BlueCross BlueShield Technology Evaluation Center, USA
BCOHTA	British Columbia Office of Health Technology Assessment, Kanada (jetzt geschlossen)
CCE	The Centre for Clinical Effectiveness, Australien
CHUM	Centre Hospitalier de l'Université de Montréal, Kanada
CTFPHC	Canadian Task Force on Preventive Health Care, Kanada
ECRI	USA
ETESA	Unidad De Tecnologias De Salud, Chile
HAYES Inc.	USA
HSURC	Health Services Utilization and Research Committee, Kanada (jetzt geschlossen)
HTAC	Health Technology Advisory Committee, USA (jetzt geschlossen)
ICES	Institute for Clinical Evaluative Sciences, Kanada
ICSI	The Institute for Clinical Systems Improvement, USA
MHH	Hannover Medical School, Medizinische Hochschule Hannover, Deutschland
MHTAU	Malaysian Health Technology Assessment Unit, Malaysia
MUHC	Technology Assessment Unit of the McGill University Health Centre, Kanada
NICE	National Institute for Health and Clinical Excellence, UK
OTA	Office of Technology Assessment, USA (jetzt geschlossen)
SHPIC	Scottish Health Purchasing Information Centre, UK (jetzt geschlossen)
TA-SWISS	Swiss Science and Technology Council/Technology Assessment, Schweiz
TNO	TNO Prevention and Health, Niederlande
UHC	University HealthSystem Consortium, USA
WIHRD	Wessex Institute for Health Research and Development: STEER Reports (herausgegeben von Bazian Ltd), UK
WMHTAC	West Midlands Health Technology Assessment Collaboration, UK
WorkSafe BC	Workers' Compensation Board of BC, Kanada
WSDLI	Washington State Department of Labor and Industries, USA

## 8.5 HTA-Berichte aus der HTA-Datenbank

1. Adelaide Health Technology Assessment on behalf of National Horizon Scanning Unit (HealthPACT and MSAC). The Meniett device for the treatment of Meniere's disease. Adelaide 2005.
2. Agencia de Evaluacion de Tecnologias Sanitarias. Health risks posed by silicone implants in general, with special attention to breast implants. IPE-00/23 (Public report). Madrid 2000.
3. Agency for Healthcare Research and Quality. Prevention of venous thromboembolism after injury. Rockville 2000.
4. Agency for Healthcare Research and Quality. Percutaneous kyphoplasty for vertebral fractures caused by osteoporosis and malignancy and percutaneous vertebroplasty for vertebral fractures caused by osteoporosis and malignancy. Rockville 2005.
5. Agency for Healthcare Research and Quality. Point of care testing of hemoglobin A1C. Rockville 2005.
6. Aguado RJ, Llanos MA. Vaccine for primary prophylaxis against human papillomavirus infection. Sevilla 2006.
7. Alberta Heritage Foundation for Medical Research. Vagus nerve stimulation for refractory epilepsy. Edmonton 1998.
8. Amin M. Independence iBOT 3000 mobility system: a stair-climbing wheelchair. Ottawa 2004.
9. Andalusian Agency for Health Technology Assessment. Use of the drug eluting stents - review. Sevilla 2004.
10. Andalusian Agency for Health Technology Assessment. Use of flash sterilization. Sevilla 2005.
11. Andalusian Agency for Health Technology Assessment. Stabilised hyaluronic acid in the treatment of osteoarthritis. Sevilla 2006.
12. Australian Health Technology Advisory Committee. Minimal access surgery. Canberra 1996.
13. Australian Safety and Efficacy Register of New Interventional Procedures - Surgical. A systematic review of laparoscopic-assisted resection of colorectal malignancies. Canberra 2000.
14. Ball CM. Hyperbaric oxygen therapy for multiple sclerosis. Anonymous. Anonymous. CO1: United Kingdom: 2002. Report. 2004 .  
[www.wihrd.soton.ac.uk/projx/signpost/welcome.htm](http://www.wihrd.soton.ac.uk/projx/signpost/welcome.htm)
15. Banken R. Low-intensity ultrasound (Exogen (TM)) for the treatment of fractures. Montreal 2004.
16. Basque Office for Health Technology Assessment. Analysis of the scientific evidence about effectiveness and efficiency of the different hip prostheses - primary research, systematic review. Vitoria-Gasteiz 1999.
17. Baxter-Jones ADG, Helms PJ, Russell R, Grant A, Ross S. Early asthma prophylaxis, natural history, skeletal development and economy (EASE): a pilot randomised controlled trial. Southampton 2000.
18. Berry E, Kelly S, Westwood ME, Davies LM, Gough MJ, Bamford JM, Meaney JFM, Airey CM, Cullingworth J, Barbieri M, Jackson A, Smith MA. The cost-effectiveness of magnetic resonance angiography for carotid artery stenosis and peripheral vascular disease: a systematic review. Southampton 2002.
19. Blue Cross Blue Shield Association. Extracorporeal shock wave treatment for chronic plantar fasciitis. Chicago 2005.
20. Blue Cross Blue Shield Association. Off-label uses of bevacizumab: renal cell carcinoma and other miscellaneous non-colorectal cancer indications. Chicago 2006.

21. Bowen J, Hopkins R, He Y, Blackhouse G, Lazzam C, Tu J, Cohen E, Tarride JE, Goeree R. Systematic review and cost-effectiveness analysis of drug eluting stents compared to bare metal stents for percutaneous coronary interventions in Ontario. Interim Report for the Ontario Ministry of Health and Long-term Care. Toronto 2005.
22. Brooks L. Overview of the regulation of acupuncture in Alberta. Edmonton 2001.
23. Brophy J, Chen J. Use of carmustine implants (GLIADEL wafer) in patients with malignant glioma at The McGill University Health Centre. 2004.
24. Bryan E. The benefit of moving antenatal classes out of the hospital into the community. Clayton 2002.
25. Canadian Agency for Drugs and Technologies in Health. Aliskiren. Ottawa 2006.
26. Canadian Coordinating Office for Health Technology Assessment. Finasteride: clinical and economic impacts - summary. Ottawa 1996.
27. Canadian Coordinating Office for Health Technology Assessment. Gliasite radiation therapy. No. 9. Canada: 2001.  
www.cadth.ca (07.04.2009)
28. Canadian Coordinating Office for Health Technology Assessment. Radio frequency energy. No. 12. Canada: 2002.  
www.cadth.ca (07.04.2009)
29. Canadian Coordinating Office for Health Technology Assessment. Adalimumab and rheumatoid arthritis. Canada: 2003. Report No. 39, 12.05.2003.  
www.cadth.ca (07.04.2009)
30. Canadian Coordinating Office for Health Technology Assessment. Voriconazole. Anonymous. Anonymous. Canada: 2003. Report. 24.02.2003.  
www.cadth.ca (07.04.2009)
31. Canadian Coordinating Office for Health Technology Assessment. Bevacizumab for advanced colorectal cancer. Canada, 2004. Report.  
www.cadth.ca (07.04.2009)
32. Canadian Coordinating Office for Health Technology Assessment. An assessment of bisphosphonate drugs to manage pain secondary to bone metastases. Ottawa: Canadian Coordinating Office for Health Technology Assessment (CCOHTA):14, 2004.
33. Centre for Reviews and Dissemination. Inhaler devices for the management of asthma and COPD. York 2003.
34. Centre for Reviews and Dissemination. Effectiveness of antihypertensive drugs in black people. York 2004.
35. Cooter R, Babidge W. Ultrasound-assisted lipoplasty. Canberra 1999.
36. Corabian P. In vitro fertilization and embryo transfer as a treatment for infertility. Ottawa 1997.
37. Corbillon E, Poullie AI, Doublet JD. Laparoscopic versus open surgery live-donor nephrectomy. Paris 2004.
38. Costa V, Brophy J, McGregor M. Vacuum-assisted wound closure therapy (V.A.C (R)). 2005
39. Coster S, Gulliford MC, Seed PT, Powrie JK, Swaminathan R. Monitoring blood glucose control in diabetes mellitus: a systematic review. Southampton 2000.
40. Cummins C, Marshall T, Burls A. Percutaneous endoscopic gastrostomy (PEG) feeding in the enteral nutrition of dysphagic stroke patients. Birmingham 1999.
41. Dalziel K, Round A, Stein K, Garside R, Price A. Effectiveness and cost-effectiveness of imatinib for first-line treatment of chronic myeloid leukaemia in chronic phase: a systematic review and economic analysis. Southampton 2004.
42. Danish Centre for Evaluation and Health Technology Assessment. Alefacept (Amevive(R)) for the treatment of severe psoriasis - Early Warning on New Health Technology 2003 2(3). Copenhagen 2003.

43. Dretzke J, Cummins C, Sandercock J, Fry-Smith A, Barrett T, Burls A. Autoantibody testing in children with newly diagnosed type 1 diabetes mellitus. Southampton 2004.
44. Ebrahim S. Detection, adherence and control of hypertension for the prevention of stroke: a systematic review. Southampton 1998.
45. Einarson TR, Iskedjian M. Novel antipsychotics for patients with attention-deficit hyperactivity disorder: a systematic review. Ottawa 2002.
46. Erickson L. Assessment of photodynamic therapy using porfimer sodium for esophageal, bladder and lung cancers. Montreal 2004.
47. French S. Phototherapy in the home for jaundiced neonates. Clayton 2003.
48. Garces K, McCormick A, McGahan L and Skidmore B. Botulinum toxin A in the treatment of upper and lower limb spasticity: a systematic review of randomized controlled trials. Ottawa 2006.
49. Gaynes BN, West SL, Ford C, Frame P, Klein J and Lohr KN. Screening for suicide risk: a systematic evidence review for the U.S. Preventive Services Task Force. Rockville 2004.
50. Gebo K, Jenckes M and Chander G. Management of chronic hepatitis C. Rockville 2002.
51. German Agency for Health Technology Assessment at the German Institute for Medical Documentation and Information. Autologous bone marrow transplantation versus high dose chemotherapy in metastatic breast carcinoma - systematic review. Köln 1999.
52. German Agency for Health Technology Assessment at the German Institute for Medical Documentation and Information. Evaluation of optical coherence tomography in the diagnosis of age related macular degeneration compared with fluorescence angiography. Köln 2007.
53. German Agency of Health Technology Assessment at German Institute for Medical Documentation and Information (DAHTA@DIMDI). Importance of Nuclear Magnetic Resonance in caring for patients with back pain. Köln 2002.
54. Gibis B, Busse R, Reese E, Richter K, Schwartz FW and Kobberling J. Mammography screening as a method for the early detection of breast cancer. Baden Baden 1998.
55. Gilbert FJ, Grant AM, Gillan MGC, Vale L, Scott NW, and Campbell MK. Does early magnetic resonance imaging influence management or improve outcome in patients referred to secondary care with low back pain? A pragmatic randomised controlled trial. Southampton 2004.
56. Habl C, Bodenwinkler A and Sturzlinger H. Root treatment of molars. Köln 2005.
57. Harris R, Lohr KN, Beck R, Fink K, Godley P, and Bunton AJ. Screening for prostate cancer. Rockville 2002.
58. Haute Autorite de Sante/French National Authority for Health. Refractory obsessive compulsive disorders: conventional treatments and deep brain stimulation. Paris 2006.
59. Haute Autorite de Sante/French National Authority for Health. Value of mutation detection and of the activated protein C resistance assay in inherited thrombophilia. Paris 2007.
60. Hernandez JJ and Rueda JR. Haemodilution and perioperative blood recovery. Vitoria-Gasteiz: 2002.
61. Ho C, Hailey D, Warburton R, MacGregor J, Pisano E and Joyce J. Digital mammography versus film-screen mammography: technical, clinical and economic assessments. Ottawa 2002.
62. HTA Unit of the Institute of Technology Assessment. Colorectal cancer screening: Integrating colonoscopy/sigmoidoscopy into the Austrian preventive medical check up - unsystematic review. Wien 2003.
63. HTA Unit of the Institute of Technology Assessment of the Austrian Academy of Sciences. Chronic hepatitis C: implications for therapy and economic resources in Austria. Wien 2004.

64. Huriri-Driscoll A, Rasmussen P and Day P. Mental health services for older people: a critical appraisal of the literature. Christchurch 2004.
65. Institute for Clinical Systems Improvement. Blood brain barrier disruption chemotherapy. Bloomington 2001.
66. Institute for Clinical Systems Improvement. Off-pump coronary artery bypass grafting. Bloomington 2003.
67. Institute for Clinical Systems Improvement. Antiviral treatment for chronic hepatitis C. Bloomington 2005.
68. International Network of Agencies for Health Technology Assessment. Background paper one. Methods used in the measurement of bone density (INAHTA Joint Project). Stockholm 1996.
69. L'Agence Nationale d'Accreditation d'Evaluation en Sante (ANAES). Cardiac defibrillators. Paris 1998.
70. L'Agence Nationale d'Accreditation d'Evaluation en Sante (ANAES). Clinical evaluation of transmyocardial laser revascularisation techniques. Paris 1999.
71. L'Agence Nationale d'Accreditation d'Evaluation en Sante (ANAES). Treatment of age related macular degeneration. Paris 2001.
72. Lessard C and Framarin A. Endometrial ablation techniques in the treatment of dysfunctional uterine bleeding. Montreal 2002.
73. Lewis R, Bagnall AM, Forbes C, Shirran E, Duffy S, Kleijnen J, Riemsma R and Riet ter G. The clinical effectiveness of trastuzumab for breast cancer: a systematic review. Southampton 2002.
74. Liu Z, Albon E and Hyde C. The effectiveness and cost effectiveness of immunoglobulin replacement therapy for primary immunodeficiency and chronic lymphocytic leukaemia: a systematic review and economic evaluation. Southampton 2005.
75. Long A, McFadden C and DeVine D. Management of allergic and nonallergic rhinitis. Rockville 2002.
76. Makni H, St-Hilaire C, Robb L, Larouche K and Blancquaert I. Tandem mass spectrometry and neonatal blood screening in Quebec. Montreal 2007.
77. Malaysian Health Technology Assessment Unit. Ozone therapy. Kuala Lumpur 2005.
78. Mar J, Rueda JR and Duran J. Cost-effectiveness analysis of treatments of sleep apnea syndrome in the Basque Country. Vitoria-Gasteiz 2000.
79. Medical Services Advisory Committee. Brachytherapy for the treatment of prostate cancer. Anonymous. Canberra 2000.
80. Medical Services Advisory Committee. Positron emission tomography. Canberra 2000.
81. Medical Services Advisory Committee. Liquid based cytology for cervical screening. Anonymous. Canberra 2002.
82. Medical Services Advisory Committee. Placement of artificial bowel sphincters in the management of faecal incontinence: assessment report. Canberra 2003.
83. Medical Services Advisory Committee. Positron emission tomography (PET) for epilepsy. Canberra 2004.
84. Membe S, McGahan L, Cimon K, Gawel M, Giammarco R and Mierzwinski-Urban M. Triptans for acute migraine: comparative clinical effectiveness and cost-effectiveness. Ottawa 2007.
85. Moga C, Harstall C and Tang Z. Celecoxib for the treatment of pain in osteoarthritis and rheumatoid arthritis. Edmonton 2005.
86. Mundy L and Merlin T. Ultrafast magnetic resonance imaging (MRI) for patients undergoing conventional MRI. Horizon Scanning Prioritising Summary - Volume 2. Canberra 2003.

87. National Coordinating Centre for Health Technology Assessment. The clinical effectiveness and cost-effectiveness of strontium ranelate for the prevention of osteoporotic fragility fractures in postmenopausal women (Vol 11.4). Southampton 2007.
88. National Horizon Scanning Centre. Memantine for Alzheimer's disease and vascular dementia - horizon scanning review. Birmingham 2001.
89. National Horizon Scanning Centre. Preventing restenosis after PTCA - horizon scanning review. Birmingham 2001.
90. National Horizon Scanning Centre. Afelimomab for sepsis - horizon scanning review. Birmingham 2002.
91. National Horizon Scanning Centre. High intensity focused ultrasound for prostate cancer - horizon scanning review. Birmingham 2002.
92. National Horizon Scanning Centre. Nesiritide for acute heart failure - horizon scanning review. Birmingham 2002.
93. National Horizon Scanning Centre. Pegaptanib for age-related macular degeneration - horizon scanning review. Birmingham 2002.
94. National Horizon Scanning Centre. Porfimer sodium photodynamic therapy for high-grade dysplasia of Barrett's oesophagus - horizon scanning review. Birmingham 2003.
95. National Horizon Scanning Centre. aTL (autologous tumour vaccine) for stage I-III renal cell carcinoma - horizon scanning review. Birmingham 2004.
96. National Horizon Scanning Centre. Bivalirudin for percutaneous coronary interventions (PCI) - horizon scanning review. Birmingham 2004.
97. National Horizon Scanning Centre. Leflunomide (Arava) for adult psoriatic arthritis - horizon scanning review. Birmingham 2004.
98. National Horizon Scanning Centre. Citrasate dialysate for patients with chronic or acute renal failure requiring long-term haemodialysis: horizon scanning technology briefing. Birmingham 2006.
99. National Horizon Scanning Centre. Febuxostat for the management of hyperuricaemia in patients with gout: horizon scanning review. Birmingham 2006.
100. National Institute for Clinical Excellence. Improving outcomes in urological cancers. London 2002.
101. National Institute for Clinical Excellence. Balloon kyphoplasty for vertebral compression fractures. London 2003.
102. National Institute for Clinical Excellence. Guidance on the use of glitazones for the treatment of type 2 diabetes. London 2003.
103. National Institute for Clinical Excellence. Stentgraft placement in abdominal aortic aneurysm. London 2003.
104. National Institute for Clinical Excellence. Balloon angioplasty of pulmonary vein stenosis in infants. London 2004.
105. National Institute for Clinical Excellence. Endovascular closure of atrial septal defect. London 2004.
106. National Institute for Clinical Excellence. Free fluid thermal endometrial ablation. London 2004.
107. National Institute for Clinical Excellence. Minimally invasive two-incision surgery for total hip replacement. London 2005.
108. National Institute for Health and Clinical Excellence. Living-donor liver transplantation. London 2006.
109. National Institute for Health and Clinical Excellence. Inhaled insulin for the treatment of type 1 and type 2 diabetes. London 2006.

110. National Institute for Health and Clinical Excellence. High dose rate brachytherapy for carcinoma of the cervix. London: National Institute for Health and Clinical Excellence (NICE):2, 2006.
111. National Institute for Health and Clinical Excellence. Peginterferon alfa and ribavirin for the treatment of mild chronic hepatitis C. London 2006.
112. National Institute for Health and Clinical Excellence. Rituximab for the treatment of follicular lymphoma. London 2006.
113. National Institute for Health and Clinical Excellence. Septostomy with or without amnioreduction for the treatment of twin-to-twin transfusion syndrome. London 2006.
114. National Institute for Health and Clinical Excellence. Stent-graft placement in abdominal aortic aneurysm. London 2006.
115. National Institute for Health and Clinical Excellence. Cryotherapy for renal cancer. London 2007.
116. NHS Centre for Reviews and Dissemination. The treatment and management of chronic fatigue syndrome/myalgic encephalomyelitis in adults and children. York 2007.
117. Quality NHS, I. The use of surgery to aid weight reduction for people with morbid obesity. Glasgow 2002.
118. Quality NHS, I. The use of tension-free vaginal tape (Gynaecare TVT) for stress incontinence. Glasgow 2003.
119. Quality NHS, I. Rituximab for the treatment of follicular lymphoma. Glasgow 2006.
120. Quality NHS, I. Fludarabine monotherapy for first line treatment of chronic lymphocytic leukaemia. Glasgow 2007.
121. Ontario Ministry of Health and Long-Term Care. Endovascular repair of abdominal aortic aneurysm. Ottawa 2002.
122. Ontario Ministry of Health and Long-Term Care. Energy delivery systems for treatment of benign prostatic hyperplasia: health technology policy assessment. Toronto: Medical Advisory Secretariat, Ontario Ministry of Health and Long-Term Care (MAS):108, 2006.
123. Ontario Ministry of Health and Long-Term Care. Negative pressure wound therapy: health technology literature review. Toronto 2006.
124. Ontario Ministry of Health and Long-Term Care. Gastric electrical stimulation: health technology policy assessment. Toronto 2006.
125. Patterson J. Outcomes of abdominoplasty. London: Bazian Ltd (Editors), Wessex Institute for Health Research and Development, University of Southampton: 9, 2003.
126. Riviere A, Pichon A, Augustovski F, Cernadas C, Ferrante D, Regueiro A, and Garcia MS. Brachytherapy for prostate cancer treatment. Ciudad de Buenos Aires 2003.
127. Riviere A, Pichon A, Augustovski F, Alcaraz A, Bardach A, Ferrante D, Garcia MS, Glujovsky D, Lopez A, Regueiro A and Colantonio L. Etanercept, infliximab and adalimumab for the treatment of rheumatoid arthritis. Ciudad de Buenos Aires 2006.
128. Rodgers M, Fayter D, Richardson G, Ritchie G, Lewin R and Sowden AJ. The effects of psychosocial interventions in cancer and heart disease: a review of systematic reviews. York 2005.
129. Calvo A, Saenz A, Conde Olasagasti JL, Iglesia Imaz I and Hernandez TA. Effectiveness and safety of penile prosthesis. IPE-98/15 (Public report). Madrid 1998.
130. Sarria-Santamera A, Timoner-Aguilera J. Assessment of the effectiveness of health promotion and disease prevention technologies. Madrid 2003.
131. Shekelle P, Rich M and Morton S. Pharmacologic management of heart failure and left ventricular systolic dysfunction: effect in female, black, and diabetic patients, and cost-effectiveness. Rockville 2003.

132. Shukla VK, Chen S, Boucher M, Mensinkai S, and Dales R. Long-acting beta2 agonists for the maintenance treatment of chronic obstructive pulmonary disease in patients with reversible and non-reversible airflow obstruction: a systematic review of clinical effectiveness. Ottawa 2006.
133. Simpson S and Burls A. A systematic review of the effectiveness and cost-effectiveness of palivizumab (Synagis(R)) in the prevention of respiratory syncytial virus (RSV) infection in infants at high risk of infection. Birmingham 2001.
134. Stacey D, Noorani HZ, Fisher A, Robinson D, Joyce J and Pong RW. Telephone triage services: systematic review and a survey of Canadian call centre programs. Ottawa 2003.
135. Sutherland LM, Williams JAR, Padbury RTA, Gotley DG, Stokes B and Maddern GJ. A systematic review of radiofrequency ablation for the treatment of liver tumours. Australien: Oktober 2002. Systematic review.  
www.surgeons.org/asernip-s (07.04.2009)
136. Toohar RL. A systematic review of laparoscopic live donor nephrectomy (2nd update and re-appraisal). Canberra 2003.
137. Unidad de Evaluacion de Tecnologias Sanitarias Agencia Lain Entralgo. Immediate breast reconstruction (IBR). Systematic review and outcome assessment of an IBR unit in Madrid (Spain). Madrid 2005.
138. Waugh N, Black C, Walker S, McIntyre L, Cummins E and Hillis G. The effectiveness and cost-effectiveness of computed tomography screening for coronary artery disease: systematic review. Southampton 2006.
139. Wilby J, Kainth A, Hawkins N, Epstein D, McIntosh H and McDaid C. Clinical effectiveness, tolerability and cost-effectiveness of newer drugs for epilepsy in adults: a systematic review and economic evaluation. Southampton 2005.
140. Woodroffe R, Yao GL, Meads C, Bayliss S, Ready A and Raftery J. Clinical and cost-effectiveness of newer immunosuppressive regimens in renal transplantation: a systematic review and modelling study. Southampton 2005.

## 8.6 Deutsche HTA-Berichte

- BÄK/KBV001 Laserinduzierte interstitielle Thermotheapie (LITT) bei malignen Tumoren; Laser-induced interstitial thermotherapy (LITT) for the treatment of malignant tumors
- BÄK/KBV002 Fundus-Fotografie zur Früherkennung diabetesassoziierter Schädigungen des Augenhintergrundes (diabetische Retinopathie und Makulopathie); Fundus photography for early detection of diabetes associated damages of the eyeground
- BUA001 Pulsierende Signaltherapie (PST); Treatments using pulsed electromagnetic fields
- BUA002 Neurotopische Therapie nach Desnizza und ähnliche Therapien mit Kochsalzlösungsinjektionen; Neurotopical therapy according to Desnizza and similar therapies using saline solution injections
- BUA003 Autologe Chondrozytenimplantation (ACI); Autologous chondrocyte implantation and transplantation
- BUA004 Hyperbare Sauerstofftherapie (HBO); Hyperbaric oxygen therapy (HOT)
- BUA005 Magnetresonanz-Tomographie der weiblichen Brust (MRM); Magnetic resonance imaging (MRI) of the female breast
- BUA006 Akupunktur; Acupuncture

BUA007	Photodynamische Therapie (PDT) mit Verteporfin bei altersabhängiger feuchter Makuladegeneration mit subfoveolären klassischen chorioidalen Neovaskularisationen; Photodynamic therapy
BUA008	Uterus-Ballon-Therapie; Uterine ballon therapy
BUA009	Ambulante Balneophototherapie - 1. Nicht-synchrone Photosoletherapie, 2. Bade-PUVA; Balneophototherapy in ambulatory care 1. non-simultaneously photosoletherapy 2. external (bath) PUVA therapy
BUA010	Osteodensitometrie; Osteodensitometry
BUA011	Behandlung mit ionisiertem Sauerstoff; Treatments using ionized oxygen
BUA012	Sauerstoff-Mehrschritt-Therapie nach Ardenne (SMT); Multistep oxygen therapy according to Ardenne
BUA013	Extrakorporale Stoßwellentherapie (ESWT); Extracorporeal shock wave therapy
BUA014	Niedrigdosierter, gepulster Ultraschall (NGU); Therapies using low-intensity pulsed ultrasound
BUA015	Positronen-Emissions-Tomographie (PET); Positron emission tomography (PET)
BUA016	Atlastherapie nach Arlen; Treatment of the atlas vertebra according to Arlen
BUA017	Selektive UVA1-Bestrahlung; Selective UVA1 irradiation
BUA018	Behandlung mit CO <sub>2</sub> -Insufflationen (Quellgasbehandlung); Treatments using carbon dioxide insufflations
BUA019	Hämatogene Oxidationstherapie (HOT); Hematogenous oxydation therapy
BUA020	Oxyvenierungstherapie nach Regelsberger synonym u. a.: intravenöse Sauerstoffinsufflation, Sauerstoff- Infusions-Therapie (SIT), Komplexe intravenöse Sauerstofftherapie (KIS); Intravenous oxygen therapy according to Regelsberger
BUA021	Ozon-Therapie, Ozon-Eigenbluttherapie, Sauerstoff-Ozon-Eigenbluttherapie, Oxyontherapie, Hyperbare Ozontherapie; Ozone therapy, oxygen-ozone autohaemotherapy, oxyon therapy, hyperbaric ozone therapy
BUA022	Ultraviolettbestrahlung des Blutes (UVB); Irradiation of blood with ultraviolet waves
BUA023	Screening auf Gestationsdiabetes; Screening for gestational diabetes
DAHTA001	Stellenwert von Ultraschallverfahren im Rahmen der Osteoporoseversorgung (Früherkennung des Frakturrisikos). Ökonomischer Kurz-HTA; The value of ultrasound diagnostic technologies in the prevention of fractures in patients with osteoporosis
DAHTA003	Bewertung unterschiedlicher Hygienekonzepte zur Kontrolle von MRSA (Methicillin-resistente Staphylococcus aureus); Evaluation of different concepts of hygiene to control MRSA (methicillin-resistant staphylococcus aureus)
DAHTA006	Bewertung der Möglichkeiten und Verfahren zur Aufbereitung medizinischer Einwegprodukte; Assessment for the possibility and technologies of medical single-use products

- DAHTA007 Glycoprotein IIb/IIIa-Rezeptorantagonisten in der Therapie akuter koronarer Syndrome - ein gesundheitsökonomischer HTA-Bericht; Glycoprotein IIb/IIIa receptor blocker in the treatment of coronary syndrome patients economic efficiency
- DAHTA008 Medizinische Effektivität beim Einsatz von GP- IIb/IIIa-Rezeptorantagonisten in der Therapie von akuten Koronarsyndromen; Effectiveness of glycoprotein IIb/IIIa receptor antagonists in the treatment of acute coronary syndroms (ACS)
- DAHTA009 Antivirale Therapie bei Patienten mit chronischer Hepatitis C in Deutschland - medizinische und ökonomische Evaluation der initialen Kombinationstherapie mit Interferon/Peginterferon und Ribavirin; Antiviral therapy for patients with chronic hepatitis C in Germany - evaluation of effectiveness and cost-effectiveness of initial combination therapy with interferon/peginterferon plus ribavirin
- DAHTA011 Ultraschall in der Schwangerschaft. Beurteilung der routinemäßigen Schwangerschafts-ultraschalluntersuchungen unter Maßgabe der Mutterschaftsrichtlinien; Ultrasound in pregnancy - Evaluation of the routine ultrasound examinations according to the The German Maternity Code
- DAHTA013 Stroke Units - Update des HTA Berichts „Die Evaluation von Stroke Units als medizinische Technologie“; Stroke Units: update of HTA report "Assessment of stroke units as medical technology"
- DAHTA015 Sonographische Diagnostik beim akuten Abdomen bei Kindern und Erwachsenen; Sonographic diagnosis of acute abdomen in children and adults
- DAHTA017 Minimal-invasive Verfahren zur Behandlung des Bandscheibenvorfalles; Minimally invasive surgical procedures for the treatment of lumbar disc herniation
- DAHTA018 Wurzelbehandlung an Molaren; Endodontic Treatment of Molars
- DAHTA019 Pap-Test zum Screening auf Zervixkarzinom. Einfluss verschiedener Untersuchungsintervalle; PAP-Test for the screening of cervical carcinoma: impact of different examination intervals
- DAHTA021 Bestimmung der Albuminausscheidung im Urin bei Diabetikern zur Vorsorge und Kontrolle der diabetischen Nephropathie; Determination of albuminuria in the urine of diabetics for prevention and control of diabetic nephropathy
- DAHTA022 Medizinische und ökonomische Effektivität der Pneumokokkenimpfung für Säuglinge und Kleinkinder; Effectiveness and Cost-Effectiveness of general immunisation of infants and young children with the heptavalent conjugated pneumococcal vaccine
- DAHTA023 Bluthochdruckleitlinien und ihre Auswirkungen auf das Gesundheitssystem; Hypertension Guidelines and their Effects on the Health System
- DAHTA024 Neuraminidasehemmer in der Therapie und Postexpositionsprophylaxe der Influenza; Neuraminidase inhibitors in the treatment and post exposition prophylaxis of the influenza
- DAHTA025 Kognitives Training bei Demenzen und andere Störungen mit kognitiven Defiziten; Cognitive training for dementia

DAHTA026	Senkung der Restenoserate durch Einsatz beschichteter Stents bei koronarer Herzkrankheit. Systematische Übersicht zur medizinischen Wirksamkeit und gesundheitsökonomische Bewertung zum Vergleich von beschichteten gegenüber unbeschichteten Stents; Coated stents to prevent restenosis in coronary heart disease
DAHTA028	Screening auf urogenitale Chlamydia trachomatis-Infektionen; Screening on urogenital Chlamydia trachomatis
DAHTA029	Extrakorporale artifizielle Leberunterstützungssysteme bei akutem Leberversagen oder einer akuten Dekompensation eines chronischen Leberleidens; Extracorporal hemodialysis with acute or decompensated chronicl hepatic failure
DAHTA029	Thrombozytenaggregationshemmer zur Primär- und Sekundärprävention des ischämischen Schlaganfalls; Platelet aggregation inhibitors in primary and secondary prevention of ischemic stroke
DAHTA030	Die elastisch stabile intramedulläre Nagelung bei instabilen kindlichen Unterarmschaffraktururen; Elastic robust intramedullary nailing for forearm fracture in children
DAHTA031	Bedeutung der intensivierten Pflege; Value of intensified nursing
DAHTA032	Monitoring von Herzfunktionen mit Telemetrie; Monitoring heart functions using telemetry
DAHTA033	Bypassmaterialien in der Gefäßchirurgie; Bypass materials in vascular surgery
DAHTA034	Therapie der mittelschweren und der schweren Psoriasis; Therapy of moderate and severe psoriasis
DAHTA035	Einsatz von Statinen in der Primärprävention; The use of statins in primary prevention
DAHTA036	Koronarkalkbestimmung mit CT-Verfahren bei asymptomatischen Risikopatienten; Computed Tomography for the measurement of coronary calcification in asymptomatic risk patients
DAHTA037	Prävention rezidivierender Rückenschmerzen- Präventionsmaßnahmen in der Arbeitsplatzumgebung; Prevention of relapsing backache
DAHTA038	Orchiektomie versus medikamentöse Therapie mit LH-RH-Analoga zur Behandlung des fortgeschrittenen Prostatakarzinoms; Orchiectomy versus medical therapy with LH-RH analogues for the treatment of advanced prostatic carcinoma
DAHTA039	Antioxidative Vitamine zur Prävention kardiovaskulärer Erkrankungen nach Nierentransplantation und bei chronischer Niereninsuffizienz; Antioxidative vitamins for prevention of cardiovascular disease for patients after renal transplantation and patients with chronic renal failure
DAHTA040	Bypassoperation am schlagenden Herzen im Vergleich zur Operation mit Unterstützung durch die Herz-Lungen-Maschine; Bypass operation of the beating heart in comparison to the operation with support of the heart-lung machine
DAHTA041	Verfahren zur Steigerung der Teilnehmerate an Krankheitsfrüherkennungsprogrammen; Interventions for increasing uptake in screening programmes
DAHTA042	Therapie des unspezifischen Tinnitus ohne Ursache; Therapy of unspecific tinnitus without organic cause

- DAHTA043 Misteltherapie als begleitende Behandlung zur Reduktion der Toxizität der Chemotherapie maligner Erkrankungen; Mistletoe treatments for minimising side effects of anticancer chemotherapy
- DAHTA044 Therapie der Neurodermitis; Therapy of atopic eczema
- DAHTA045 Hörscreening für Neugeborene - Update; Screening of the hearing of newborns - Update
- DAHTA046 Welche Auswirkung hat die Kataraktoperation auf das Entstehen oder das Fortschreiten einer altersbedingten Makuladegeneration (AMD)?; What Effects has the Cataract Surgery on the Development and Progression of Age-Related Macular Degeneration (AMD)?
- DAHTA047 Laparoskopische vs. offene Appendektomie - Systematische Übersicht zur medizinischen Wirksamkeit und gesundheitsökonomische Analyse; Laparoscopic versus open appendectomy - A systematic review and health economic analysis
- DAHTA048 Erkennungsgüte und Kosteneffektivität von Screeningverfahren zur Erfassung von primären Offenwinkelglaukomen; Validity and Cost-Effectiveness of Methods for Screening of Primary Open Angle Glaucoma
- DAHTA049 Wertigkeit der optischen Kohärenztomographie im Vergleich zur Fluoreszenzangiographie in der Diagnostik der altersbedingten Makuladegeneration (AMD); Evaluation of optical coherence tomography in the diagnosis of age related macula degeneration compared with fluorescence angiography
- DAHTA050 Hormone zur Therapie von Beschwerden im Klimakterium und zur Primärprävention von Erkrankungen in der Postmenopause; Hormones for therapy of climacteric afflictions
- DAHTA051 Langzeitsubstitutionsbehandlung Opioidabhängiger; Long term substitution treatment (maintenance treatment) of opioid dependent persons
- DAHTA052 Episiotomie bei der vaginalen Geburt; Episiotomy in Vaginal Birth: a Health Technology Assessment
- DAHTA053 Drotrecogin alfa (aktiviert) bei der Behandlung der schweren Sepsis; Efficacy and effectiveness of recombinant human activated protein C in severe sepsis of adults
- DAHTA054 Nutzen und Risiken hormonaler Kontrazeptiva bei Frauen; Benefits and risks of hormonal contraception for women
- DAHTA055 Welchen medizinischen und gesundheitsökonomischen Nutzen hat die Kolposkopie als primäres Screening auf das Zervixkarzinom?; Review on the medical and health economic evidence for an inclusion of colposcopy in primary screening programs for cervical cancer
- DAHTA056 Dekubitusprophylaxe und -therapie; Decubital prophylaxis and therapy
- DAHTA058 HPV-DNA-Diagnostik zur Zervixkarzinomfrüherkennung; Assessment of effectiveness and cost-effectiveness of HPV testing in primary screening for cervical cancer
- DAHTA059 Künstliche Ventrikel bei fortgeschrittener Herzinsuffizienz; Ventricular Assist Devices for Heart Failure

- DAHTA060 Topische antientzündliche Behandlung der Neurodermitis im Kindesalter; Corticosteroid therapy in the treatment of pediatric patients with atopic dermatitis
- DAHTA061 Wie ist der derzeitige Stellenwert der Homozysteinbestimmung im Blut als Risikofaktor für die koronare Herzkrankheit (KHK)?; The role of Homocysteine as a predictor for coronary heart disease.
- DAHTA062 Die Masern-Mumps-Röteln-Impfung aus gesundheitspolitischer und ökonomischer Sicht; The Measels-Mumps-Rubella Vaccination from a health political and economical point of view
- DAHTA065 Maßnahmen zur Verbesserung der Compliance bzw. Adherence in der Arzneimitteltherapie mit Hinblick auf den Therapieerfolg; Interventions for enhancing medication compliance/adherence with benefits in treatment outcomes
- DAHTA066 Mundgesundheit nach kieferorthopädischer Behandlung mit festsitzenden Apparaten; Oral health with fixed appliances orthodontics
- DAHTA067 Vergleich der laparoskopischen Narbenhernioplastik und der konventionellen Operation mit und ohne Netzeinlage – Effektivität und Kostennutzenrelation; Health Technology Assessment of laparoscopic compared to conventional surgery with and without mesh for incisional hernia repair regarding safety, efficacy and cost-effectiveness
- DIMDI004 Biochemisches Screening für fetale Chromosomenanomalien und Neuralrohrdefekte - eine Verfahrensbewertung; Biochemical screening for fetal chromosomal abnormalities and neural tube defects - a technology assessment
- DIMDI008 Evaluation präoperativer Routinediagnostik (Röntgenthorax, EKG, Labor) vor elektiven Eingriffen bei Erwachsenen; Preoperative routine diagnostics (chest radiography, ECG, and laboratory tests) for elective surgery
- DIMDI010 Operative Eingriffe an der lumbalen Wirbelsäule bei bandscheibenbedingten Rücken- und Beinschmerzen - eine Verfahrensbewertung; Surgical treatment of lumbar spine for leg and back pain caused by the disc syndrome - a health technology assessment
- DIMDI011 Bewertung von Verfahren zur Diagnostik der akuten Sinusitis maxillaris bei Erwachsenen; Diagnosis of acute maxillary sinusitis in adults - systematic review and meta-analysis
- DIMDI013 Die Rolle der Osteodensitometrie im Rahmen der Primär-, Sekundär- und Tertiärprävention/Therapie der Osteoporose; The role of osteodensitometry in primary, secondary, and tertiary prevention, and in treatment of osteoporosis
- DIMDI014 Evaluation arthroskopischer Operationen bei akuten und degenerativen Meniskusläsionen; The evaluation of arthroscopic surgery in acute and degenerative meniscus lesions
- DIMDI015 Die Evaluation von Stroke Units als medizinische Technologie; Assessment of stroke units as medical technology
- DIMDI016 Hochdosis-Chemotherapie mit autologer Stammzelltransplantation zur Therapie des metastasierenden Mammakarzinoms; High dose chemotherapy with autologous bone marrow transplantation in metastatic breast carcinoma

- DIMDI017 Vergleichende Effektivität und Differentialindikation von Ballondilatation (PTCA) versus Bypasschirurgie bei Ein- und Mehrgefäßerkrankungen der Herzkranzgefäße; Coronary artery bypass graft surgery versus percutaneous transluminal coronary angioplasty in the treatment of coronary vessel diseases
- DIMDI018 Hüftgelenkendoprothetik bei Osteoarthritis; Hip replacement with osteoarthritis
- DIMDI019 PSA-Screening beim Prostatakarzinom - systematischer gesundheitsökonomischer Review; PSA screening for prostate carcinoma - systematic health economic review
- DIMDI020 Ökonomische Evaluationen der Positronen-Emissions-Tomographie; Positron emission tomography - the economic efficacy
- DIMDI021 Das Schlaf-Apnoe-Syndrom - systematische Übersichten zur Diagnostik, Therapie und Kosten-Effektivität; Diagnosis and treatment of sleep apnea - systematic review of diagnostics, therapy, and cost-effectiveness
- DIMDI022 Stellenwert der Magnet-Resonanz-Tomographie im Rahmen der Versorgung von Patienten mit Rückenschmerzen - Kurz-HTA: Update einer Best-Evidence-Synthese; Evaluation of magnetic resonance imaging in health care of patients with back pain - short HTA report updating a best evidence synthesis
- DIMDI024 Die Wertigkeit der Streßechokardiographie in der Primärdiagnostik der koronaren Herzkrankheit; Quality rating of the stress echocardiography in the primary diagnostics of coronary artery disease
- DIMDI025 Knochen- und Knochenersatzmaterialien zur parodontalen Regeneration oder zum Knochenaufbau für Implantate - ein gesundheitsökonomisches HTA; Bone and bone substitute for periodontal rehabilitation and for bone reconstruction before implantation - systematic review of health economics
- DIMDI026 Die Verwendung des Excimer Lasers in der refraktiven Augenchirurgie; The use of the excimer laser in refractive surgery
- DIMDI027 Wertigkeit von Tumor-Nekrose-Faktor alpha-Antagonisten in der Behandlung der rheumatoiden Arthritis (Medizinischer Teil); Tumor necrosis factor alpha antagonists in the treatment of rheumatoid arthritis - medical efficacy
- DIMDI028 Dünnschichtpräparationen und computergestützte Untersuchungen von Zervixabstrichen - Medizinische Effektivität, gesundheitsökonomische Evaluation und systematische Entscheidungsanalyse; Liquid-based slide preparations and automated devices in examining cervix smears - systematic review of medical efficacy and health economics, and decision analysis
- DIMDI029 Stenting versus Ballondilatation bei koronarer Herzkrankheit systematische Übersicht zur medizinischen Effektivität; Stenting versus balloon angioplasty in coronary heart disease a systematic review of clinical effectiveness
- DIMDI030 Wertigkeit des Einsatzes der genotypischen und phänotypischen HIV-Resistenzbestimmung im Rahmen der Behandlung von HIV-infizierten Patienten; Evaluation of genotypic and phenotypic definition of resistance in the treatment of patients infected by HIV - medical efficacy and economic efficiency

- DIMDI031 Positronen-Emissions-Tomographie systematische Übersichten zur Wirksamkeit bei ausgewählten Indikationen; Positron emission tomography (PET) - systematic review of efficacy in selective medical indications
- DIMDI032 Wertigkeit des Einsatzes der intravaskulären Ultraschallbildgebung (IVUS) im Rahmen von diagnostischen und therapeutischen Herzkatheteruntersuchungen - ein HTA-Bericht zur medizinischen Effektivität; The value of intravascular ultrasound (IVUS) for diagnostic and therapeutic coronary angiography - a medical health technology assessment
- DIMDI033 Wertigkeit des Einsatzes der intravaskulären Ultraschallbildgebung (IVUS) im Rahmen von diagnostischen und therapeutischen Herzkatheteruntersuchungen - ein gesundheitsökonomischer HTA-Bericht; Value of intravascular ultrasound (IVUS) in diagnostic and therapeutic heart catheterisation in patients with coronary heart disease - an economic health technology assessment
- DIMDI034 Verfahrensbewertung der CT-Angiographie, MR-Angiographie, Doppler-Sonographie und Szintigraphie bei der Diagnose von Nierenarterienstenosen; Assessment of the diagnostic value of computed tomography (CT) angiography, magnetic resonance (MR) angiography, Doppler sonography and scintigraphy for detecting renal artery stenosis
- DIMDI035 Wertigkeit von Tumor-Nekrose-Faktor-alpha-Antagonisten in der Behandlung der Rheumatoiden Arthritis; Tumor necrosis factor alpha antagonists in the treatment of rheumatoid arthritis - economic efficiency
- DIMDI036 Ein Vergleich verschiedener chirurgischer Verfahren zur elektiven Leistenhernieoperation bei Erwachsenen - ein Health Technology Assessment; The open versus laparoscopic methods in surgery of inguinal hernias - a systematic review
- DIMDI037 Stenting versus Ballondilatation bei koronarer Herzkrankheit; Stenting versus balloon angioplasty for coronary heart diseases
- DIMDI038 Knochen- und Knochenersatzmaterialien zur parodontalen Regeneration und zum Knochenaufbau für Implantate - eine systematische Bewertung der medizinischen Wirksamkeit; Bone and bone substitute for periodontal rehabilitation and for bone reconstruction before implantation - systematic review of medical efficacy
- DIMDI040 Hörscreening für Neugeborene - ein Health Technology Assessment der medizinischen Effektivität und der ökonomischen Effizienz; Screening of the hearing of newborns - systematic review of medical efficacy and economic efficiency
- HTAI002 Magnetic Resonance Guided Focused Ultrasound Surgery for the Treatment of Uterine Fibroids; Magnetresonanz-gesteuerte Ultraschallchirurgie für die Behandlung von Uterusmyomen
- HTAI003 Klinische und ökonomische Evaluation des diagnostischen Verfahrens ‚Kapselendoskopie‘ zur Untersuchung des Dünndarms; Clinical and Economic Evaluation of the Diagnostic Method of Capsule Endoscopy for Examination of the Small Intestine
- HTAI004 Bewertung des Mehrnutzens der Therapie mit Glitazonen im Vergleich zur Therapie mit Metformin oder Sulfonylharnstoffen bei der Behandlung des Typ 2 Diabetes mellitus; Assessment of the Additional Benefits of Glitazones Compared to Metformin or Sulfonylureas in the Treatment of Type 2 Diabetes Mellitus

HTAI005	Health Technology Assessment: Heptavalenter Pneumokokkenkonjugat-Impfstoff (PCV7); Health Technology Assessment: Heptavalent pneumococcal conjugate vaccine (PCV7)
HTAI001	Klinische und gesundheitsökonomische Relevanz von Statinen in der medikamentösen Therapie der Hypercholesterinämie; Clinical relevance and significance to health economics of statins in the medical treatment of hypercholesterinaemia
KBV001	Kurzassessment Extrakorporale Photophorese (ECP); Short assessment extracorporeal photophoresis
IQWiG001	Clopidogrel versus Acetylsalicylsäure in der Sekundärprophylaxe vaskulärer Erkrankungen
IQWiG002	Kurzwirksame Insulinanaloga zur Behandlung des Diabetes mellitus Typ 1
IQWiG003	Kurzwirksame Insulinanaloga zur Behandlung des Diabetes mellitus Typ 2
IQWiG004	Fixe Kombinationen aus Kortikosteroiden und lang wirksamen Beta-2-Rezeptoragonisten zur inhalativen Anwendung bei Patienten mit Asthma bronchiale
IQWiG005	Leukotrien-Rezeptor-Antagonisten bei Patienten mit Asthma bronchiale
IQWiG006	Cholinesterasehemmer bei Alzheimer Demenz
IQWiG007	Nutzenbewertung nichtmedikamentöser Behandlungsstrategien bei Patienten mit Bluthochdruck: A-Gewichtsreduktion
IQWiG008	Bewertung des therapeutischen Nutzens von inhalativem Insulin
IQWiG009	Ultraschall-Screening in der Schwangerschaft: Testgüte hinsichtlich der Entdeckungsrates fetaler Anomalien
IQWiG010	Stammzelltransplantation bei den Indikationen Akute lymphatische Leukämie (ALL) und Akute myeloische Leukämie (AML) bei Erwachsenen
IQWiG011	Früherkennungsuntersuchung von Hörstörungen bei Neugeborenen
IQWiG012	Früherkennungsuntersuchung von Sehstörungen bei Kindern bis zur Vollendung des 6. Lebensjahres
IQWiG013	Stammzelltransplantation bei erworbener schwerer aplastischer Anämie
IQWiG014	Bewertung des therapeutischen Nutzens von Exenatide - Rapid Report
IQWiG015	Vakuumversiegelungstherapie von Wunden – Rapid Report
IQWiG016	Hyperbare Sauerstofftherapie bei Brandwunden
IQWiG017	Interstitielle Brachytherapie beim lokal begrenzten Prostatakarzinom
IQWiG018	Balneophototherapie

- IQWiG019 Nichtmedikamentöse lokale Verfahren zur Behandlung der benignen Prostatahyperplasie
- IQWiG020 Vakuumversiegelungstherapie von Wunden
- IQWiG021 Hyperbare Sauerstofftherapie bei der idiopathischen Femurkopfnekrose des Erwachsenen



Die systematische Bewertung medizinischer Prozesse und Verfahren, *Health Technology Assessment* (HTA), ist mittlerweile integrierter Bestandteil der Gesundheitspolitik. HTA hat sich als wirksames Mittel zur Sicherung der Qualität und Wirtschaftlichkeit im deutschen Gesundheitswesen etabliert.

Seit Einrichtung der Deutschen Agentur für HTA des DIMDI (DAHTA) im Jahr 2000 gehören die Entwicklung und Bereitstellung von Informationssystemen, speziellen Datenbanken und HTA-Berichten zu den Aufgaben des DIMDI.

Im Rahmen der Forschungsförderung beauftragt das DIMDI qualifizierte Wissenschaftler mit der Erstellung von HTA-Berichten, die Aussagen machen zu Nutzen, Risiko, Kosten und Auswirkungen medizinischer Verfahren und Technologien mit Bezug zur gesundheitlichen Versorgung der Bevölkerung. Dabei fallen unter den Begriff Technologie sowohl Medikamente als auch Instrumente, Geräte, Prozeduren, Verfahren sowie Organisationsstrukturen. Vorrang haben dabei Themen, für die gesundheitspolitischer Entscheidungsbedarf besteht.