

Aufbau einer Datenbasis
'Evaluation medizinischer Verfahren und Technologien'
in der Bundesrepublik Deutschland

PSA-Screening beim Prostatakarzinom

Systematischer gesundheitsökonomischer Review.

Entwicklung und Anwendung eines Instrumentariums zur systematischen Beschreibung
und Bewertung gesundheitsökonomischer Studien.

Verfasser:

Uwe Siebert, M.P.H. postgrad., M.Sc.¹; Dr. Nikolai Mühlberger, M.P.H. postgrad.²;
Corinne Behrend, M.P.H. postgrad.¹; Prof. Dr. Jürgen Wasem³

¹ Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie,
Postgradualer Studiengang Öffentliche Gesundheit und Epidemiologie, Ludwig-
Maximilians-Universität München

² GSF-Forschungszentrum für Umwelt und Gesundheit, Institut für Medizinische
Informatik und Systemforschung, Neuherberg

³ Lehrstuhl für Allgemeine Betriebswirtschaftslehre und Gesundheitsmanagement,
Ernst-Moritz-Arndt-Universität Greifswald

München, Dezember 1999

Inhaltsverzeichnis

A Abstract	1
B Executive Summary	3
C Hauptdokument	12
C.1 Policy Question	13
C.2 Einführung/Hintergrund.....	15
C.2.1 Methoden gesundheitsökonomischer Evaluation.....	15
C.2.2 Screening als medizinische Technologie	30
C.2.3 Beschreibung der Zielkondition "Prostatakarzinom".....	39
C.2.4 Beschreibung der Technologie "PSA-Bestimmung".....	52
C.2.5 Beschreibung der Intervention "PSA-Screening beim Prostatakarzinom"	56
C.3 Forschungsfragen.....	68
C.4 Methodik	69
C.4.1 Entwicklung des Bewertungsinstrumentariums für gesundheitsökonomische Studien	71
C.4.2 Zielpopulation und Ergebnisparameter	76
C.4.3 Datenquellen und Recherche.....	79

C.4.4	Bewertung der Information/Studienqualität	82
C.4.5	Datenextraktion ökonomischer Parameter	82
C.5	Ergebnisse	86
C.5.1	Bewertungsinstrumentarium für gesundheitsökonomische Studien	86
C.5.2	Ergebnis der Literaturrecherche	95
C.5.3	Standardisierte Berichte zu berücksichtigten Publikationen und qualitative Informationssynthese	98
C.5.4	Bewertung der Information/Studienqualität	112
C.5.5	Qualitative Informationssynthese ökonomischer Parameter	126
C.6	Diskussion	135
C.6.1	Bewertungsinstrumentarium für gesundheitsökonomische Studien	135
C.6.2	PSA-Screening beim Prostatakarzinom	144
C.6.3	Überlegungen zur Übertragbarkeit der Ergebnisse nach Deutschland	159
C.7	Schlussfolgerungen und Empfehlungen	170
C.8	Quellenangaben	173

D Anhang	184
D.1 Anhang 1: Dokumentationen	184
D.2 Anhang 2: Deskriptionskatalog	187
D.3 Anhang 3: Erläuterungen zu den gesundheitsökonomischen Kriterienkatalogen.....	208
D.4 Anhang 4: Standardisierte Berichte	219

Verzeichnis der Tabellen

Tabelle 1:	Methoden der Evaluation von Leistungen der Gesundheitsversorgung.....	23
Tabelle 2:	Kriterien zur Beurteilung der Effektivität von Screening-Programmen.....	32
Tabelle 3:	Staging beim Prostatakarzinom nach dem Whitmore-Jewett- und TNM-System.....	44
Tabelle 4:	Grading beim Prostatakarzinom nach Gleason (1992).....	47
Tabelle 5:	Panel zur Entwicklung eines Instrumentariums für die Bewertung von gesundheitsökonomischer Studien im Rahmen von Health Technology Assessment	71
Tabelle 6:	In die Evaluation eingeschlossene gesundheitsökonomische Studientypen.....	78
Tabelle 7:	Internet-Seiten der berücksichtigten HTA vornehmenden bzw. unterstützenden Institutionen	80
Tabelle 8:	Kostenarten in Screening-Programmen für Tumorerkrankungen.....	84
Tabelle 9:	Transparenzkatalog. Kriterienkatalog zur Bewertung der formalen Transparenz und Nachvollziehbarkeit gesundheitsökonomischer Studien (entwickelt im Konsensusverfahren von den gesundheitsökonomischen Projektgruppen München, Hannover, Ulm)	88
Tabelle 10:	Qualitätskatalog. Kriterienkatalog zur Bewertung der inhaltlichen und methodischen Studienqualität gesundheitsökonomischer Studien (entwickelt im Konsensusverfahren von den gesundheitsökonomischen Projektgruppen München, Hannover, Ulm)	92

Tabelle 11:	Ausschluss von Referenzen bei der manuellen Vorauswahl (Datenbanken EMBASE und MEDLINE EXPRESS). Stand 15.9.1998	95
Tabelle 12:	In die gesundheitsökonomische Evaluation eingeschlossenen Publikationen	97
Tabelle 13:	In der gesundheitsökonomischen Evaluation nicht berücksichtigte Publikationen mit Angabe des Ausschlussgrundes	98
Tabelle 14:	Standardisierte Berichte. Dokumentationsstruktur für Original- und Synthesestudien	99
Tabelle 15:	Extrahierte Merkmale und Ergebnisparameter aus den berücksichtigten gesundheitsökonomischen Untersuchungen zum Screening des Prostatakarzinoms	103
Tabelle 16:	Beurteilung der PSA-Studien nach dem Erfüllungsgrad der Transparenzkriterien.....	113
Tabelle 17:	Beurteilung der PSA-Studien nach dem Erfüllungsgrad der Transparenzdimensionen	117
Tabelle 18:	Beurteilung der PSA-Studien nach dem Erfüllungsgrad der Bewertungskriterien	120
Tabelle 19:	Beurteilung der PSA-Studien nach dem Erfüllungsgrad der Bewertungsdimensionen	124
Tabelle 20:	Übersicht über die berücksichtigten Kostenarten in den evaluierten gesundheitsökonomischen Studien	127
Tabelle 21:	Kosten und Kosten-Effektivität der untersuchten PSA-Screening-Verfahren. Ohne Währungskonversion.	130
Tabelle 22:	Kosten und Kosten-Effektivität der untersuchten PSA-Screening-Verfahren. Währungskonversion über Kaufpreisparitäten im Gesundheitssektors des Jahres 1993 (Quelle: OECD Health Data 1997).	132

Tabelle 23: Konsens und Dissens zu methodischen Fragen der ökonomischen Evaluation.....	138
Tabelle 24: Deskriptionskatalog. Kriterienkatalog zur deskriptiven Erfassung und Dokumentation von Merkmalen und Inhalten gesundheitsökonomischer Studien (entwickelt im Konsensusverfahren von den gesundheitsökonomischen Projektgruppen München, Hannover, Ulm)	187

Verzeichnis der Abbildungen

Abbildung 1: Verzahnung der Entwicklung des Bewertungsinstrumentariums mit dessen praktischer Anwendung bei der Bewertung der gesundheitsökonomischen Studien zum PSA-Screening beim Prostatakarzinom.....	70
--	----

Verzeichnis der Abkürzungen

2. GKV-NOG	Zweites Neuordnungsgesetz der gesetzlichen Krankenversicherung
AHCPR	Agency for Health Care Policy and Research
AHTAC	Australian Health Technology Advisory Committee
AUG	Ausscheidungsurogramm
BCOHTA	British Columbia Office of Health Technology Assessment
BMAE	Bewertungsmaßstab für kassenärztliche Leistungen
BMG	Bundesministerium für Gesundheit
BPH	Benigne Prostatahyperplasie
CCOHTA	Canadian Coordinating Office for Health Technology Assessment
CETS	Conseil d'Évaluation des Technologies de la Santé du Québec
CPT-4	Current Procedural Terminology, 4 th edition
CT	Computertomographie
DARE	Database of Abstracts of Reviews of Effectiveness
DEALE	Declining Exponential Approximation of Life Expectancy
DRG	Diagnosis Related Groups
DRU	Digitale rektale Untersuchung (engl.: DRE - digital rectal examination)
EBM	Einheitlicher Bewertungsmaßstab
EGO	Ersatzkassen-Gebührenordnung
ELISA	Enzyme-linked Immuno Sorbent Assay

HRQL	Health-related Quality of Life
HTA	Health Technology Assessment
HYE	Healthy Year Equivalent
IRMA	Immuno Radiometric Assay
ISTAHC	International Society of Technology in Health Care
NCCHTA	National Coordinating Centre for Technology Assessment
NEED	NHS Economic Evaluation Database
NHS	National Health Services (GB)
NMR (MTR)	Kernspinresonanztomographie
OTA	Office of Technology Assessment, Congress of the United States
PIN	Prostatische intraepitheliale Neoplasie
ProPSA	Prospective Payment System Bill
PSA	Prostata-spezifisches Antigen
PSA-D	PSA-Dichte
PSA-V	PSA im Zeitverlauf (PSA-velocity)
QALY	Quality-adjusted Life Year
Q-PSA	Quotient aus freiem und gesamtem PSA
RBRVS	Resource Based Relative Value Score (Medicare)
RCT	Randomized Clinical Trial
RIA	Radio Immuno Assay
RT-PCR	Reverse Transkriptase Polymerase Kettenreaktion
SBU	The Swedish Council on Technology Assessment in Health Care

SGB V	Sozialgesetzbuch fünftes Buch
TNM	Tumor - Nodi (Lymphknoten) - Metastasen
TRUS	Transrektale Ultrasonographie
TUR-P	Transurethrale Resektion der Prostata

A Abstract

PSA-Screening beim Prostatakarzinom. Systematischer gesundheitsökonomischer Review

Uwe Siebert, Nikolai Mühlberger, Corinne Behrend, Jürgen Wasem

Fragestellung. Das Prostatakarzinom ist in Deutschland mit etwa 11% (1996) die dritthäufigste tumorbedingte Mortalitätsursache bei Männern und besitzt eine wichtige medizinische, gesellschaftliche und ökonomische Bedeutung. Obwohl der Leistungskatalog der Krankenkassen dies nicht vorsieht, ist das Screening mittels der PSA-Bestimmung de facto in Deutschland bereits verbreitete Praxis.

Ziel der Arbeit war (i) die Entwicklung eines Instrumentariums zur systematischen Bewertung gesundheitsökonomischer Studien und (ii) die Anwendung dieses Instrumentariums zur Beurteilung der gesundheitsökonomischen Evidenz bezüglich der medizinischen Technologie PSA-Screening beim Prostatakarzinom.

Methodik. In einem gemeinsamen konsensusbildenden Prozess wurde von einem achtköpfigen Panel aus den drei gesundheitsökonomischen Projektgruppen München, Hannover und Ulm unter der Federführung Münchens ein Instrumentarium zur Beurteilung gesundheitsökonomischer Studien entwickelt.

Zur Erschließung des gesundheitsökonomischen Datenmaterials zum Thema "PSA-Screening bei Prostatakarzinom" wurde eine Literaturrecherche in den Datenbanken MEDLINE, EMBASE, DARE, CC, Econ Lit, PSYNDEXplus, sociofile, HealthSTAR, CancerLit, SOMED, und NEED durchgeführt, die gesundheitsökonomische Originalpublikationen und Reviews einschloss. Die Recherche wurde ergänzt durch Referenzlisten, Informationen von den HTA vornehmenden Einrichtungen und "graue" Literatur.

Ergebnisse

a) Bewertungsinstrumentarium. Das vom Panel der gesundheitsökonomischen Projektgruppen entwickelte Bewertungsinstrumentarium besteht aus (i) einem Deskriptionskatalog zur systematischen Erfassung von Einzelinformationen, (ii) einem Transparenzkatalog zur Beurteilung der formalen Transparenz und

Nachvollziehbarkeit der angewendeten Methoden und berichteten Ergebnisse und (iii) einem Qualitätskatalog zur Beurteilung der inhaltlichen und methodischen Studienqualität. Der Qualitätskatalog erwies sich als praktikables und überlegenes Instrument. Voraussetzung für seine Anwendung ist die Expertise eines interdisziplinären Gutachterteams. Ferner wurde ein Raster zur standardisierten Berichterstattung von gesundheitsökonomischen Studien erstellt und als Datenbank aufbereitet.

b) PSA-Screening beim Prostatakarzinom. Insgesamt erfüllten 16 Publikationen die Einschlusskriterien der Literaturrecherche, darunter 14 Primärstudien und zwei Reviews. Die Studien weisen eine große Heterogenität auf bezüglich der Wahl der zu vergleichenden Screening-Strategien, des Zeithorizontes, der berücksichtigten Kostenarten und weiteren Rahmenbedingungen. Die medizinische Evidenz des PSA-Screenings ist umstritten und die ermittelten Kosten-Effektivitäts-Relationen sind heterogen. Die für den Entscheidungskontext relevantesten Studien haben als Ergebnis einen geringen positiven inkrementellen Effekt des Screenings auf die Restlebenserwartung und einen negativen Effekt auf die qualitätsadjustierten Lebenserwartung. Insgesamt sind die Autoren-Schlussfolgerungen bezüglich der Empfehlung eines PSA-Screenings kontrovers. Eine Übertragbarkeit der Ergebnisse auf Deutschland ist u.a. aufgrund mangelhafter interner Validität, fehlender deutscher Präferenzdaten und nicht spezifizierter Mengengerüste, fehlender Aussagen zu Kosten und der Kosten-Effektivität eines wiederholten Screenings und der Nichtberücksichtigung relevanter Kosten nicht ohne weiteres möglich.

Schlussfolgerungen. Das im Konsens erarbeitete Bewertungsinstrumentarium erwies sich als praktikabel und themenübergreifend anwendbar. Voraussetzung für seine Anwendung ist die Expertise eines interdisziplinären Gutachterteams.

Bei einem PSA-Massen-Screening wäre mit zusätzlichen Kosten für das deutsche Gesundheitssystem bei fraglichen Gewinnen an Lebensjahren und Lebensqualität zu rechnen.

Aufgrund der methodischen Probleme bei der Übertragbarkeit halten wir als effiziente methodische Vorgehensweise die Anpassung extern entwickelter entscheidungsanalytischer Modelle (Entscheidungsbäume und Markov-Modelle) hoher Qualität und Flexibilität an die bundesdeutschen Verhältnisse mit anschließender Re-Analyse für notwendig. Dieses Vorgehen soll dem Entscheidungsträger innerhalb eines kalkulierbaren Zeitrahmens anwendungsorientierte Ergebnisse zu Allokationsentscheidungen zur Verfügung stellen.

B Executive Summary

PSA-Screening beim Prostatakarzinom.

Systematischer gesundheitsökonomischer Review

Uwe Siebert, Nikolai Mühlberger, Corinne Behrend, Jürgen Wasem

Fragestellung

Das Prostatakarzinom ist in Deutschland mit etwa 11% (1996) die dritthäufigste tumorbedingte Mortalitätsursache bei Männern und gleichzeitig die häufigste Todesursache unter den urologischen Tumoren. Im Jahr 1996 verstarben in Deutschland 11 916 Männer an dieser Erkrankung¹³⁶. Aufgrund der hohen Prävalenz in den höheren Altersgruppen würde ein Massen-Screening zu einer enormen finanziellen Belastung der entsprechenden Kostenträger führen. Damit kommt dem Prostatakarzinom eine wichtige medizinische, gesellschaftliche und ökonomische Bedeutung und eine hohe Public-Health-Relevanz zu.

Das Verfahren zur Bestimmung des prostataspezifischen Antigens (PSA-Test) wurde 1994 als Gebührenziffer in den EBM aufgenommen. Bezüglich des Prostatakarzinoms besteht jedoch Abrechnungsfähigkeit nur bei einer Anwendung zur Bestätigung und Monitoring von Malignität. Ein routinemäßiger Einsatz im Rahmen eines Screening-Verfahrens ist bisher nicht vorgesehen.

De facto ist allerdings in der Praxis eine Zunahme des Einsatzes des PSA-Tests als Verfahren zur Diagnose und Früherkennung von Prostatakarzinomen festzustellen. Auch kommen die internationale Literatur und Guidelines verschiedener Organisationen zu kontroversen Aussagen bezüglich des medizinischen Nutzens eines solchen Screenings. Da diesem Thema aufgrund der hohen Prävalenz insbesondere in den hohen Altersgruppen ferner eine große ökonomische Bedeutung zukommt, stellt sich die Frage, wie die PSA-Bestimmung in Verbindung mit einem routinemäßigen Screening für Prostatakarzinom bezüglich des kurz- und langfristigen medizinischen Nutzens, den Kosten und der Relation zwischen medizinischem Nutzen und Kosten zu bewerten ist. Dabei sind immer auch kurz- und langfristige Risiken und die gesundheitsbezogene Lebensqualität aller Screening-Teilnehmer einzubeziehen.

Die vorliegende Arbeit verfolgt zwei grundsätzliche Ziele:

- Die Entwicklung eines Instrumentariums zur systematischen Bewertung gesundheitsökonomischer Studien
- Anwendung dieses Instrumentariums zur Beurteilung der gesundheitsökonomischen Evidenz bezüglich der medizinischen Technologie PSA-Screening beim Prostatakarzinom

Dabei wurde besonderer Wert auf das Aufzeigen methodischer Probleme bei der Technologiefolgenabschätzung von Screening-Verfahren gelegt.

Methodik

In einem gemeinsamen konsensusbildenden Prozess wurde von einem achtköpfigen Panel aus den drei gesundheitsökonomischen Projektgruppen München, Hannover und Ulm unter der Federführung Münchens ein Instrumentarium zur Beurteilung gesundheitsökonomischer Studien entwickelt, welches zum einen die standardisierte Erfassung und Deskription der Studieninhalte und zum anderen die systematische Bewertung der Transparenz und der inhaltlichen und methodischen Studienqualität als Ziel hatte. Das Bewertungsinstrumentarium sollte von jeder der drei Projektgruppen exemplarisch anhand von einem Beispielthema angewandt werden:

Projektgruppe München: PSA-Screening beim Prostatakarzinom.

Projektgruppe Hannover: Autologe Knochenmarktransplantation und hochdosierte Chemotherapie bei metastasierendem Mammakarzinom.

Projektgruppe Ulm: Ballondilatation (PTCA) vs. Stenting bei koronarer Herzkrankheit.

Im Rahmen des Vorhabens "Bestandsaufnahme, Bewertung und Vorbereitung der Implementation einer Datensammlung 'Evaluation medizinischer Verfahren und Technologien' in der Bundesrepublik Deutschland" sollte das bezüglich der Effektivität in einem separaten Bericht untersuchte Thema "PSA-Screening beim Prostatakarzinom" um die ökonomische Komponente ergänzt werden. Neben Parametern zum medizinischen Nutzen wurden deshalb als Ergebnisparameter die

Kosten und die Kosten-Effektivität der medizinischen Technologie "PSA-Screening beim Prostatakarzinom" untersucht.

Als ökonomische Ergebnisparameter wurden folgende Maßzahlen verwendet: (1) Kosten pro Screening-Teilnehmer, (2) Kosten pro identifiziertes Prostatakarzinom, (3) Kosten pro gerettetes Leben, (4) Kosten pro gewonnenes Lebensjahr, (5) Kosten pro gewonnenes qualitätsadjustiertes Lebensjahr (QALY).

Zur Erschließung des gesundheitsökonomischen Datenmaterials wurde eine Literaturrecherche in den Datenbanken MEDLINE, EMBASE, DARE, CC, Econ Lit, PSYNDEXplus, sociofile, HealthSTAR, CancerLit, SOMED, und NEED durchgeführt, die gesundheitsökonomische Originalpublikationen und Reviews einschloss (1986 bis August 1998). Neben den klassischen gesundheitsökonomischen Evaluationsstudien zur simultanen Betrachtung von Kosten und Konsequenzen wurden auch reine Kostenbetrachtungen eingeschlossen. Die Recherche wurde ergänzt durch eine Durchsicht der Referenzlisten aller zusammengetragenen Publikationen einschließlich der HTA-Berichte zum Themengebiet und relevanter Fachzeitschriften.

Einbezogen wurden ferner Informationen von den HTA vornehmenden Einrichtungen (AHCPR, AHTAC, BCOHTA, CCOHTA, CHEPA, Danish Institute for Health Technology Assessment, ECRI, ISTAHC, McMaster University, Ministère de la Santé et des Services sociaux du Québec, NCCHTA, OTA, SBU). Parallel dazu wurden 12 Einrichtungen und Organisationen in Deutschland angeschrieben, um Hintergrundinformationen zur Situation in Deutschland zu erhalten. Durch Befragung von Experten und Kooperationspartnern wurde versucht, das vorhandene Material um "graue" Literatur zu ergänzen.

Ausgeschlossen wurden "personal opinions", Abstracts ohne entsprechende Volltextpublikation und HTA-Berichte zum Themengebiet, sofern letztere nicht als Hintergrundinformation für Primärpublikationen mit eigenen Kostenberechnungen herangezogen wurden.

Ergebnisse

a) Bewertungsinstrumentarium

Im Rahmen der Bewertung medizinischer Technologien wurde von dem Panel der gesundheitsökonomischen Projektgruppen (München, Hannover, Ulm) im

Konsensusverfahren ein Instrumentarium zur Beurteilung gesundheitsökonomischer Studien entwickelt und erprobt. Dabei wurden sowohl die internationalen Standards und Guidelines zur Durchführung und Berichterstattung von gesundheitsökonomischen Studien berücksichtigt als auch die Spezifika in Deutschland bezüglich Terminologie, Methodik und der Relevanz verschiedener methodischer Kriterien anhand der deutschen Literatur und Expertise zu diesem Themenkomplex berücksichtigt.

Das in allen drei gesundheitsökonomischen Projektgruppen angewandte Bewertungsinstrumentarium umfasst drei Kriterienkataloge, welche jeweils die Bereiche Fragestellung/Evaluationsrahmen, Analysemethoden und Modellierung, Gesundheitseffekte, Kosten Diskontierung, Ergebnispräsentation, Behandlung von Unsicherheiten, Diskussion und Schlussfolgerungen abdecken:

- (1) den Deskriptionskatalog. Er dient der detaillierten und umfassenden Erfassung von Einzelinformationen. Er ist insbesondere als Schulungsinstrument zur Standardisierung und Systematisierung der spezifischen Methoden und Terminologie zu verstehen. Seine Kenntnis dient als Hilfe bei der Anwendung der beiden folgenden Kriterienkataloge.
- (2) den Transparenzkatalog. Er dient der Beurteilung der formalen Transparenz und Nachvollziehbarkeit der angewandten Methoden und berichteten Ergebnisse und Schlussfolgerungen.
- (3) den Qualitätskatalog. Er dient der Beurteilung der inhaltlichen und methodischen Studienqualität. Dabei wird diese im Hinblick darauf bewertet, ob die angewandten Methoden und die Ergebnisberichterstattung bezüglich der Fragestellung und der vorgegebenen Rahmenbedingungen der gesundheitsökonomischen Evaluation inhaltlich und methodisch adäquat gewählt, durchgeführt und dargestellt wurden.

Der Vergleich der Resultate nach der praktischen Anwendung von Transparenz- und Qualitätskatalog bei den ökonomischen Untersuchungen zum PSA-Screening zeigt, dass die Überprüfung formaler Transparenzkriterien nicht ausreicht, um auf die kontextabhängige interne Validität einer Studie und damit auf deren methodisch-inhaltlichen und entscheidungsrelevanten Informationswert schließen zu können. Für letzteres ist die Anwendung des Qualitätskataloges eine notwendige Bedingung.

Für die Anwendung des Qualitätskataloges ist ein interdisziplinäres Gutachterteam Voraussetzung, das Expertise in folgenden Bereichen abzudecken hat: Medizin,

Gesundheitsökonomie, klinische Epidemiologie, Biometrie, formale Entscheidungsanalyse und Lebensqualitätsmessung.

In Anlehnung an internationale Datenbanken zur Bewertung von medizinischen Technologien wurde von der Münchner Projektgruppe ein den Spezifika gesundheitsökonomischer Studien gerecht werdendes Raster (standardisierte Berichte) zur standardisierten Berichterstattung gesundheitsökonomischer Studien entwickelt und beispielhaft für das Thema "PSA-Screening beim Prostatakarzinom" eingesetzt. Dieses wurde in einem zweiten Schritt zu einer Datenbank komprimiert, welche zu jeder aufgenommenen Studie die schnelle, gezielte und systematische Extraktion der wichtigsten Einzelinformationen zu Fragestellung/verglichene Screening-Strategien, Evaluationsrahmen, Zielpopulation, Methodik, Ergebnissen zu Effektivität, Kosten und Kosteneffektivität, Diskontierungsraten, Behandlung von Unsicherheiten, Autorenschlussfolgerungen und Limitierungen bezüglich interner Validität, Generalisierbarkeit und Übertragbarkeit zulässt.

b) PSA-Screening beim Prostatakarzinom

insgesamt erfüllten 16 Publikationen die Einschlusskriterien der Literaturrecherche. Davon waren 14 Primärstudien und zwei Reviews. Davon erhoben acht Studien eigene gesundheitsbezogenen Effekte (Screening-/Diagnosestudien) und acht Studien griffen bezüglich der Gesundheitseffekte auf Literaturdaten zurück (Synthesestudien). Als Hintergrunddokumente bzw. technische Reports zu den Primärstudien fungierten weitere sieben Dokumente. Ein Teil der Studien stellte hauptsächlich die Diagnose von Prostatakarzinom im asymptomatischen Stadium in den Vordergrund, ein anderer mehr die therapeutischen und gesundheitsbezogenen Konsequenzen der Diagnose. Als gesundheitsökonomischer Studientyp kommen elf Kosten-Effektivitäts-Analysen, eine Kosten-Konsequenzen-Analyse und zwei Kosten-Nutzwert-Analysen vor. Zwei Studien führten sowohl eine Kosten-Effektivitäts- als auch eine Kosten-Nutzen-Analyse durch. Ein großer Teil der Studien setzte entscheidungsanalytische Modelle ein.

Nur drei der Studien erreichten über 50% des möglichen Summenscores bei der Bewertung der inhaltlich-methodischen Studienqualität.

Als wesentliche Ergebnisse seien aufgeführt:

- Die Studien unterscheiden sich insbesondere bezüglich der Wahl der zu vergleichenden Screening-Strategien, der Festlegung der Basis-Annahmen, der Mess- und Bewertungskonzepte von Ressourcenverbrauch und gesundheitsbezogenen Effekten, der einbezogenen Kostenarten, der Zeitpunkte der Datenerhebung und damit der Stadien im Lebenszyklus (Entwicklungsstadien) der Technologien, der regionalen und nationalen Strukturen der Gesundheitsversorgung, der Darstellung der Ergebnisse in Form von Durchschnittswerten oder inkrementellen Werten.
- Die medizinische Evidenz des PSA-Screenings ist umstritten.
- Die qualitativ hochwertigen Studien mit langfristigem Zeithorizont kommen unter verschiedenen Annahmen zu einem positiven inkrementellen Effekt des Screenings auf die Restlebenserwartung, der je nach Altersgruppe und untersuchter Screening-Strategie im Bereich von 0 bis 16 Tagen pro Screening-Teilnehmer liegt⁸¹. Die Analyse der qualitätsadjustierten Lebenserwartung deutet allerdings in allen untersuchten Altersgruppen und Strategien darauf hin, dass ein Screening insgesamt mehr Schaden als gesundheitsbezogenen Nutzen bringt und höhere Kosten verursacht, als die herkömmliche Diagnosepraxis.
- Je nach Studie liegen die währungskonvertierten inkrementellen Kosten pro Screeningteilnehmer zwischen 87 DM und 988 DM und die Kosten pro identifiziertem Krebsfall zwischen 6 545 DM und 18 579 DM. Die Kosten pro gewonnenem Lebensjahr variieren in Abhängigkeit von den berücksichtigten Kostenarten und Altersgruppen zwischen 14 115 DM und 823 770 DM.
- Selbst bei der Restriktion auf die zahlreichen nordamerikanischen Studien zeigt sich eine große Heterogenität der Kosten-Effektivitäts-Quotienten. Dies weist auf das große Potential systematischer Fehler (Bias) und die daraus resultierende eingeschränkte interne Validität der Studienergebnisse hin.
- Drei der Autoren sprechen sich auf der Basis der jeweiligen Datenlage gegen ein PSA-Screening aus. Fünf Autoren legen weiteren Klärungsbedarf nahe. Vier Autoren erachten das Screening als realisierbar bzw. empfehlen es, während vier weitere auf die bei Früherkennungsprogrammen zu beachtende unterschiedliche diagnostische Kosten-Effektivität verschiedener Screening-Strategien hinweisen. Das Spektrum der Autoren-Schlussfolgerungen spiegelt

damit die Standpunkte in der derzeitigen Kontroverse bezüglich der Früherkennung beim Prostatakarzinom wider.

- Die Übertragbarkeit der Ergebnisse auf Deutschland ist aufgrund mangelhafter interner Validität der Studien, umstrittener medizinischer Effektivität, mangelnder Datenlage zu den Präferenzen und der Lebensqualität, nicht spezifizierter Mengengerüste, fehlender Aussagen zu Kosten und der Kosten-Effektivität eines wiederholten Screenings und der Nichtberücksichtigung relevanter Kosten nicht ohne weiteres möglich.

Schlussfolgerungen

Die Zielsetzung dieses Berichtes war die Entwicklung eines Bewertungsinstrumentariums für gesundheitsökonomische Studien und dessen exemplarische Anwendung am Thema "PSA-Screening beim Prostatakarzinom". Damit sollte der Prozess der medizinischen Technologiefolgenabschätzung um die ökonomische Komponente ergänzt werden.

a) Bewertungsinstrumentarium

Die Anwendung und Verfeinerung des Bewertungsinstrumentariums anhand von drei Beispielthemen aus verschiedenen Bereichen medizinischer Technologien spricht für die globale, d.h. themenübergreifende Anwendbarkeit des entwickelten Instrumentariums.

Das im Konsens erarbeitete Bewertungsinstrumentarium leistet einen unterstützenden Beitrag bei Public-Health-relevanten Entscheidungsprozessen zum Einsatz medizinischer Technologien und der damit verbundenen Ressourcenallokation. Insbesondere geschieht dies im Hinblick auf die kritische Beurteilung der vorhandenen Datenlage und die Identifikation kritischer Faktoren bei der Generalisierbarkeit und auf Deutschland zu übertragenden gesundheitsökonomischen Evidenz internationaler Studienergebnisse.

b) PSA-Screening beim Prostatakarzinom

Bei einem Massen-Screening wäre mit zusätzlichen Kosten für das deutsche Gesundheitssystem bei fraglichen Gewinnen an Lebensjahren und Lebensqualität zu rechnen. Aufgrund der großen Heterogenität der Kosten-Effekt-Relationen in den ausgewerteten Studien und der skizzierten erheblichen Probleme einer Übertragbarkeit auf die Bundesrepublik, lassen sich die zu erwartenden Kosten und Kosten-Effektivitäts-Relationen gegenwärtig für Deutschland nicht hinreichend präzise quantifizieren.

Vor einem Routineeinsatz des PSA-Screenings mit Aufnahme in den Leistungskatalog der gesetzlichen Krankenversicherung sollte eine stärkere Evidenz für die Abnahme der gesellschaftlichen Krankheitslast herrschen. Zur Beurteilung von Compliance, Effektivität und Kosten von einem Massen-Screening sind die Ergebnisse aus randomisierten Studien abzuwarten.

Der individuellen Patientenentscheidung zur Durchführung eines PSA-Tests sollte grundsätzlich ein ausführliches ärztliches Gespräch vorausgehen, welches u.a. über die Risiken von screening-bedingten Komplikationen bei zweifelhaftem medizinischen Nutzen informiert.

Die Empfehlung, bei der derzeitigen Evidenzlage von einem routinemäßigen Einsatz des PSA-Screenings abzusehen, deckt sich weitgehend mit der Schlussfolgerung der vorausgegangenen Arbeit von Pientka, in der die Ergebnisse von HTA-Berichten, Leitlinien/Konsensuspapieren und systematischen Reviews aufarbeitet und zusammengefasst wurden.¹¹⁶ Dort konnte gezeigt werden, dass in erster Linie methodisch schwächere Dokumente (vor allem Konsensuspapiere) zu einer positiven Beurteilung des Screenings gelangen, dass jedoch mehr wissenschaftlich- und evidenzorientierte Dokumente (HTA-Berichte, Leitlinienpapiere) den Nutzen des Screenings als zweifelhaft und unzureichend belegt erachten.

c) Weiterer Forschungsbedarf

Aufgrund der eingehenden Betrachtungen zur internen Validität und Übertragbarkeit der Studienergebnisse halten wir bei entscheidungsrelevanten Themenfeldern folgende methodischen Vorgehensweise für sachgerecht und notwendig:

- extern entwickelte entscheidungsanalytische Modelle (Entscheidungsbäume oder Markov-Modelle) hoher Qualität und Flexibilität werden innerhalb eines

Forschungsprojektes bezüglich Struktur und Daten an die bundesdeutschen Verhältnisse angepasst,

- Studienergebnisse zu Kosten und Kosten-Effektivitäts-Relationen werden mit Blick auf die bundesdeutschen Verhältnisse re-analysiert.

Hiermit zeichnet sich ein effizientes Forschungsprogramm ab, dass dem gesundheitspolitischen Entscheidungsträger anwendungsorientierte Ergebnisse zu Allokationsentscheidungen zur Verfügung stellen kann, die in eine bundesdeutsche Datenbasis "Evaluation medizinischer Verfahren und Technologien" aufgenommen werden können.

C Hauptdokument

PSA-Screening beim Prostatakarzinom.

Systematischer gesundheitsökonomischer Review

Uwe Siebert, Nikolai Mühlberger, Corinne Behrend, Jürgen Wasem

C.1 Policy Question

Das Prostatakarzinom ist in Deutschland mit etwa 11% (1996) die dritthäufigste tumorbedingte Mortalitätsursache bei Männern und gleichzeitig die häufigste Todesursache unter den urologischen Tumoren.¹³⁶

Das Verfahren zur Bestimmung des prostataspezifischen Antigens (PSA-Test) wurde 1994 als Gebührenziffer in den EBM aufgenommen. Abrechnungsfähigkeit besteht jedoch nur bei einer Anwendung zur Bestätigung und zum Monitoring von Malignität. Ein Test-Einsatz ohne vorherigen Verdacht auf ein Prostatakarzinom ist somit bisher nicht vorgesehen.

Allerdings ist in der Praxis eine Zunahme des Einsatzes des PSA-Tests als Verfahren zur Diagnose und Früherkennung von Prostatakarzinomen festzustellen. Es stellt sich die Frage, wie dieses Verfahren in Verbindung mit einem routinemäßigen Screening auf Prostatakarzinom bezüglich des kurz- und auch langfristigen medizinischen Nutzens für die Patienten zu bewerten ist. Dabei muss der medizinische Nutzen gegen die kurz- und langfristigen Risiken für alle Screening-Teilnehmer abgewogen werden.

Um den Leistungszuwachs der sich weiter entwickelnden medizinische Technologien und der alternden Bevölkerung auch in Zukunft finanzieren zu können, müssen angesichts der Begrenztheit der Ressourcen im deutschen Gesundheitswesen, die etwa im Grundsatz der Beitragssatzstabilität ihren Ausdruck gefunden hat, medizinische Technologien systematisch und vergleichbar nicht nur bezüglich medizinischer Effektivität, sondern auch bezüglich ihrer gesundheitsökonomischen Effizienz bewertet werden.

Nach den bestehenden gesetzlichen Grundlagen (§2 Abs. 1 SGB V) haben Qualität und Wirksamkeit der erbrachten Leistungen im Gesundheitswesen dem allgemein anerkannten Stand der medizinischen Erkenntnisse zu entsprechen und den medizinischen Fortschritt zu berücksichtigen. Das Wirtschaftlichkeitsgebot (§12 Abs. 1 SGB V) gebietet, dass die Leistungen das Maß des Notwendigen nicht überschreiten dürfen. Neue Untersuchungs- und Behandlungsmethoden in der ambulanten vertrags(zahn)ärztlichen Versorgung sind nur dann abrechnungsfähig, wenn der Bundesausschuss der Ärzte und Krankenkassen deren diagnostischen oder therapeutischen Nutzen sowie deren medizinische Notwendigkeit und Wirtschaftlichkeit - auch im Vergleich zu bereits zu Lasten der Krankenkassen

erbrachten Methoden - anerkannt hat (§135 Abs.1 SGB V). In einer Richtlinie hat der Bundesausschuss der Ärzte und Krankenkassen das Verfahren zur Überprüfung der Voraussetzungen für die Aufnahme neuer und die Überprüfung bestehender vertragsärztlicher Leistungen geregelt.¹⁹

Zum Thema "PSA-Bestimmung beim Screening für Prostatakarzinom" liegen inzwischen mehrere Beobachtungsstudien und zusammenfassende Reports aus dem Ausland vor, die den medizinischen Nutzen der Technologie kritisch hinterfragen. In Deutschland kam ein in der ersten Phase des BMG-Projekts "Aufbau einer Datenbasis zur Evaluation medizinischer Verfahren und Technologien in Deutschland" verfasster Bericht zur medizinischen Effektivität des PSA-Screenings¹¹⁶ zu der Schlussfolgerung, dass zum jetzigen Zeitpunkt keine hinreichende Evidenz für die medizinische Wirksamkeit einer routinemäßigen Anwendung des Tests zur Screening-Untersuchung vorliegt.

Die vorliegende Nachfolgearbeit (2. Phase des BMG-Projekts) befasst sich mit den gesundheitsökonomischen Konsequenzen des PSA-Screenings und ergänzt auf diese Weise den Prozess der Technologiefolgenabschätzung. Die zum heutigen Zeitpunkt verfügbare Evidenz wird anhand der verfügbaren Literatur dargestellt, zusammengefasst und bewertet, wobei neben der Effektivität der Technologie insbesondere die Kosten und die Relation zwischen Kosten und Effektivität Berücksichtigung finden. Um einen Anhaltspunkt für die Übertragbarkeit der Ergebnisse liefern zu können, werden die Daten um Hintergrundzahlen aus Deutschland ergänzt. Im Sinne des Gesamtprojekts verfolgt die vorliegende Arbeit zwei grundsätzliche Ziele:

- Die Entwicklung eines Instrumentariums zur systematischen Bewertung gesundheitsökonomischer Studien
- Anwendung dieses Instrumentariums zur Beurteilung der gesundheitsökonomischen Evidenz bezüglich der medizinischen Technologie PSA-Screening beim Prostatakarzinom

Dabei wurde besonderer Wert auf das Aufzeigen methodischer Probleme bei der Technologiefolgenabschätzung von Screening-Verfahren gelegt.

Insgesamt soll mit dem vorliegenden Bericht ein Beitrag im Rahmen des anstehenden Entscheidungsprozesses geleistet werden.

C.2 Einführung/Hintergrund

Die folgende Einführung gibt zunächst einen allgemeinen Überblick über die Methoden der gesundheitsökonomischen Evaluation und die Besonderheiten von Screening-Verfahren als medizinische Technologien. Es folgt die Beschreibung der Zielerkrankung (Zielkondition) "Prostatakarzinom", der Technologie "PSA-Bestimmung" und der evaluierten Intervention "PSA-Screening beim Prostatakarzinom".

C.2.1 Methoden gesundheitsökonomischer Evaluation

Die vorliegende Arbeit befasst sich in erster Linie mit den ökonomischen Konsequenzen von Technologieanwendungen. Gesundheitsökonomische Evaluationen können zur Unterstützung gesellschaftlicher und einzelwirtschaftlicher Entscheidungen über die Nutzung medizinischer Technologien herangezogen werden. Als alleinige Entscheidungsgrundlage dürfen und wollen sie jedoch nicht gelten.

Eine umfassende Darstellung und fundierte Analyse der ökonomischen Evaluationsmethodik und ihrer spezifischen Probleme bei Anwendungen auf dem Feld der Gesundheitsversorgung kann im Rahmen dieser Arbeit nicht geleistet werden. Als Standardwerk zu diesem Thema können die Monographien von Luce & Elixhauser,⁹⁴ Sloan¹³² und Gold, Siegel, Russel & Weinstein⁴⁸ angesehen werden. Als deutschsprachige Einführungen und Empfehlungen zur Durchführung gesundheitsökonomischer Studien seien Schöffski et al.,¹²⁶ Leidl,^{86,87} König et al.,⁷⁶ Konsensgruppe "Gesundheitsökonomie"⁷⁷ und Hessel et al.⁶⁵ genannt. Um aber auch dem in diesem Bereich nicht bewanderten Leser eine Orientierungshilfe zu geben und die Grundlagen dieser Arbeit zu skizzieren, soll hier ein grober Überblick über einige wichtige Elemente der methodischen Konzeptionen zur ökonomischen Evaluation gegeben werden. Dieser Überblick wird um Hinweise auf Schwachpunkte in den Verfahren der ökonomischen Evaluation ergänzt. Methodische Probleme primär epidemiologischer bzw. statistischer Art werden an dieser Stelle nicht erläutert.

Zunächst sei auf die Vielschichtigkeit von Technologiefolgen hingewiesen. Neben der Frage nach den bei der Implementierung und Durchführung von Maßnahmen der

Gesundheitsversorgung entstehenden Kosten stellt sich die Frage nach der medizinischen Wirksamkeit, die sich in einer positiven Veränderung des Gesundheitszustands der Zielpopulation widerspiegelt. Positive Veränderungen können grundsätzlich als verringerte Wahrscheinlichkeit unerwünschter Gesundheitszustände (z.B. Vermeidung oder Linderung von Morbidität) oder als erhöhte Wahrscheinlichkeit erwünschter Gesundheitszustände (z.B. Erhöhung der gesundheitsbezogenen Lebensqualität) erfasst werden. Bei der Bewertung dieser Veränderungen sind unterschiedliche Gesichtspunkte zu berücksichtigen. Zunächst besitzt Gesundheit einen von individuellen Präferenzen abhängigen subjektiven Wert, was dazu führt, dass gleiche Änderungen des Gesundheitszustandes in unterschiedlichen Populationen unterschiedlich zu werten sind. Weiterhin schlägt sich eine Veränderung des Gesundheitszustandes häufig in einer Änderung der Produktivität nieder, da gesündere Menschen in der Regel in der Lage sind, länger und produktiver zu arbeiten. Schließlich beeinflusst die Veränderung des Gesundheitszustandes die Inanspruchnahme von Versorgungsleistungen. So können gesündere Menschen weniger, infolge einer Verlängerung des Lebens jedoch möglicherweise auch mehr medizinische Leistungen in Anspruch nehmen. Aus der Komplexität der Zusammenhänge geht hervor, dass die Ergebnisse einer gesundheitsökonomischen Evaluation in hohem Maße vom Wissen um das Ausmaß der gesamten Folgen einer Intervention bestimmt werden. Ohne eine hinreichende klinische und epidemiologische Wissensbasis ist eine sinnvolle ökonomische Evaluation daher unmöglich.

C.2.1.1 Begriffserklärungen

Für das Verständnis gesundheitsökonomischer Evaluationsstudien sind eine Reihe verschiedener Begriffe und Konzepte von grundlegender Bedeutung. Sie werden im folgenden kurz erläutert.

Kosten

Unter Kosten wird der monetär bewertete Verbrauch von Ressourcen oder Inputs verstanden, welcher für das Erzielen eines bestimmten Outputs oder die Erreichung eines bestimmten Zustands erforderlich ist. Entscheidend ist dabei, dass die Bewertung des Ressourcenverbrauchs auf Basis von Opportunitätskosten erfolgt.^{43,48,87,94,132} Diesem Bewertungsprinzip liegt die Überlegung zugrunde, dass

mit der Verwendung einer Ressource für einen bestimmten Zweck diese Ressource gleichzeitig immer auch der Verwendung für einen anderen Zweck entzogen wird. Damit geht notwendigerweise der Verzicht auf einen alternativen Output einher, dessen Wert aus der Sicht der Ökonomie die wahren Kosten der entsprechenden Ressource darstellt. Theoretisch betrachtet ist die Bewertung mit Opportunitätskosten in einer Evaluationsstudie deshalb (zumindest bei Einnahme einer gesellschaftlichen Perspektive) zwingend, weil sich eine ökonomische Evaluation im Kern immer auf eine Wahlentscheidung zwischen alternativen Verwendungsmöglichkeiten von Ressourcen bezieht.^{43,48,76,87,94,132} Kosten sind aus ökonomischer Sicht notwendigerweise immer mit Ressourcenverbrauch verknüpft. Im Kontext gesundheitsökonomischer Evaluation bedeutet dies, dass zu den tatsächlichen ökonomischen Kosten nur der Wert jener Ressourcen zu rechnen ist, die als Resultat von Krankheit und Krankheitsbehandlung verbraucht werden oder verlorengehen. So sind an Krankheit, Behinderung und Tod anknüpfende Transferzahlungen (z.B. Krankengeld, Invalidenrenten oder Sterbegeld) aus gesamtgesellschaftlicher Sicht nicht als Kosten zu betrachten, weil sie lediglich eine Umverteilung von Kaufkraft, nicht aber den Verbrauch von Ressourcen darstellen. Aus einzelwirtschaftlicher Sicht dagegen können Transferkosten sowohl einen Verlust (z.B. für den Kostenträger) als auch einen Gewinn (z.B. für den Zahlungsempfänger) an Ressourcen bedeuten.⁴⁸

Eingebürgert hat sich in der Literatur die Unterscheidung zwischen den direkten, den indirekten und den sog. intangiblen Kosten. Unter den direkten Kosten ist der monetär bewertete Ressourcenverbrauch zu verstehen, welcher zum einen unmittelbar an den Einsatz der betrachteten Technologie geknüpft ist und zum anderen durch Veränderungen der Inanspruchnahme von Leistungen des Gesundheitswesens als mittelbare Auswirkung des Technologieeinsatzes entsteht.^{38,48,76,87,94,126,132,137} Unter den indirekten Kosten ist der Geldwert der Veränderungen der Produktivität (z.B. bei verminderter Arbeitsfähigkeit oder bei Erwerbsunfähigkeit) eines Individuums infolge Morbidität oder vorzeitiger Mortalität zu verstehen.^{43,48,76,87,94,126,132} Intangible Kosten, auch psychosoziale Kosten genannt, schließen u.a. Veränderungen der psychischen Belastung, des Schmerzes, der Angst, der Selbsteinschätzung und der Fähigkeit zur Wahrnehmung sozialer Rollen ein.^{43,48,87,126,132} Sie sind monetär schwer zu bewerten, weshalb manche Autoren empfehlen, die Arten der intangiblen Kosten lediglich aufzuzählen. Sie können jedoch auf der Effektseite, bei der Lebensqualitätsmessung, erfasst werden.^{43,87}

Konsequenzen, Effekte

Unter den Konsequenzen bzw. Effekten (outcomes) einer Maßnahme der Gesundheitsversorgung werden alle Ergebnisse und Ereignisse verstanden, welche auf den Einsatz dieser Maßnahme zurückzuführen sind. Dabei stehen üblicherweise (1) die klinischen bzw. medizinischen Konsequenzen und (2) deren Auswirkung auf die gesundheitsbezogene Lebensqualität (health-related quality of life, HRQL) im Mittelpunkt der evaluatorischen Betrachtung:

(1) Klinische bzw. medizinische Konsequenzen schließen Maße der Morbidität (z.B. Heilungsraten, Nebenwirkungshäufigkeiten) und Mortalität (z.B. Restlebenserwartung, Überlebensraten), aber auch intermediäre Ergebnisse (z.B. Häufigkeit korrekter Diagnosen, gesundheitsrelevante Verhaltensänderungen) ein.^{76,87,126} Eine wesentliche Beschränkung der Verwendung ausschließlich medizinischer (und hier insbesondere intermediärer) Ergebnismaße besteht in der mangelnden Vergleichbarkeit von Studien mit unterschiedlichen Effektparametern.^{76,87} Diese Beschränkung kann durch Umrechnung der klinischen Konsequenzen in standardisierten Einheiten (z.B. DM oder QALYs) abgemildert werden.

(2) Gesundheit ist eine mehrdimensionale Größe, für die es keine ausdifferenzierte, allgemein akzeptierte Definition gibt.^{87,126} Es liegt jedoch eine Reihe von Konzepten vor, Gesundheit und ihre Dimensionen zu beschreiben und zu messen. In der Diskussion haben sich einige zentrale Komponenten der Gesundheit - und damit der gesundheitsbezogenen Lebensqualität, welche die für ein Individuum neben der Lebensdauer wesentlichen Aspekte der Gesundheit erfasst - herauskristallisiert. So wird Gesundheit in Kategorien des physischen, psychischen und sozialen Befindens und Funktionierens beschrieben. Krankheit wird in Folge über die Einschränkungen dieser Befindlichkeiten und Funktionen definiert. Im Rahmen einer Evaluation können die Kategorien der Gesundheit als Einzelitems, gruppiert oder in einer Gesamtschätzung bewertet werden.^{87,126}

Wichtige Ansätze zur Messung und Beurteilung von Gesundheitszuständen sind zum einen Methoden zur quantitativen Messung gesundheitsbezogener Lebensqualität und zum anderen Verfahren der Nutzen- bzw. Nutzwertmessung.

Krankheitsspezifische Maße richten sich auf diejenigen Gesundheitsaspekte, welche mit einem bestimmten Krankheitsbild verbunden sind. Maße der Lebensqualität, die Gesundheitsaspekte übergreifend über den jeweiligen Untersuchungsbezug abbilden, heißen generische Gesundheitsmaße. Ihre weitgehende Vergleichbarkeit kann auf Kosten der Feststellung krankheitsspezifischer Veränderungen gehen. Gesundheitsprofile beschreiben einzelne Aspekte und fassen diese gegebenenfalls zu Dimensionen zusammen. Beispiele für generische Profilinstrumente sind das Sickness Impact Profile (SIP), der Medical Outcome Study 36-Item Short Form Health Survey (SF-36) und das Nottingham Health Profile (NHP). Gesundheitsindizes dagegen integrieren alle Indikatoren nach einem vorgegebenen Verrechnungsschema in einer einzigen Maßzahl. Zu den generischen Indexinstrumenten zählen beispielsweise der Karnofsky-Index, der EuroQol und die Quality of Well-Being Scale (QWB). Weiterhin wird bei der Messung der Lebensqualität nach Ordinal- oder Kardinalskalierung unterschieden. Anhand einer Ordinalskala kann die Rangfolge von Gesundheitszuständen definiert werden. Ungewiss bleibt jedoch, um wieviel besser oder schlechter ein bestimmter Gesundheitszustand im Vergleich zu einem anderen eingeschätzt wird. Bei Kardinalskalen dagegen gewinnt auch der Abstand zwischen zwei Lebensqualitätsstufen an Bedeutung. Um die Ergebnisse von Lebensqualitätsmessungen gut einschätzen zu können, muss die methodische Brauchbarkeit der verschiedenen Instrumente bei jedem Untersuchungskontext aufgezeigt werden. Zu den wichtigsten methodischen Prüfkriterien gehören dabei die Objektivität, die Reliabilität, die Validität und die Reagibilität des eingesetzten Messinstrumentes wie auch die Praktikabilität der Messung.^{57,87,114,126}

Wird die Gesamtheit der Effekte einer Maßnahme zusammengefasst und nach individuellen oder kollektiven Präferenzen gewichtet, spricht man von Nutzwerten (utilities). Nutzenbewertungen von Gesundheitszuständen beruhen demnach auf individuellen bzw. kollektiven Gesamtwertungen der Zustände in Einheiten einer kardinalen Nutzenfunktion, mit der das mehrdimensionale Konzept der Gesundheit in einem skalaren Index (relativer bzw. inkrementeller Nutzwert, typischerweise auf einer Skala mit den Endpunkten 0 für Tod und 1 für vollständige Gesundheit) abgebildet wird.^{87,126} Ein Ansatz für eine solche Nutzenmessung besteht darin, zunächst einen Gesundheitsindex als eindimensionale Maßzahl zu entwickeln und in einem zweiten Schritt für diesen Index eine Bewertungsfunktion zu ermitteln, welche jedem Indexwert eine Nutzengröße zuordnet.^{18,126} Alternativ können Gesundheitszustände auch, ohne Umwege über einen Gesundheitsindex, direkt bewertet werden.

Zur Ermittlung individueller Präferenzen haben sich drei Verfahren etabliert: das Einstufungsverfahren (rating scale method), die Methode der Standardlotterie (standard gamble method) und das Verfahren der zeitlichen Abwägung (time trade-off method):

(1) Bei Anwendung des Einstufungsverfahrens wird das Individuum typischerweise mit einer Linie mit klar definierten Endpunkten konfrontiert. Die beiden Endpunkte dieser Rating-Skala markieren den besten bzw. schlechtesten Gesundheitszustand einer Reihe relevanter Gesundheitszustände (z.B. völlige Gesundheit und Tod). Ziel ist es, die zu bewertenden Gesundheitszustände gemäß der individuellen Präferenz zwischen den Extremwerten zu positionieren.^{48,87,126,132}

(2) Beim Verfahren der Standardlotterie werden die Befragten für jeden zu bewertenden Gesundheitszustand vor die Wahl zwischen dem sicheren Verbleib im betreffenden Gesundheitszustand und einem Wettspiel gestellt. Das Wettspiel beinhaltet mit der Wahrscheinlichkeit p vollkommene Gesundheit und mit der Wahrscheinlichkeit $1-p$ den sofortigen Tod. Der Wert für p wird solange variiert, bis ein Wert gefunden ist, bei dem das Individuum zwischen beiden Optionen indifferent ist. Die Wahrscheinlichkeit p am Indifferenzpunkt entspricht dem Nutzwert des zu bewertenden Gesundheitszustands.^{18,45,48,87,126,132,142-144}

(3) Bei der Methode der zeitlichen Abwägung wird das Individuum für jeden spezifizierten Gesundheitszustand vor die Wahl zwischen dem Verbleib im betreffenden Gesundheitszustand für eine Restlebensdauer T und vollständiger Gesundheit für eine kürzere Restlebensdauer t gestellt. Anschließend wird t solange variiert, bis ein Wert gefunden ist, bei dem das Individuum zwischen beiden Optionen indifferent ist. Der Nutzwert des Zustands ergibt sich aus t/T .^{18,45,48,87,126,132,142-144}

Mit Hilfe der ermittelten Nutzwerte kann für jeden gegebenen Gesundheitszustand die Zeitspanne in idealer Gesundheit wiedergegeben werden, die diesem Zustand nutzenmäßig äquivalent ist. Damit werden die infolge einer Maßnahme der Gesundheitsversorgung auftretenden Veränderungen in der (Rest-)Lebensdauer und in der gesundheitsbezogenen Lebensqualität in einem aggregierten Effektmaß

zusammengefasst. So werden beispielsweise die sog. qualitätsbereinigten Lebensjahre (quality-adjusted life years, QALYs) als der Summe der Produkte aus ermittelten Präferenzwerten für die Gesundheitszustände und der jeweils darin verbrachten Zeit ermittelt.^{48,87,126,132,142} In der tatsächlichen Anwendung der Verfahren haben sich jedoch unterschiedliche Ergebnisse gezeigt. Dies kann u.a. an den technischen Schwierigkeiten liegen, für ein so komplexes Thema ein reliables und valides Befragungsinstrument zu entwickeln. Zusätzlich weisen experimentelle Forschungsergebnisse darauf hin, dass die Gültigkeit der den Modellen zugrunde liegenden Verhaltensannahme der Maximierung des Erwartungsnutzens angezweifelt werden kann. Keines der beschriebenen Bewertungsverfahren kann den Anspruch auf Allgemeingültigkeit erheben; einen Goldstandard für die Messung der Lebensqualität gibt es nicht. Vielmehr wird diskutiert, welches Verfahren unter welchen Umständen geeignet ist.^{87,126} Das Problem einer vergleichenden Nutzenbewertung von Gesundheitszuständen kann demnach gegenwärtig weder in seinen praktisch-technischen Aspekten noch in seiner theoretischen Grundlegung als gelöst gelten.

Inkrementelle Analyse

Das Prinzip der inkrementellen Analyse - teils auch uneinheitlich Grenzwertanalyse oder Marginalanalyse genannt - ist in der Ökonomie von zentraler Bedeutung. Nach diesem Prinzip können einer bestimmten Aktivität solange zusätzliche Ressourcen zugeführt werden können, wie der Nutzen einer zusätzlichen Input-Einheit deren Kosten übersteigt. In diesem Fall nimmt auch der Gesamtnutzen dieser Aktivität zu. Übersteigt demnach der Grenznutzen einer Maßnahme - d.h. der Zusatznutzen, der aus einer zusätzlichen Input-Einheit resultiert - die Grenzkosten, dann kann der Gesamtnutzen durch eine Ausdehnung der Aktivität erhöht werden. Ist der Grenznutzen dagegen geringer als die Grenzkosten, führt die Fortsetzung der Aktivität zu einem Nettoverlust.^{18,43,87,94,109,126,132,137} Eines der grundlegenden Probleme in der Medizin besteht nicht in der Frage, ob bestimmte Maßnahmen völlig überflüssig sind, sondern darin, bis zu welchem Punkt Diagnose oder Therapie vorangetrieben werden sollen. Bei Zugrundelegen von Durchschnittskosten- und Durchschnittsnutzenbetrachtungen können, wenn nicht falsche, so doch zumindest äusserst irreführende Ergebnisse erzielt werden, da Durchschnitts- und Marginalwerte unter Umständen extrem voneinander abweichen. Die Bedeutung des inkrementellen Ansatzes belegt beispielsweise die oft angeführte Studie von Neuhauser & Lewicki (1975)¹⁰⁶, in welcher die Kosten-Effektivität der Wiederholung

eines bestimmten Tests beim Kolonkarzinom-Screening im Hinblick auf eine Minimierung falsch-negativer Testresultate untersucht wurde. In dieser Studie wurde gezeigt, dass bei der sechsten Durchführung des Tests die Durchschnittskosten pro entdecktem Krebsfall scheinbar vertretbar um ca. US-\$ 200 auf US-\$ 2,451 anstiegen; die für den Übergang von fünf auf sechs Tests zusätzlich anfallenden Kosten pro zusätzlich diagnostiziertem Karzinom lagen allerdings bei über US-\$ 47 Millionen, da nur wenige weitere Fälle identifiziert werden konnten.

Diskontierung

Fundamental für jede ökonomische Bewertung ist die genaue Analyse des Zeitprofils der Kosten und der Konsequenzen der zu evaluierenden Gesundheitstechnologien. Zwischen den zu vergleichenden Alternativen gibt es in der Regel Unterschiede hinsichtlich des Zeitprofils. Aus diesen Unterschieden können unterschiedliche Bewertungen alternativer Interventionen resultieren, obwohl sich diese Alternativen bei zeitunabhängiger Betrachtung in ihren Kosten und Konsequenzen nicht unterscheiden. Dies steht im Zusammenhang damit, dass bezüglich der Zeitpunkte des Kosten- und Nutzenanfalls unterschiedliche Präferenzen bestehen: Positive Effekte werden im allgemeinen zu einem frühen Zeitpunkt, negative Effekte zu einem späten Zeitpunkt bevorzugt. Diese Zeitpräferenz lässt sich u.a. aus der Unsicherheit über die Dauer des eigenen Lebens erklären.

Der Zeitpräferenz muss in einer ökonomischen Bewertung Rechnung getragen werden, indem zukünftig anfallende Kosten und Nutzen in ihre Gegenwartswerte umgerechnet werden. Erst durch diese, als Diskontierung bezeichnete, Umrechnung werden die im Rahmen einer Evaluationsstudie betrachteten Alternativen vergleichbar.^{38,43,48,76,87,94,126,132} Der erforderliche Umfang der Diskontierung (jährliche Diskontierungsrate) ist allerdings noch umstritten.^{48,132} Auch wurde die Frage aufgeworfen, ob für die Kosten des mit der Gesundheitstechnologie verbundenen Ressourcenverbrauchs nicht eine andere Diskontierungsrate verwendet werden sollte als für die gesundheitlichen Konsequenzen.¹¹³ Ferner wurde eine Veränderung des Diskontsatzes in Abhängigkeit von der Zeit diskutiert.⁶⁰

C.2.1.2 Grundtypen der ökonomischen Evaluation

Tabelle 1 stellt verschiedene Formen oder Typen der Evaluation von Leistungen der Gesundheitsversorgung gegenüber.

Tabelle 1: Methoden der Evaluation von Leistungen der Gesundheitsversorgung

Untersuchte Faktoren	Untersuchungen ohne Vergleich	Untersuchungen mit Vergleich
Kosten	Kostenbeschreibung (cost description, economic dossier) Krankheitskosten-Studie (cost-of-illness analysis)	Kosten-Analyse (cost analysis)
Konsequenzen	Effektbeschreibung (outcome description)	Wirksamkeits-Analyse (efficacy evaluation) Effektivitäts-Analyse (effectiveness evaluation)
Kosten und Konsequenzen	Kosten-Konsequenzen-Beschreibung (cost outcome description)	Kosten-Konsequenzen-Analyse (cost-consequence analysis) Kosten-Minimierungs-Analyse, Kosten-Kosten-Analyse (cost-minimisation analysis) Kosten-Effektivitäts-Analyse (cost-effectiveness analysis) Kosten-Nutzwert-Analyse (cost-utility analysis) Kosten-Nutzen-Analyse (cost-benefit analysis)

Tabelle nach Stoddart & Drummond (1984)¹³⁷

Die Bewertung von Wirtschaftlichkeit bzw. Effizienz im Gesundheitswesen erfordert Verfahren, welche die Kosten und Nutzen von Leistungen der Gesundheitsversorgung messen und miteinander vergleichen. Vollständige ökonomische Evaluationen zeichnen sich durch zwei Bedingungen aus: (1) Es müssen mindestens zwei alternative Technologien verglichen werden, wovon eine auch der bisherige Versorgungsstandard ("Status quo" bzw. "Keine Intervention") oder eine Placebobehandlung sein kann. (2) Es müssen sowohl Kosten als auch Effekte der alternativen Technologien bestimmt und in Relation gesetzt werden.^{43,48,76,87,94,132,137} Je nachdem, ob die Konsequenzen einer Leistung der Gesundheitsversorgung in natürlichen, morbiditäts- oder mortalitätsbezogenen Einheiten (z.B. klinische Parameter wie der Blutdruck, Restlebensdauer, Lebensqualität) betrachtet werden, ob den Konsequenzen ein Nutzwert zugeschrieben wird oder ob die Konsequenzen monetär bewertet werden, lassen sich verschiedene Evaluationstypen beschreiben.^{43,48,76,87,137}

Kosten-Konsequenzen-Analysen

In Kosten-Konsequenzen-Analysen werden die Kosten und Konsequenzen der untersuchten Alternativen der Gesundheitsversorgung lediglich in disaggregierter Form aufgeführt (z.B. als Arzneimittelkosten, Krankenhauskosten, andere Kosten, verhinderte Erkrankungsfälle, kurzfristige Nebenwirkungen, langfristige Nebenwirkungen) und einander gegenübergestellt. Jegliche Gewichtung der einzelnen Komponenten sowie deren Aggregation wird dem Leser bzw. Nutzer der Studie überlassen.^{20,48}

Kosten-Minimierungs-Analysen; Kosten-Kosten-Analysen

Kosten-Minimierungs-Analysen, auch Kosten-Kosten-Analysen genannt, sind dadurch gekennzeichnet, dass die Kosten von zwei oder mehr Interventionen miteinander verglichen werden, um die kostengünstigste Alternative zu ermitteln. In der Praxis besteht zwischen der Kosten-Minimierungs-Analyse und der Kosten-Analyse kein Unterschied, wenn die Effekte der Alternativen identisch sind.^{48,87,94,126,137} Die implizite Unterstellung gleicher Effektivität sollte jedoch sorgfältig geprüft werden. Im Prinzip bedarf eine Kosten-Minimierungs-Analyse der Evidenz, dass Unterschiede in den Effekten der verschiedenen Alternativen nicht existierend bzw. unbedeutend sind.¹³⁷

Kosten-Effektivitäts-Analysen

Bei Kosten-Effektivitäts-Analysen werden sowohl Kosten als auch Konsequenzen erhoben. Dabei werden die Effekte der zu bewertenden Interventionen mengenmäßig in ihren natürlichen Einheiten (klinische Parameter, Lebensdauer, Lebensqualität) ausgedrückt.^{43,76,87,94,126} Das Verhältnis der im Vergleich zu einer Basisalternative zusätzlich anfallenden Kosten und Effekten wird als (inkrementeller) Kosten-Effektivitäts-Quotient (cost-effectiveness ratio) bezeichnet. Aus medizinischer Perspektive hat die Kosten-Effektivitäts-Analyse den Vorteil, die für die klinische Bewertung unmittelbar relevanten Wirkungsdimensionen zu erfassen, soweit diese einer quantitativen Messung zugänglich sind. Die Beschränkungen dieses Verfahrens sind vor allem darin zu sehen, dass nur Interventionen miteinander verglichen werden können, deren Konsequenzen sich in gleichen natürlichen, eindimensionalen Einheiten ausdrücken lassen.^{87,126}

Kosten-Nutzwert-Analysen

Kosten-Nutzwert-Analysen folgen den Prinzipien der Kosten-Effektivitäts-Analyse, drücken jedoch verschiedene medizinische Effektdimensionen auf einer gemeinsamen Skala in aggregierten Nutzeneinheiten aus.^{43,48,87,94,126,132} Unter den Ansätzen zur aggregierten Beurteilung von Lebensdauer und Lebensqualität dominiert das QALY-Konzept. Kosten-Nutzwert-Analysen werden von manchen Autoren ebenfalls als Kosten-Effektivitäts-Analysen bezeichnet.⁴⁸ Kosten-Nutzwert-Analysen zeichnen sich dadurch aus, dass sie prinzipiell alle Allokationsalternativen innerhalb des Gesundheitssektors einer durchgängig vergleichenden Bewertung zugänglich machen. Die Ergebnisse verschiedener QALY-Analysen werden vielfach zu QALY league tables zusammengefasst.^{42,48,142} Hierunter wird die Zusammenstellung der Kosten von gesundheitlichen Versorgungsmaßnahmen je gewonnenem qualitätsbereinigtem Lebensjahr in einer Rangfolge verstanden. Eine unreflektierte Interpretation dieser Ranglisten ist allerdings problematisch. So werden Kosten und Nutzwerte bei der Berechnung der QALYs vielfach unterschiedlich abgegrenzt; auch beruhen die Berechnungen teilweise auf Durchschnitts-, teilweise auf Grenzwertbetrachtungen. Zudem müssen die Studien, auf denen die QALY-Ranglistenwerte beruhen, im Hinblick auf ihre Generalisierbarkeit bzw. Übertragbarkeit geprüft werden.⁴²

Kosten-Nutzen-Analysen

Kosten-Nutzen-Analysen unterziehen alle Effekte einer Maßnahme der Gesundheitsversorgung einer monetären Bewertung. Das Resultat einer solchen Analyse ist der Nettonutzen (bzw. bei negativem Nettonutzen die Nettokosten) der evaluierten Intervention(en).^{18,43,48,87,126,132} Ein Vorteil dieser Analyseform liegt in der prinzipiell uneingeschränkten Vergleichbarkeit von Interventionen bzw. Aufwendungen innerhalb wie auch außerhalb des Gesundheitssektors.⁸⁷ Die Kosten-Nutzen-Analyse ist demzufolge eine Form der ökonomischen Evaluation, mit deren Hilfe auch über die Frage entschieden werden könnte, ob Ressourcen für die Gesundheitsversorgung eingesetzt oder anderen Verwendungszwecken zugeführt werden sollten. Darüber hinaus ist sie die einzige Form der ökonomischen Bewertung, welche die Frage nach Nettokosten bzw. Nettoeinsparungen einer Maßnahme beantworten kann (Kosten-Nutzen-Differenz).^{41,43,138}

Der Anwendung der Kosten-Nutzen-Analyse zur Evaluierung der Gesundheitsversorgung stehen jedoch erhebliche Probleme entgegen, die vor allem mit der monetären Bewertung des menschlichen Lebens - oder, genauer, der Verlängerung statistischer Menschenleben - zusammenhängen.^{18,87} Für die Monetarisierung medizinischer Konsequenzen existieren zwei sehr unterschiedliche Konzepte: (1) der Humankapitalansatz (human capital approach) und (2) der Ansatz der Zahlungsbereitschaft (willingness-to-pay approach).

(1) Nach dem Humankapitalansatz ist der Wert des Lebens durch den Beitrag gegeben, den ein Mensch noch zum Sozialprodukt leisten könnte. Dieser Ansatz hat den Vorzug, relativ leicht operationalisierbar zu sein. Er zielt darauf ab, den vollständigen Wegfall an Produktionspotential als Folge von Krankheit und frühzeitigem Tod zu ermitteln, wobei das aufgrund von Morbidität und Mortalität entgangene Einkommen, berechnet aus der Zeit des Produktionsausfalls und dem Lohnsatz, herangezogen wird.^{18,87,132} Eine Weiterentwicklung des Humankapitalansatzes, der Friktionskostenansatz, sucht nicht den potentiellen, sondern den tatsächlichen Produktionsausfall zu erfassen.⁷⁸ Bei einem längeren Arbeitsausfall durch Krankheit oder Tod fällt in Ländern mit dauerhaft hoher Arbeitslosigkeit die Produktion nicht um die gesamte entgangene Arbeitszeit aus. Beim Friktionskostenansatz werden langfristige Produktionsausfälle durch Erwerbsunfähigkeit und Tod, die beim Humankapitalansatz zu hohen Kostenschätzungen führen, auf die Zeit, die bis zur Einstellung eines neuen Arbeitnehmers vergeht (Friktionsperiode), beschränkt. Im Ergebnis bietet dieser Ansatz besonders in Ländern mit dauerhaft hoher Arbeitslosigkeit eine realistischere Bewertung von Produktionsverlusten. Allerdings entzündeten sich ethische Bedenken an der Implikation beider Ansätze, dem Leben nicht erwerbsfähiger Menschen jeglichen Wert abzusprechen.^{18,87,132} Aus ökonomischer Sicht ist außerdem zu kritisieren, dass bei beiden Ansätzen das Sozialprodukt zu einem validen Indikator für das Wohlergehen der Gesellschaft erklärt wird. Eine Begründung hierfür ist aus der ökonomischen Theorie, welche sonst durchweg auf individuelle Wertungen abstellt, nicht ableitbar.¹⁸

(2) Dem Ansatz der Zahlungsbereitschaft zufolge wäre der Wert des Lebens grundsätzlich aus der Höhe des Geldbetrags abzuleiten, den ein Individuum für eine in ihrem Ausmaß definierte Verringerung der Sterbewahrscheinlichkeit innerhalb eines gegebenen Zeitraums bzw. für eine in ihrem Ausmaß definierte Verlängerung der erwarteten Restlebensdauer zu zahlen bereit ist.¹⁸ Indem individuelle oder kollektive Präferenzen herangezogen werden, ist der Ansatz der

Zahlungsbereitschaft zwar fest in der subjektiven Wertlehre der Ökonomie verankert, stößt jedoch in der praktischen Anwendung auf Schwierigkeiten. Generell gibt es zwei unterschiedliche Ansätze zur Messung der Zahlungsbereitschaft. Zum einen können die Individuen direkt befragt werden (Befragungen, Fragebogenstudien; contingent valuation approach), zum anderen kann ihre Zahlungsbereitschaft indirekt aus der Beobachtung ihres Verhaltens geschlossen werden (Auswertung von Marktdaten bzw. Messung offenbarer Präferenzen; revealed preferences approach).^{18,132} Beide Methoden haben ihre spezifischen Vorzüge und Probleme. So weisen die Ergebnisse empirischer Studien zur Zahlungsbereitschaft aus Umfragen und Verhaltensbeobachtungen teilweise Inkonsistenzen auf und zeichnen sich vor allem durch eine große Streubreite der ermittelten Bewertungen statistischer Leben aus.¹⁸

Ein theoretisch und empirisch unproblematischer methodischer Zugang zur monetären Bewertung von Unterschieden in der Lebenserwartung steht derzeit nicht zur Verfügung. Dies gilt dementsprechend auch für die monetäre Bewertung von Änderungen des Gesundheitszustands, selbst wenn es gelänge, die oben angedeuteten Probleme bei der Nutzensgewichtung von Gesundheitszuständen zu lösen.

Krankheitskosten-Studien

In Krankheitskosten-Studien wird lediglich die Kostenseite untersucht, wobei der Ressourcenverbrauch im Zusammenhang mit einer oder mehreren Krankheiten ermittelt wird.^{43,87,137} Im Unterschied zu den bisher genannten Formen ökonomischer Bewertung beinhalten Krankheitskosten-Studien keinen Vergleich alternativer Interventionen sondern untersuchen die ökonomischen Konsequenzen der Krankheit selbst. Sie identifizieren kostenintensive Bereiche der Gesundheitsversorgung und quantifizieren das potentielle Einsparpotential therapeutischer und präventiver Maßnahmen.⁷³ In der Praxis unterscheiden sich Krankheitskosten-Studien häufig hinsichtlich des Spektrums der einbezogenen Kostenarten und -elemente wie auch in ihrer entweder inzidenz- oder prävalenzorientierten Betrachtungsweise. Valide Vergleiche der Krankheitslast unterschiedlicher Krankheiten sind jedoch nur dann möglich, wenn die Studien diesbezüglich identisch sind.⁷³

C.2.1.3 Schlüsselemente ökonomischer Evaluation

Zu den für die Aussagefähigkeit der Ergebnisse maßgeblichen Elementen ökonomischer Evaluationsstudien zählen die Wahl der Studienperspektive und die Auswahl der in die Bewertung einzubeziehenden Technologiealternativen.

Die Perspektive

Unter der Perspektive einer Studie wird der Standpunkt bzw. Blickwinkel verstanden, aus dem heraus die Studie durchgeführt wird. Dabei kann eine gesellschaftliche Betrachtungsweise, die Sichtweise eines Kostenträgers oder Leistungserbringers, der Standpunkt eines Arbeitgebers oder die Perspektive der Patienten bzw. ihrer Angehörigen in Frage kommen. Mit der Wahl der Perspektive wird maßgeblich über die in die Bewertung einzubeziehenden Kosten und Konsequenzen einer Gesundheitstechnologie sowie über die Schlussfolgerungen entschieden, die aus einer Studie gezogen werden können.^{20,38,43,48,76,87,94,126,132,138,141}

Die Technologiealternativen

Vollständige ökonomische Evaluationen sind ihrer Natur nach vergleichende Studien. "Beste" oder "günstigste" Interventionsalternativen sind Technologien immer nur bezüglich der Art und Menge jener Technologien, die in die vergleichende Bewertung mit eingegangen sind. Deshalb kommt der Auswahl der Alternativen überragende Bedeutung zu. Grundsätzlich gilt, dass alle in der Realität sinnvollen und praktikablen Technologien in den Vergleich einbezogen werden sollten.^{20,38,43,48,87,94,132,138,141} Eine Forderung, der vor allem bei der Durchführung randomisierter Studien, bei denen eine zusätzliche ökonomische Evaluation durchgeführt wird, oftmals nicht nachgekommen werden kann.

C.2.1.4 Formale Entscheidungsanalyse

Außerhalb des Gesundheitswesens hat die Verwendung von Modellen zur vereinfachten Darstellung komplexer Zusammenhänge, zur Simulation von Experimenten und zur Extrapolation alternativer Szenarien eine lange Tradition. Auch im Gesundheitswesen bieten sich Modellierungen an, wenn Entscheidungen unter Unsicherheit- d.h. unter Unkenntnis des zu erwartenden (Netto-)Nutzens bzw.

(Netto-)Schadens der zur Auswahl stehenden Alternativen - getroffen werden sollen.^{35,115,126} Aufgrund der Nutzung von Sekundärdaten ist die Erstellung von Modellen im Vergleich zur Datensammlung in klinischen Prüfungen relativ schnell und kostengünstig.^{115,126} Modellierungen ermöglichen es, komplexe Entscheidungsprozesse rational begründet und nachvollziehbar zu strukturieren und die Auswirkungen verschiedenster Parameteränderungen zu untersuchen.^{97,115} Die am häufigsten verwendeten Methoden zur Evaluation von Technologien im Zuge der Modellbildung sind klassische Entscheidungsbaumanalysen und Szenariomodelle wie Markov-Verfahren.

Entscheidungsbaumanalysen

Entscheidungsbäume bilden Krankheitsverläufe in Abhängigkeit von verschiedenen Interventionsalternativen graphisch ab. Dabei werden mögliche Abläufe klinischer Ereignisse und Entscheidungen getrennt für jede Alternative (Strategie) durch Knoten (Zufalls- bzw. Entscheidungsknoten) und Linien bis zu ihren Endpunkten (Endknoten) dargestellt. Ferner werden die Eintrittswahrscheinlichkeiten der verschiedenen Ereignisse eingetragen. Die Endpunkte repräsentieren die Kosten und medizinischen Konsequenzen (Effekte) der jeweiligen Strategien. Als Effekteinheiten können Geldeinheiten, Effektparameter oder Nutzwerte (utilities) angegeben werden. Der Baum dient der Veranschaulichung aller vom Entscheidungsträger in Betracht zu ziehenden Strategien sowie der Berechnung der strategiespezifischen Erwartungswerte. Letztere sind die mit den Pfadwahrscheinlichkeiten gewichteten Durchschnittswerte aller strategiespezifischen Endpunkte (folding back and averaging out).^{35,36,82,97,100}

Markov-Verfahren

Markov-Modelle werden ebenfalls dazu verwendet, Ereignis- und Entscheidungsabläufe zu strukturieren, empirische Daten zu synthetisieren und klinische sowie ökonomische Ergebnisse zu evaluieren. Markov-Modelle sind dann anstelle von einfachen Entscheidungsbäumen anzuwenden, wenn (i) der Zeitpunkt des Ereigniseintritts eine große Rolle spielt, (ii) zeitabhängige Risiken vorliegen oder (iii) Ereignisse wiederholt auftreten können. Dabei wird eine bestimmte Anzahl diskreter Gesundheitszustände definiert, Übergangswahrscheinlichkeiten zwischen den einzelnen Gesundheitszuständen in Abhängigkeit von Zeitintervallen gegebener

Länge festgelegt, die Kosten und Nutzwerte für jeden Gesundheitszustand ermittelt und schließlich die Berechnung durchgeführt. Wesentliche Eigenschaft der Markov-Verfahren ist, dass die Übergangswahrscheinlichkeit von einem zum anderen Zeitabschnitt ausschließlich vom aktuell eingenommenen Gesundheitszustand abhängig ist; alle zuvor durchlaufenen Zustände sind dafür irrelevant (Markov-Eigenschaft).^{9,101,133}

Die schwierigste Aufgabe liegt in der Festlegung der Übergangswahrscheinlichkeiten. Als Näherungswerte werden die in klinischen und epidemiologischen Studien ermittelten Häufigkeiten verwendet. Die Übergangswahrscheinlichkeiten sind im Fall von Markov-Ketten konstant; sie können aber auch zeitabhängig definiert werden und werden dann in Markov-Prozessen verarbeitet. Im komplexeren stochastischen Markov-Modell hängen die Übergangswahrscheinlichkeiten nicht nur von unterschiedlichen Zeitabständen, sondern auch vom Patientenalter sowie anderen Parametern wie beispielsweise dem Krankheitsrisiko ab. Ausgewertet werden kann ein Markov-Modell mittels Matrixalgebra (bei Markov-Ketten), als Kohorten-Simulation oder als Monte-Carlo-Simulation.^{9,98,101,133}

Aufgrund der Komplexität von Struktur und Annahmen sind Modelle häufig unübersichtlich und intransparent, was die Nachvollziehbarkeit ihrer Validität oft erschwert. Deshalb werden als Voraussetzung für die Akzeptanz von Modellen Einfachheit vor Komplexität, absolute Transparenz, eine Qualitätsüberprüfung der verwendeten Daten, Sensitivitätsanalysen sowie die Validierung des Modells gefordert.¹²⁶

C.2.2 Screening als medizinische Technologie

Unter Screening (Früherkennung, Vorfelddiagnostik) wird eine auf eine bestimmte Krankheit gerichtete diagnostische Maßnahme ("Test") verstanden, welche darauf abzielt, noch symptomlose Krankheitsträger möglichst im Frühstadium zu erkennen und sie einer effektiven Behandlung zuzuleiten. Durch die Früherkennung können weniger aggressive Behandlungsmethoden angewandt, der Krankheitsverlauf abgeschwächt oder das Sterberisiko vermindert werden. Hauptsächlicher Zweck von Früherkennungsmaßnahmen ist die Senkung der krankheitsspezifischen Morbidität und/oder Mortalität.^{64,139}

Prinzipiell lassen sich verschiedene Typen von Screening unterscheiden. Das Massen-Screening (mass screening) zieht die Gesamtbevölkerung oder zumindest

einen ganzen Bevölkerungsteil mit ein, während das selektive Screening (selective screening) auf bestimmte Hochrisikogruppen in der Bevölkerung abzielt. In der Praxis ist diese Unterscheidung allerdings oft nicht möglich, da sich Früherkennungsmaßnahmen bei wegen altersabhängiger Unterschiede in der Inzidenz und Prävalenz von Erkrankungen bei bestimmten Neoplasien auf bestimmte Altersgruppen beschränken müssen. Nur bei bestimmten Tumorerkrankungen bestehen mit den derzeit vorhandenen Früherkennungsinstrumenten Möglichkeiten, Hochrisikogruppen anhand anderer Merkmale als dem Alter und Geschlecht zu definieren.¹³⁹ Einfacher lässt sich zwischen organisiertem Screening und opportunistischem bzw. spontanem Screening (opportunistic screening, case finding) unterscheiden. Das organisierte Screening ist populationsbezogen, wobei bestimmte Bevölkerungsteile oder Subpopulationen (z.B. Arbeitsplatzuntersuchungen, Immigranten), beispielsweise in Form von öffentlichen Kampagnen, zu einer Früherkennungsuntersuchung eingeladen bzw. aufgefordert werden. Die Initiative zur Früherkennung geht hier von den medizinisch Verantwortlichen aus. Das opportunistische Screening dagegen ist, beispielsweise als periodische Gesundheitsuntersuchung, Bestandteil der medizinischen Routineversorgung. Es findet, unabhängig vom Grund der ärztlichen Konsultation, durch Kontaktaufnahme der Patienten mit den medizinischen Versorgungsstrukturen und damit auf deren Initiative hin statt. Das opportunistische Screening kommt beispielsweise bei Blutdruckmessungen im Zusammenhang mit kardiovaskulären Erkrankungen, bei Abstrichen im Zusammenhang mit Gebärmutterhalskrebs oder bei digitalen rektalen Untersuchungen im Zusammenhang mit Prostata- und Mastdarmkrebs zum Tragen.^{64,139}

Konzeptionen zu Früherkennungsmaßnahmen basieren auf einer von zwei impliziten Annahmen. Gemäß der ersten kann die Entdeckung und Elimination von Risikofaktoren oder Vorformen einer Krankheit deren Ausbruch verhindern. Ein Beispiel hierzu ist die endoskopische Entdeckung und Entfernung von Polypen im Bereich des Dickdarms. Gemäß der zweiten Annahme kann die frühzeitige Entdeckung und Behandlung einer Krankheit in der präklinischen Phase die Prognose verbessern. Hierzu kann als Beispiel die Mammographie zur Früherkennung von Brustkrebs genannt werden. Den Diskussionen hinsichtlich der Früherkennung von Prostatakrebs liegt das zweite Motiv zugrunde.^{64,139}

C.2.2.1 Effektivitäts-Grundsätze des Screenings

Die Weltgesundheitsorganisation¹⁶¹ hat zehn Kriterien definiert, welche zur Effektivitätsbeurteilung eines Früherkennungsprogrammes in Betracht gezogen werden sollten und an denen sich auch der bundesdeutsche Gesetzgeber bei der Definition von Voraussetzungen für Früherkennungsuntersuchungen (§25 SGB V) teilweise orientiert hat.

Tabelle 2: Kriterien zur Beurteilung der Effektivität von Screening-Programmen

-
- Die Erkrankung sollte ein bedeutendes Gesundheitsproblem darstellen.
 - Die Erkrankung sollte ein diagnostisch erfassbares latentes oder präsymptomatisches Stadium aufweisen.
 - Zum natürlichen Verlauf der Erkrankung einschließlich des Übergangs vom latentem zum manifesten Stadium sollten adäquate Erkenntnisse vorliegen.
 - Für Patienten mit diagnostizierter Erkrankung sollte eine anerkannte Therapiemöglichkeit bestehen.
 - Für die Früherkennung der Erkrankung sollte ein ausreichend eindeutiges Test- bzw. Untersuchungsverfahren vorhanden sein.
 - Der Test bzw. die Untersuchung für die Früherkennung der Erkrankung sollte für die Zielpopulation annehmbar sein.
 - Es sollte Einvernehmen darüber bestehen, wer als Patient zu behandeln ist.
 - Die für die Früherkennung, Diagnose und Therapie benötigte Infrastruktur sollte vorhanden sein.
 - Die Kosten pro identifiziertem Erkrankungsfall (einschließlich Diagnose und Therapie) sollte in einem ökonomisch ausgewogenen Verhältnis zu den möglichen Gesamtausgaben im Gesundheitswesen stehen.
 - Die Früherkennung von Erkrankungsfällen sollte ein kontinuierlicher Prozess (Inzidenz-Screening) und kein Einmal-Projekt (Prävalenz-Screening) sein.
-

Als zentral sind insbesondere die folgenden Punkte anzusehen.

1. Bedeutung des Gesundheitsproblems

Aufgrund ethischer, praktischer und ökonomischer Aspekte sollte das Screening ausschließlich für Erkrankungen in Betracht gezogen werden, welche vom gesellschaftlichen Standpunkt her Gewicht haben. Dies bedeutet, dass die betreffenden Gesundheitsprobleme in einer Population verbreitet und/oder deren morbiditäts- und leidensbezogenen Auswirkungen schwerwiegend sein sollten. Zur Quantifizierung der gesellschaftlichen wie auch der individuellen Krankheitslast sind u.a. die Krankheitsprävalenz, das Erkrankungsalter bzw. die Altersverteilung der

Krankheit, die stadiumspezifische Prognose, Behandlungskomplikationen und Kosten zu berücksichtigen.^{64,139}

2. Diagnostische Methoden

Diagnostische Methoden zur Früherkennung sollten einfach anwendbar, kostengünstig und trennscharf sein. Eine weitere Voraussetzung für Screening-Tests ist eine gute Akzeptanz innerhalb der Zielgruppe. Sie sollten nicht als schmerzhaft, beschwerlich oder in anderer Weise unangenehm empfunden werden.

3. Behandlungsmöglichkeiten

Sollen Früherkennungsmaßnahmen erfolgreich sein, bedarf es nicht nur zuverlässiger und akzeptierter (differential-)diagnostischer Methoden, sondern auch wirksamer Behandlungsmöglichkeiten. Nach allgemeinem Konsens sollte die Effektivität von Therapien durch randomisierte kontrollierte Studien (randomized clinical trials, RCTs) belegt werden. Anhand solcher Studien sollte auch bestimmt werden, ob eine im Frühstadium eingeleitete Behandlung weniger aggressiv ausfällt als die Therapie nach klinischer Manifestation.^{64,139}

4. Kosten

Die Kosten für Screening, diagnostische Abklärung und Therapie sollten in angemessenem Verhältnis zur erzielten Wirkung stehen. Für die ökonomische Bewertung von präventiven Maßnahmen ist allgemein anzuführen, dass zwischen dem unverzüglichen Anfall von Kosten und dem Auftreten eines potentiellen Nutzens beachtliche Zeit vergehen kann. Gesamtgesellschaftliche Kosten werden leicht unterschätzt, da sie zum großen Teil, in Form von Zeitkosten, Transportkosten, physischen und psychischen Unannehmlichkeiten u.ä., von den Screening-Teilnehmern selbst getragen werden. Dieser Ressourcenverbrauch sollte in Analysen aus gesamtgesellschaftlicher Sicht berücksichtigt werden.¹³⁹

In Deutschland legt §25 Abs. 3 SGB V folgende gesetzliche Kriterien für Leistungen zur Früherkennung von Krankheiten fest:¹³⁴

1. die Krankheit muss wirksam behandelt werden können

2. das Vor- oder Frühstadium der Krankheit muss durch diagnostische Maßnahmen erfassbar sein
3. die Krankheitszeichen müssen medizinisch-technisch genügend eindeutig erfassbar sein
4. genügend Ärzte und Einrichtungen müssen vorhanden sein, um die aufgefundenen Verdachtsfälle eingehend zu diagnostizieren und zu behandeln.

C.2.2.2 Gütekriterien von Tests (Sensitivität, Spezifität, prädiktiver Wert)

Ideale Tests erfordern eine hohe Sicherheit, sowohl Erkrankte (Sensitivität) als auch Gesunde (Spezifität) zuverlässig zu klassifizieren.⁶⁴ Sensitivität und Spezifität sind Testcharakteristika, die für jedes diagnostische Einsatzgebiet und die jeweils dafür vorgesehene Zielpopulation neu beurteilt werden müssen. Dies geschieht in der Regel durch den Vergleich des Testverfahrens mit dem besten verfügbaren diagnostischen Referenzverfahren (Goldstandard). Die Angaben zur diagnostischen Validität des Tests hängen demnach auch von der Qualität des Goldstandards ab.¹⁵⁴ Die Ermittlung der Testcharakteristika von Screening-Tests gestaltet sich schwierig, wenn der Goldstandard wegen übermäßiger Invasivität nicht ohne begründeten Verdacht durchgeführt werden kann.^{27,61,64,71,118,139} Dieser Situation begegnet man häufig im Bereich des Tumor-Screenings.

Besonderes Augenmerk ist im Bereich des Screenings auf die Spezifität zu richten. Da die Zielpopulation von Screening-Maßnahmen per Definition aus symptomlosen und damit überwiegend gesunden Personen besteht, führt geringe Spezifität zwangsläufig zu einer großen Zahl falschpositiver Testergebnisse. Dies bedeutet wiederum, dass die Sicherheit, mit der aus dem Auftreten eines positiven Testresultates auf das Vorliegen der Erkrankung geschlossen werden kann (prädiktiver Wert des positiven Tests oder positiver prädiktiver Wert, positive predictive value, PPV), gering ausfällt. Negative Folgen von falschpositiven Tests bestehen im zusätzlichen Aufwand und Gefährdungspotential von abklärenden diagnostischen Verfahren und in der mit einem positiven Befund verbundenen psychischen Belastung.^{64,139}

Es sei darauf hingewiesen, dass Sensitivität und Spezifität als Funktion des gewählten Grenzwerts zur Trennung von Gesunden und Kranken (diagnostic cut-off) zu verstehen sind.⁶⁴ Durch eine Verschiebung des Grenzwerts kann auf beide

Charakteristika Einfluss genommen werden. Allerdings wird die Verbesserung einer Größe in der Regel mit der Verschlechterung der anderen Größe erkaufte. Bei der Festsetzung eines Grenzwertes muss daher entschieden werden, inwieweit der Sensitivität oder der Spezifität eines Tests Vorrang gegeben wird, was ein Abwägen zwischen den Folgen von Fehlzuordnungen, d.h. von falschpositiven und falschnegativen Testergebnissen erfordert.⁶⁴

Durch Kombination von Tests kann zusätzlicher Einfluss auf das Trennverhalten genommen werden. Hier bieten sich parallele oder sequentielle Teststrategien an. Bei der parallelen Anwendung von Tests wird das Ergebnis als positiv gewertet, wenn entweder alle Tests positiv sind - hier wird die Sensitivität gegenüber den Einzeltests niedriger, die Spezifität höher - oder mindestens ein Test positiv ausfällt - hier steigt die Sensitivität gegenüber den Einzeltests, während die Spezifität abnimmt.^{64,154} Bei der sequentiellen Anwendung von Tests wird der Einsatz weiterer diagnostischer Maßnahmen vom Ergebnis des jeweils vorangehenden Tests abhängig gemacht. Normalerweise wird bei negativem Ausfall des Tests auf eine weitere diagnostische Abklärung verzichtet, womit das sequentielle Testen zu einer Erhöhung der Spezifität führt.^{64,154}

C.2.2.3 Systematische Fehler

Sowohl in Studien zur Ermittlung von Testcharakteristika als auch bei der Evaluation von Screening-Programmen gilt es, systematische Fehler zu vermeiden. Im folgenden werden die wichtigsten Fehlerquellen kurz besprochen.

Verification bias oder (diagnostic) work-up bias

Wie oben erwähnt sollte bei der adäquaten Ermittlung diagnostischer Gütekriterien das Vorliegen oder Nichtvorliegen von Erkrankung bei allen Mitgliedern der Untersuchungspopulation durch ein diagnostisches Referenzverfahren gesichert werden. Beeinflussen die Ergebnisse des Tests die Durchführung der weiteren diagnostischen Abklärung des Krankheitsstatus, wird das Referenzverfahren letztlich lediglich an einer selektierten Untergruppe der Zielpopulation angewandt. Hier kommt es zu einem verification bias (work-up bias), wenn Sensitivität und Spezifität des Tests entweder lediglich für die Untergruppe mit referenzgesichertem Erkrankungsstatus, oder für die gesamte Zielpopulation unter der fraglichen Annahme, dass Probanden mit negativem Testergebnis gesund sind, ermittelt

werden.^{27,61,64,69,71,84,118,139} Im ersten Fall werden Sensitivität und Spezifität des Tests über- oder unterschätzt. Im zweiten Fall kann die Spezifität reliabel sein, da falschpositive Testergebnisse durch das Abklärungsverfahren aufgedeckt werden, jedoch wird die Sensitivität überschätzt. Unter bestimmten Umständen kann dieser Bias kontrolliert werden.^{27,69}

Diagnostic suspicion bias oder diagnostic review bias

Die Ergebnisse eines Tests und des Referenzverfahrens sollten unabhängig voneinander und mit Verblindung der Untersucher hinsichtlich der Resultate des jeweils anderen Untersuchungsverfahrens ermittelt werden. Damit soll vermieden werden, dass das Ergebnis bzw. die Interpretation des einen Verfahrens die Durchführung, Interpretation und/oder Resultatermittlung des jeweils anderen beeinflusst, was als diagnostic suspicion bias oder diagnostic review bias bezeichnet wird. Dieser bias ist entweder mit einer Überschätzung oder einer Unterschätzung der Testcharakteristika verbunden.^{61,69,71,118}

Selection bias oder volunteer bias

Teilnehmer an einer Früherkennungsmaßnahme können sich von der Zielpopulation in wesentlichen Aspekten unterscheiden. Teilnehmer eines Screening-Programms haben tendenziell ein ausgeprägteres Gesundheitsbewusstsein. Dieses steht entweder im Zusammenhang mit einem gesünderen Lebensstil, welcher sich in einem unterdurchschnittlichen krankheitsspezifischen Mortalitätsrisiko ausdrücken kann; oder aber es bezieht sich aufgrund bestimmter genetischer, klinischer oder lebensstilbezogener Faktoren auf die Kenntnis des Vorliegens eines erhöhten Risikos, in welchem Fall, unabhängig von der Wirksamkeit von Früherkennungsmaßnahmen, mit einem überdurchschnittlichen Mortalitätsrisiko zu rechnen ist. Richtung und Größe des aufgrund von Eigenschaften von Teilnehmern bzw. Nichtteilnehmern entstehenden potentiellen systematischen Fehlers sind schwierig zu bestimmen.⁶⁴

Lead time bias

Mit lead time wird der Zeitraum bezeichnet, der sich durch die screeningbedingte Vorverlegung des Diagnosezeitpunkts ergibt.^{15,16,64,139} Lead time hängt im

Einzelfall vom natürlichen Verlauf der Erkrankung ab und ist individuell nicht bestimmbar. Allerdings kann die Verteilung von lead time durch den Vergleich der Inzidenz von Krankheitsdiagnosestellungen in unterschiedlichen Zeiträumen bei gescreenten und nichtgescreenten Populationen ermittelt werden.¹³⁹ Diese Verteilung ist abhängig von der Sensitivität des betreffenden Screening-Tests, von der diagnostischen Praxis vor der Implementierung des betreffenden Screening-Programms, von der Krankheitsinzidenz sowie von der individuell unterschiedlichen Dauer des präklinischen Stadiums, während derer eine Früherkennung der Erkrankung überhaupt möglich ist (detectable preclinical phase).¹³⁹

Durch die Nichtberücksichtigung von lead time kann es beim Vergleich der Überlebenszeiten von gescreenten und nichtgescreenten Populationen zu einem systematischen Fehler kommen, da eine Vorverlegung der Diagnosestellung zu einer scheinbaren Verlängerung der Restlebenszeit führt, selbst wenn das Screening ohne Einfluss auf die Überlebenszeit ist. Aus diesem Grund kann aus (unbereinigten) Überlebenszeitkurven keine reliable Information über den Nutzen einer Screening-Maßnahme gezogen werden.^{15,16,64,139}

Length bias

Die Wahrscheinlichkeit für die Entdeckung einer Erkrankung mittels Screening-Maßnahme ist proportional der Dauer des präklinischen Stadiums, welche ihrerseits umgekehrt proportional der Entwicklungsrate der Erkrankung ist. Wird das Screening wiederholt durchgeführt, ist die Wahrscheinlichkeit, während des Intervalls zwischen zwei Früherkennungsuntersuchungen das symptomatische Stadium erreicht zu haben, für schnell fortschreitende Erkrankungen höher als für sich langsam entwickelnde. Unter der Annahme, dass Entwicklungsrate und biologische Aggressivität bzw. Prognose korrelierende Phänomene sind - wofür in der Praxis meist keine eindeutigen Beweise vorliegen -, würde dies bedeuten, dass Früherkennungsmaßnahmen überrepräsentativ Erkrankungen mit geringer Entwicklungsrate und guter Prognose erfassen.^{15,16,64,139} Beim Vergleich der Überlebenszeiten von gescreenten und nichtgescreenten Populationen kann daher ein systematischer Fehler, der length bias, entstehen. Als Extrem sind hier Erkrankungsfrühformen zu nennen, welche kein oder allenfalls geringes Entwicklungs- bzw. Progressionspotential aufweisen und in ihrem natürlichen Verlauf zu Lebzeiten des Individuums im präklinischen, asymptomatischen Stadium verbleiben. Einige Autoren bezeichnen diese Fälle als "Pseudoerkrankung" und

betrachten den Aspekt von length bias in diesem Zusammenhang gesondert als overdiagnosis (bias).¹⁵ Ob und wie solche latenten Erkrankungsfälle durch Früherkennungsmaßnahmen entdeckt bzw. unterschieden werden können, ist allerdings unklar. Lead time bias, length bias und overdiagnosis können lediglich in RCTs adäquat kontrolliert werden, welche eine Schätzung ihres kombinierten Effektes erlauben.¹⁵

Stage migration bias

Auch im Zusammenhang mit der Stadieneinteilung von Erkrankungen kann sich eine Art von lead time bias und length bias auswirken. Durch modernere, intensivere und/oder systematischere Diagnoseverfahren bzw. -standards können mehr Erkrankungen in prognostisch günstigeren Phasen erfasst werden, als dies mittels der älteren Technologien möglich war. Dies bedeutet auch, dass es bei einer Übertragung der neueren Diagnosestandards auf die Stadieneinteilung der älteren Technologien zu einer Verschiebung von nunmehr als prognostisch ungünstiger diagnostizierbaren Fällen in das jeweils ungünstigere Stadium kommt (upward stage migration). Wird diese aufgrund von Fortschritten in der Diagnostik erfolgende Stadienbereinigung mit prognostisch besserer Klassifikation beim Vergleich von Kohorten im säkularen Verlauf nicht berücksichtigt, kommt es zu einem stage migration bias. Dabei werden die stadienspezifischen Überlebenszeiten der Kohorten, bei welchen die neueren Diagnoseverfahren angewandt wurden, günstigere Verläufe aufweisen, auch wenn mit Bezug auf das populationsbezogene Gesamtergebnis kein Effekt zu erkennen ist.^{15,44}

C.2.2.4 Effektivitätskriterien

Häufig werden gescreente und nichtgescreente Populationen hinsichtlich der Anzahl entdeckter Erkrankungsfälle, der Stadienverteilung oder der Überlebenszeit miteinander verglichen. Diese Effektivitätskriterien werden jedoch in ihrer Aussagekraft unvermeidlich durch die soeben beschriebenen systematischen Fehler beeinträchtigt.¹³⁹

Die Krankheitsinzidenz kann zur Ermittlung der Effektivität von Screening-Programmen, welche auf die Entdeckung von Vorformen der Erkrankung abzielen, eingesetzt werden. Damit der Nutzen der Entdeckung und Elimination dieser Vorformen angemessen ermittelt werden kann, muss die Intensität des

diagnostischen Follow-up für gescreente und nichtgescreente Gruppen allerdings gleich sein.¹³⁹

Als das Effektivitätsmaß mit der höchsten Validität gilt die krankheitsspezifische Mortalität. Sie ist im allgemeinen auch das vorrangige Ziel eines Screening-Programmes, wobei andere günstige Effekte wie der Möglichkeit einer weniger aggressiven Behandlung, die Verminderung von Leiden oder die Senkung der Kosten für adjuvanter und palliativer Therapie und Pflege nicht geringgeschätzt werden sollten. Auch mit der Verwendung der krankheitsspezifischen Mortalität zur Quantifizierung der günstigen Auswirkungen des Screenings sind Validitäts- und/oder Interpretationsprobleme verbunden. Als grundlegende Voraussetzung für die Validität sei hier lediglich die Anwendung gleicher Grundsätze zur Ermittlung und Klassifikation der Todesursachen bei gescreenten und nichtgescreenten Populationen genannt.¹³⁹

C.2.3 Beschreibung der Zielkondition "Prostatakarzinom"

C.2.3.1 Epidemiologie

Das Prostatakarzinom ist eine Erkrankung des höheren Lebensalters.^{2,116,127} Das Karzinom tritt selten vor dem 50. Lebensjahr auf und zeigt von allen Krebskrankheiten den extremsten Anstieg mit dem Lebensalter. Das mittlere Erkrankungsalter liegt im Saarland bei 74 Jahren, das mittlere Sterbealter bei knapp 77 Jahren.^{2,127}

Das Prostatakarzinom ist in Deutschland mit etwa 11% (1996) die dritthäufigste tumorbedingte Mortalitätsursache bei Männern und gleichzeitig die häufigste Todesursache unter den urologischen Tumoren.¹³⁶ Von den 11 916 registrierten Sterbefällen waren annähernd 10% der Verstorbenen im Alter zwischen 45 und 65 Jahren (alters- und geschlechtsspezifische Mortalität 11/100 000), 26% im Alter zwischen 65 und 75 Jahren (alters- und geschlechtsspezifische Mortalität 101/100 000) und 64% älter als 75 Jahre (alters- und geschlechtsspezifische Mortalität 510/100 000).¹³⁶ Mit jährlich mehr als 25 000 Neuerkrankungen (ca. 16% aller bösartigen Neubildungen) ist es die zweithäufigste Krebsart bei Männern.² Die jährliche Neuerkrankungsrate steigt mit dem Alter von etwa 50/100 000 bei 60jährigen bis auf über 600/100 000 bei über 80jährigen (Saarland) stark an.²

Innerhalb der Länder der Europäischen Union liegen die in Deutschland ermittelten Erkrankungsraten für Prostatakrebs im mittleren Bereich. Die höchsten Raten werden in Schweden, Finnland, den Niederlanden und Österreich verzeichnet, die niedrigsten in Griechenland, Italien, Spanien und Portugal.² Allerdings muss bei internationalen Vergleichen der Einfluss unterschiedlicher diagnostischer Methoden auf die Erkrankungsrate berücksichtigt werden.

Innerhalb der letzten Jahre ist in Deutschland statistisch ein kontinuierlicher Anstieg der Neuerkrankungsrate zu beobachten. Dem entgegen steht nur ein geringfügiger Anstieg der altersstandardisierten Mortalität, der auf die Zunahme der Sterberaten der über 85jährigen zurückzuführen ist. Es kann daher wahrscheinlich ein großer Teil des Inzidenzanstiegs auf die Einführung spezifischer Methoden in der Diagnostik zurückgeführt werden.²

Die relative 5-Jahres-Überlebensrate liegt bei 70%.² Durch den frühzeitigen Tod an Prostatakrebs gehen Männern in Deutschland jährlich 98 600 Lebensjahre verloren. Im Vergleich dazu sind bei Männern 403 000 verlorene Lebensjahre auf den frühzeitigen Tod infolge des Lungenkarzinoms zurückzuführen.

C.2.3.2 Anatomie und Pathologie

Die Prostata ist eine (im normalen Zustand) in Form und Größe kastanienähnliche, exokrine akzessorische Geschlechtsdrüse (Vorsteherdrüse), welche an der Basis der männlichen Harnblase ringförmig den oberen Teil der Harnröhre (Urethra) umschließt. Ihr Sekret mischt sich bei der Ejakulation dem Samenplasma aus den angrenzenden Samenbläschen bei und stimuliert die Beweglichkeit der Samenfäden.¹³⁹ Es werden drei anatomische Zonen unterschieden. Eine periphere Zone, welche etwa 65% der normalen Drüse umfasst, eine zentrale, etwa 25% der Drüse umfassende Zone, und eine Transitionalzone, welche etwa 10% der Drüse umfasst und in zwei bilateral-symmetrisch um die Harnröhre angelegte Lappen unterteilt ist.^{139,149}

Beim Prostatakarzinom handelt es sich in über 95% der Fälle um ein azinäres Adenokarzinom verschiedener Differenzierung.¹⁵⁸ Spezielle Formen, die insgesamt nur etwa 2% aller Karzinome ausmachen, sind Urothel- und Plattenepithelkarzinome sowie muzinöse Karzinome. Primäre Prostatasarkome machen weit unter 1% der bösartigen Prostatakarzinomen aus. Sie sind im Gegensatz zum Karzinom eine Erkrankung des jüngeren Alters und treten mit Ausnahme der Myosarkome oft schon

im Kindes- oder Jugendalter auf.^{125,148,152} Auf diese Tumorarten soll im folgenden nicht weiter eingegangen werden. Ebenfalls vom Karzinom abzugrenzen ist die benigne Prostatahyperplasie (BPH; Prostataadenom), die vornehmlich in den Transitionalanteilen der Drüse heranwächst, wobei das Wachstum als lokale Proliferation zunächst auf die periurethrale Region begrenzt ist und später das übrige Drüsengewebe zu komprimieren beginnt. Im Gegensatz dazu entwickelt sich das Prostatakarzinom vor allem in den peripheren Drüsenanteilen.^{26,125}

Das Karzinom wächst zunächst infiltrierend in die Prostata ein, wobei die Prostatakapsel primär eine gewisse mechanische Barriere darstellt. Bei organüberschreitendem Tumorwachstum kann es neben einer Invasion des periprostatichen und perirektalen Fettgewebes zur Infiltration der Samenblasen, des Blasenbodens und der Muskulatur des Beckenbodens kommen.¹⁵²

Die primäre Metastasierung erfolgt lymphogen über die perineuralen Lymphspalten in die Lymphknoten des Obturatoriusbereiches und dann in die iliakalen, sakralen und paraaortalen Lymphknoten. Das Ausmaß der Lymphknotenmetastasierung steigt mit zunehmender lokaler Tumorausbreitung und Entdifferenzierung. Etwa ein Viertel der Patienten zeigt schon lymphogene Metastasen, wenn die Prostatakapsel noch nicht überschritten ist.¹⁵² Bei Überschreiten der Kapsel werden Lymphknotenmetastasen wesentlich häufiger. Liegt eine lokale Infiltration der Samenblasen vor, muss bei der Mehrheit der Patienten mit einer lymphogenen Metastasierung gerechnet werden.^{125,152}

Die hämatogene Metastasierung betrifft vor allem das Skelettsystem. Hier sind die Knochen des Stammskeletts (Knochen des Beckens, der Lenden- und Brustwirbelsäule, Rippen) bevorzugt befallen, was auf direkte Verbindungen des regionären mit dem sakrospinalen Venensystem zurückgeführt wird.^{125,152} Viszerale Metastasen treffen bevorzugt die Lunge und Leber. Andere Organe werden erst spät befallen.^{125,152}

C.2.3.3 Ätiologie

Obwohl die eigentlichen Mechanismen der Karzinogenese - abgesehen von Fällen mit erworbener genetischer Mutation - nicht geklärt sind, scheint dem Lebensalter und endokrinen Faktoren eine bedeutende Rolle zuzukommen. Prostatakarzinome treten bei Eunuchen nicht auf, während andererseits das Androgen Dihydrotestosteron, ein in der Prostata gebildeter aktiver Metabolit von Testosteron,

in jedem Alter einen mehr oder weniger starken Wachstumsreiz auf das Prostataepithel ausübt.^{26,125,139} Auch die Tatsache, dass Prostatakarzinome zumindest vorübergehend erfolgreich durch kontrase sexuelle Therapieformen (Androgenentzug) behandelt werden können, sowie die Möglichkeit, Prostatakarzinome im Tierexperiment durch Verabreichung von testikulären Hormonen zu induzieren, stützt diese Hypothese.^{26,125,139,152} Genetische Faktoren (Familienanamnese, Ethnie), Diät und chemische Karzinogene werden ebenfalls diskutiert, doch liegen hierzu keine eindeutigen Ergebnisse vor.^{26,129,139,149,152}

C.2.3.4 Klinik und Verlauf

Der Wissensstand zum natürlichen Verlauf des Prostatakarzinoms ist unergiebig. Bei den Progressionsraten besteht ein breites Variationsspektrum. Viele maligne Veränderungen bleiben zu Lebzeiten des Erkrankungs trägers latent. Die Diskrepanz zwischen der hohen Prävalenz von Prostatakarzinom in selektierten Populationen und der relativen Seltenheit des Karzinoms als eigentliche Todesursache bzw. der Prävalenz von klinisch manifestem Karzinom wurde in mehreren Autopsiestudien mit durchschnittlichen Prävalenzen zwischen 14% und 48% festgestellt (hinsichtlich dieser Variationsbreite sei u.a. auf die Abhängigkeit der Prävalenz von der Altersverteilung in der jeweiligen Stichprobe, von der Art der histologischen Schnitte und den Selektionskriterien für die Autopsie hingewiesen.^{110,149} Auch wenn damit die Mehrheit der Prostatakarzinome nur langsam zu wachsen und keine Gefahr für das Leben darzustellen scheint, weisen einige Fälle eine schnelle Progression mit Infiltration benachbarter Gewebe und Metastasierung auf. Eine Unterscheidung von organbeschränkten malignen Veränderungen hinsichtlich ihres Wachstums- bzw. Progressionspotentials ist derzeit nicht möglich.^{26,129,139,149}

Die Symptomatik des Prostatakarzinoms variiert in Abhängigkeit von der Tumorausdehnung. Im frühen Stadium ist das Prostatakarzinom symptomlos. Erst bei lokaler Progression mit Infiltration der periurethralen Drüsenanteile oder bei einer gleichzeitig bestehenden BPH treten Blasenentleerungsstörungen auf. Pollakisurie, Dysurie, Harnverhalt und/oder Makrohämaturie gehören mit zu den ersten Symptomen. Eine Skelettmetastasierung geht zum Teil mit erheblichen Schmerzen einher, pathologische Frakturen sind möglich.^{125,152} Hier entspricht das klinische Erscheinungsbild einer Spätsymptomatik. Eine Behandlung mit kurativer Zielsetzung ist bei ihrem Auftreten nicht mehr möglich.¹⁵² Bei lokal weit fortgeschrittenem Tumor sind Defäkationsbeschwerden möglich.¹²⁵ Trotz der Möglichkeit der Vorsorge-

untersuchung haben weit über die Hälfte aller Patienten mit einem Prostatakarzinom bereits bei der Erstdiagnose ein nicht heilbares Stadium (\geq Stadium T3) erreicht.¹⁵²

C.2.3.5 Früherkennung und Diagnose

Verbreitete Verfahren zur Früherkennung bzw. Diagnose des Prostatakarzinoms sind die digitale rektale Untersuchung (DRU), die Bestimmung des PSA im Serum, die transrektale Ultraschalluntersuchung (TRUS) und die transrektale Biopsie. Diese Verfahren werden meist in einer parallelen oder sequentiellen Teststrategie kombiniert. Eine detaillierte Darstellung der einzelnen diagnostischen Verfahren und ihrem spezifischen Einsatz in Früherkennung bzw. Diagnose erfolgt weiter unten im Rahmen der Beschreibung der Technologie "PSA" und der Intervention "PSA-Screening".

C.2.3.6 Staging, Grading und weitere Prognosefaktoren

In erster Linie werden das Tumorstadium (Staging) und der histologische Differenzierungsgrad (Grading) für die Abschätzung der Prognose und die Therapieplanung herangezogen.

Tumorstadium - Staging

Hauptkriterien der klinischen und pathologischen Stadieneinteilung sind das Tumolvolumen, das Ausmaß der Tumorausbreitung und das Ausmaß der (Lymphknoten-)Metastasierung. Die Klassifikation erfolgt nach verschiedenen Einteilungssystemen. In Nordamerika wird überwiegend die Whitmore-Jewett (A bis D)-Klassifikation, in Europa vornehmlich eine modifizierte TNM-Klassifikation verwendet: T - "Tumor" (Primärtumor), N - "Nodi" (Lymphknotenmetastasen), M - "Metastasen" (Fernmetastasen). Tabelle 3 führt die Kriterien der verschiedenen Stadien auf.

Tabelle 3: Staging beim Prostatakarzinom nach dem Whitmore-Jewett- und TNM-System

	Klinisches Stadium ¹		Definition	
	Whitmore-Jewett-System	TNM-System		
		4. Aufl., 2. Revision 1992		5. Aufl. 1997
		TX T0	TX T0	Primärtumor kann nicht beurteilt werden Kein Anhalt für Primärtumor
A	T1	T1	T1	Klinisch weder durch Palpation noch in bildgebenden Verfahren erkennbarer Tumor:
A1	T1a	T1a	T1a	Zufälliger histologischer Befund (inzidentelles Karzinom) in ≤5% des resezierten Gewebes
A2	T1b	T1b	T1b	Zufälliger histologischer Befund (inzidentelles Karzinom) in >5% des resezierten Gewebes. Mässige oder geringe Differenzierung bei ≤5% des resezierten Gewebes ² .
B0	T1c	T1c	T1c	Diagnose mittels Nadelbiopsie (z.B. wegen erhöhter PSA-Werte)
B	T2	T2	T2	Tastbarer, scheinbar organbegrenzter Tumor³:
B1	T2a	T2a	T2a	Tumor befällt ≤Hälfte eines Lappens
B1	T2b	T2a	T2a	Tumor befällt >Hälfte eines Lappens, aber nicht beide Lappen
B2	T2c	T2b	T2b	Tumor befällt beide Lappen
C	T3	T3	T3	Extrakapsuläre Ausbreitung des Tumors⁴:
C1	T3a	T3a	T3a	Einseitige extrakapsuläre Ausbreitung, Samenblasen frei
C1	T3b	T3a	T3a	Beidseitige extrakapsuläre Ausbreitung, Samenblasen frei
C2	T3c	T3b	T3b	Infiltration der Samenblase(n)
C2	T4	T4	T4	Tumor fixiert oder mit Infiltration anderer benachbarter Strukturen als Samenblasen ⁵
C2	T4a	T4	T4	Infiltration von Blasenhal, Sphincter externus und/oder Rektum
C2	T4b	T4	T4	Infiltration des Levatormuskels und/oder Fixation an Beckenwand
D	N	N	N	Metastasierung - Regionale Lymphknoten^{6,7}:
	NX	NX	NX	Regionäre Lymphknoten können nicht beurteilt werden
	N0	N0	N0	Keine regionären Lymphknotenmetastasen
D1	N1	N1	N1	Regionäre Lymphknotenmetastase(n) - Metastase(n) in solitärem Lymphknoten, ≤2cm in größter Ausdehnung

D1	N2	N1	Regionäre Lymphknotenmetastase(n) - Metastase(n) in solitärem Lymphknoten, >2cm bis 5cm in größter Ausdehnung, oder in multiplen Lymphknoten, ≤5cm in größter Ausdehnung
D1	N3	N1	Regionäre Lymphknotenmetastase(n) - Metastase(n) in Lymphknoten, > 5cm in größter Ausdehnung
D	M	M	Metastasierung - Fernmetastasen:
	MX	MX	Fernmetastasen können nicht beurteilt werden
	M0	M0	Keine Fernmetastasen
D2	M1a	M1a	Fernmetastase(n) - nicht regionäre(r) Lymphknoten
D2	M1b	M1b	Fernmetastase(n) - Knochen
D2	M1c	M1c	Fernmetastase(n) - andere Lokalisation(en)

¹ Pathologische Klassifikation (pTNM): die pT-, pN- und pM-Kategorien entsprechen den T-, N- und M-Kategorien.

² Hinsichtlich der Definition des Stadiums A2 differieren die Kriterien für den histologischen Differenzierungsgrad und den Anteil rezesierten Gewebes zwischen den einzelnen Studien.

³ Ein Tumor der mittels Nadelbiopsie in einem oder beiden Lappen gefunden wird, aber weder tastbar noch in bildgebenden Verfahren erkennbar ist, wird als T1c klassifiziert.

⁴ Invasion in den Apex der Prostata oder in die Prostatakapsel; aber nicht durch diese hindurch in extrakapsuläres Gewebe, wird als T2 (nicht als T3) klassifiziert.

⁵ Die mögliche Tumorformel T4N0M0 kommt praktisch nicht vor, weil organüberschreitende Tumoren in der Regel metastasiert sind.

⁶ Regionäre Lymphknoten sind die Lymphknoten des kleinen Beckens, die im wesentlichen den Beckenlymphknoten unter der Bifurkation der Aa. iliaca communes entsprechen. Lateralität beeinflusst die N-Klassifikation nicht.

⁷ Wenn Metastasen in mehr als einer Lokalisation nachweisbar sind, wird die höchste Kategorie benutzt.

Für das lokoregionäre Prostatakarzinom (Stadium T1 und T2) ist das Tumolvolumen der wichtigste unabhängige prognostische Indikator. Für das hinsichtlich des Progressionsrisikos als relevant angesehene Tumolvolumen reichen die Schätzungen von 0.5ccm bis 1.4ccm.²⁶ Organbegrenzte Prostatakarzinome mit einem Volumen von 4.0ccm bis 6.0ccm werden (unter der Voraussetzung tumorfreier Absetzungsränder) als potentiell heilbar angesehen. Bei lokoregionär fortgeschrittenen Tumoren, insbesondere bei solchen mit pelvinem Lymphknotenbefall mit oder ohne Fernmetastasierung, ist eine kurative Behandlung mit den heutigen Therapieformen nicht möglich.^{26,129,139,149} Kein klinisches oder bildgebendes Untersuchungsverfahren ist derzeit jedoch in der Lage, das lokale Ausmaß eines Prostatakarzinoms mit ausreichender Sicherheit zu bestimmen. Ebenso wenig können pelvine Lymphknotenmetastasen mit akzeptabler Sicherheit erkannt werden.¹⁵⁸ Oft wird die Infiltration der Prostatakapsel oder der Samenbläschen sowie das Vorliegen von Metastasen unterschätzt (understaging).

Eine eindeutige Stadieneinteilung ist lediglich nach radikaler Prostatektomie möglich (pathologische Klassifikation). Präoperative Volumenangaben und klinische Stadieneinteilungen haben demnach nur eine eingeschränkte Aussagekraft. Dadurch wird die Anzahl potentiell heilbarer Karzinome häufig überschätzt.^{26,129,149}

Histologischer Differenzierungsgrad - Grading

Prostatakarzinome zeigen große Unterschiede hinsichtlich Morphologie und Zytologie und variieren dementsprechend in ihrem biologischen Verhalten.¹⁵⁸ Zur Erfassung dieser Unterschiede wurden eine Reihe verschiedener Grading-Systeme publiziert. Im angelsächsischen Raum ist das Grading nach Gleason am weitesten verbreitet. Es basiert auf der Bewertung des strukturellen Karzinomaufbaus unter Berücksichtigung unterschiedlich differenzierter Baumuster.¹⁴⁸ In einer Skala von 1 bis 5 werden die zwei vorherrschenden Strukturtypen mit Punkten entsprechend dem jeweiligen Malignitätsgrad (mit 1 bei hoher Differenzierung und 5 bei Entdifferenzierung) bewertet. Aus der Summe der Punktwerte ergibt sich der sogenannte Gleason-Score.¹⁴⁸

Tabelle 4: Grading beim Prostatakarzinom nach Gleason (1992)

Vergebene Punkte	Morphologische Kriterien
1	Runde bis ovale gleich große Einzeldrüsen, dicht nebeneinander liegend, scharf gegen die Umgebung abgegrenzt.
2	Etwas weniger uniforme Einzeldrüsen, getrennt durch geringe Mengen von Stroma, weniger scharf begrenzter Tumorrand.
3	a) Unregelmäßig große und unregelmäßig gestaltete Drüsen mit gewöhnlich reichlicherem Stroma, gelegentlich auch dicht gelagert, unregelmäßige und unscharfe Tumorgrenze. b) Papilläre oder kribriforme Strukturen, z.T. in großen gangähnlichen Bildungen.
4	a) Große unregelmäßige Epithelformationen durch Drüsenverschmelzung (engl. fused glands) sowie verzweigte Drüsen mit unregelmäßiger Infiltration in die Umgebung. b) Adenokarzinom mit ausgeprägt klarem Zytoplasma, ähnlich hellzelligen Adenokarzinomen der Niere.
5	a) Scharf begrenzte runde Epithelhaufen mit meist solidem und kribriformem Bau, gewöhnlich mit zentraler Nekrose (komedo-karzinomähnlich). b) Unregelmäßig begrenzte Formationen eines undifferenzierten Karzinoms, das nur durch gerade noch erkennbare Drüsenbildung oder Vakuolen (siegelringähnlich) als Adenokarzinom zu identifizieren ist.

Beurteilung erfolgt nach Gesamtpunktzahl (Gleason-Score). Berechnung der Gesamtpunktzahl:

- Einheitlich strukturierte Tumoren: Verdoppelung der nach obiger Tabelle vergebenen Punkte
- Unterschiedlich strukturierte Tumoren: Punkte für überwiegenden Anteil (sog. primäres Grading) + Punkte für Anteil, der zweitgrößte Komponente bildet (sog. sekundäres Grading; wenn bei Stanzbiopsien mehr als zwei unterschiedliche Anteile vorhanden sind, wird für das sekundäre Grading der Anteil mit dem ungünstigsten Punktwert benutzt).

Die Gesamtpunktzahl kann zwischen 2 und 10 liegen.

Für prognostische Zwecke werden Gleason-Scores im allgemeinen in drei Gruppen unterteilt, welche dem WHO-Grading mit einer Einteilung in die Grade G1 bis G3 (hoch- mäßig- und geringdifferenziertes Prostatakarzinom) vergleichbar sind:

G1 = Gleason-Score 2-4

G2 = Gleason-Score 5-7

G3 = Gleason-Score 8-10.

Da sich der Gleason-Score auf strukturelle und nicht zytologische Kriterien bezieht, kann er aus zytologischen (Biopsie-)Proben nicht abgeleitet werden.²⁶

Ergänzend sei hier als drittes Grading-System das im deutschen Sprachraum verwendete Grading nach den Empfehlungen des Pathologisch-urologischen Arbeitskreises Prostatakarzinom genannt. Diese Einteilung berücksichtigt in einer

Punkteskala sowohl strukturelle wie auch zytologische Kriterien und gelangt in Abhängigkeit davon zu einem Malignitätsgrad Ia, Ib, IIa, IIb, IIIa oder IIIb.^{148,158}

Es liegen einige Belege dazu vor, dass eine Reihe von Tumoren im Laufe der Zeit zu einem höher bewerteten Malignitätsgrad fortschreiten und der histologische Differenzierungsgrad demnach als guter Prädiktor für die Tumorprogression angesehen werden kann.^{26,139}

Andere Prognosefaktoren

Eine Vielzahl alternativer und neuer prognostischer Indikatoren wie DNA-Ploidie, Kernmorphometrie, Proliferationsmarker, Nachweis zirkulierender Tumorzellen mit RT-PCR, Tumor-Suppressor-Gene, Onkogene, neuroendokrine Differenzierung sowie zytogenetische Analysen wurden beim Prostatakarzinom überprüft. Bislang fehlen jedoch eindeutige Daten, die belegen, dass sie dem traditionellen Staging und Grading überlegen sind.^{26,139,149}

Prämaligne Veränderungen

Wie bei vielen maligne Tumoren epithelialer Herkunft scheint die Entstehung des Prostatakarzinoms ein komplexer, multifaktorieller Prozess zu sein, welcher, von normalem Epithel ausgehend, über mehrere Dysplasie- und Neoplasiestufen zum invasiv proliferierenden Malignom fortschreitet. Prostatistische intraepitheliale Neoplasien (PIN) werden als Präkanzerosen bzw. carcinoma in situ angesehen. Sie weisen zwar morphologische und phänotypische Ähnlichkeiten mit Karzinomzellen auf und finden sich auch oft in unmittelbarer Nachbarschaft eines Malignoms, doch besteht für eine Proliferation keine direkte Evidenz. Auch wenn angenommen werden kann, dass das PIN-Stadium der Entwicklung einiger manifester Prostatakarzinome vorausgeht, ist der Anteil dieser Fälle und die Verteilung der Zeit einer möglichen Früherkennung in dieser PIN-Phase unbekannt. Der Befund einer sog. high grade-PIN mit fragmentierter Basalzellschicht und ausgeprägter Atypie in Stanzbiopsien oder Resektatspänen erfordert eine komplette Aufarbeitung des Untersuchungsmaterials oder weitere Kontrolluntersuchungen. Ergibt sich kein Karzinom, sind aus der Diagnose einer PIN keine weitergehenden therapeutischen Konsequenzen abzuleiten.

Auch wenn sie oft in Kombination auftreten, sind Adenom (BPH) und Karzinom grundsätzlich verschiedene Erkrankungen. Die gehäufte Assoziation ist eher Ausdruck der hohen Prävalenz von BPH sowie der Tatsache, dass bei unauffälligem Primärbefund im Rahmen der histologischen Untersuchung von Präparaten nach transurethraler Resektion der Prostata (TUR-P) oder offener Operation wegen BPH oft sog. inzidentelle Karzinome diagnostiziert werden.

C.2.3.7 Therapie

Therapie mit kurativer Zielsetzung

Die Therapie des Prostatakarzinoms orientiert sich hauptsächlich am Tumorstadium, aber auch am histologischen Differenzierungsgrad sowie am Alter und Allgemeinzustand des Patienten. Für die Therapie von klinisch lokal begrenzten Karzinomen ohne Nachweis von Lymphknoten- und Fernmetastasen (Stadium T1-T2 N0 M0) kommen die abwartende Beobachtung (wait and see, watchful waiting), die radikale Prostatovesikulektomie (Prostatektomie) sowie die Strahlentherapie in Frage. Operation und Radiotherapie werden dabei in kurativer Absicht durchgeführt.^{26,139,148,152,158} Vermerkt sei hierbei, dass die Hälfte der klinisch als T2 klassifizierten Tumoren das Organ tatsächlich bereits überschritten haben und vom Pathologen als pT3 eingestuft werden.¹⁵⁸

(1) Watchful waiting

Diese Strategie wird mehrheitlich in Großbritannien und Skandinavien verfolgt.^{26,139} Sie besteht in der zumeist kontrollierten Beobachtung der Patienten über die diagnostische Abklärung von Symptomen oder durch Monitoring mittels DRU, PSA-Bestimmung, bildgebender Verfahren und/oder wiederholter Biopsien. Bei Auftreten oder Verstärkung von Symptomen oder mittels der Verlaufskontrollmethoden festgestellter Tumorprogression sind operative Eingriffe zur Beseitigung von Harnabflussbehinderung (z.B. TUR-P), Strahlenbehandlung oder eine hormonelle Therapie indiziert. Nebenwirkungen des watchful waiting können im Zusammenhang mit Miktionsbeschwerden und deren Behandlung stehen oder psychischer Natur sein.²⁶

(2) Radikale Prostatektomie

Bei der radikalen Prostatektomie werden Prostata, anhängende Samenbläschen, die zentralen Anteile beider Ducti deferentes und die prostatistische Harnröhre entfernt. Danach werden Blasen Hals und Harnröhre durch Naht wieder vereinigt.^{152,156} Am häufigsten wird der retropubische Zugang angewandt. Im Vergleich zum perinealen Vorgehen bietet er den Vorteil, dass gleichzeitig eine pelvine Lymphadenektomie durchgeführt werden kann.¹⁵⁸ Die Kombination aus laparoskopischer pelviner Lymphadenektomie und perinealer Prostatektomie soll den Vorteil der geringeren Morbidität haben. Vergleichende Studien hierzu liegen jedoch nicht vor.¹⁵⁸ Die Indikation zur radikalen Prostatektomie für das lokal weiter fortgeschrittene Karzinom (Stadium T3) wird kontrovers diskutiert.^{156,158}

Zu den typischen Komplikationen der radikalen Prostatektomie gehören Anastomosenstrikturen (6-18%), Anastomoseninsuffizienz (1.2-4%), Rektumverletzung, (2-11%) und Harninkontinenz (postoperativ zum Zeitpunkt der Entlassung bis zu 59%, milde und schwere Formen 18 Monate nach Entlassung 4-21% bzw. 0-7%).^{139,149,158} Die perioperative Mortalität liegt zwischen 0.2% und 1.2%. Häufige Nebenwirkung der radikalen Prostatektomie ist außerdem eine erektile Dysfunktion, die bei über 90% der Patienten auftritt.¹⁵⁸ Die nerverhaltende Prostatektomie ermöglicht bei einem Teil der Patienten einen Potenz-erhalt.^{125,139,149,158} Die potenzschonende radikale Prostatektomie ist den frühen Tumorstadien mit nur einseitig positiver Biopsie vorbehalten, weil sonst vermehrt mit lokalen Rezidiven gerechnet werden muss.^{156,158} Art und Schwere peri- und postoperativer Komplikationen werden durch die Patientenselektion und Erfahrung des Operateurs mitbestimmt.¹²⁵ Perioperative Morbidität und Mortalität sind alters- und komorbiditätsabhängig und steigen jenseits des 70. Lebensalters stark an.¹⁴⁸ Die Lebenserwartung sollte nach verschiedenen Empfehlungen mehr als zehn Jahre betragen, um mit der Operation einen Nutzen für den Patienten zu erzielen.^{148,149,158}

(3) Strahlentherapie

Die Strahlentherapie mit kurativer Zielsetzung wird entweder als externe Bestrahlung (Teletherapie) oder als interstitielle (oder intrakavitäre) Bestrahlung durchgeführt.^{139,149,158} Dabei werden in den meisten Patientenserien 50Gy bis 70Gy drei bis sechs Wochen auf die Prostata appliziert.^{26,149} Durch die dreidimensionale Bestrahlungsplanung und die konformierende Bestrahlungstechnik

sind die Möglichkeiten der Strahlentherapie des Prostatakarzinoms erheblich verbessert und die Akut- und Spätnebenwirkungen in Frequenz und Ausprägung gesenkt worden.¹⁵⁸

Typische Komplikationen der Bestrahlung sind akute Proktitis und Urethritis (40-60%), welche in der Regel komplett reversibel sind. Zu den chronischen Komplikationen gehören Harnröhrenstriktur (1-5.2%), Inkontinenz (2-36%) und Impotenz (25-40%).^{148,156,158}

Für den Effekt der einzelnen Behandlungsmöglichkeiten auf die Überlebenszeit liegen bislang keine Daten aus prospektiven RCTs vor. Eine allgemein akzeptierte Therapie gibt es nicht, die Effektivität der aggressiven Therapie ist fraglich.^{4,26,32,51,129,139,149,155} Der American Urological Association on Prostate Cancer Clinical Guidelines Panel ist aufgrund einer Literaturanalyse zu der Ansicht gekommen, dass es bislang keine klare Evidenz für die Überlegenheit einer dieser Optionen gibt, und dass die Meinung des entsprechend aufgeklärten Patienten als Standard angesehen werden sollte.¹⁵⁸ In Deutschland wird von den Urologen für Patienten mit einer Lebenserwartung von mehr als zehn Jahren die radikale Prostatektomie bevorzugt.¹⁵⁸

Therapie mit palliativer Zielsetzung

Bei fortgeschrittener Krebserkrankung (Stadium (p)T3-T4 N+ M+) kann nur noch palliativ therapiert werden.¹⁵²

(1) Hormone

Das Ziel einer Hormontherapie bei Patienten mit Prostatakarzinom besteht darin, den Karzinomzellen die wachstumsfördernden Androgene zu entziehen. Da jedoch nur ein Teil der Tumorzellen androgenabhängig ist, lässt sich die Tumorprogression durch Hormontherapie lediglich verzögern. Eine testikuläre Androgendeprivation kann entweder chirurgisch (bilaterale Orchiektomie) oder medikamentös (LHRH- oder GnRH-Analoga oder Östrogene) erfolgen. Zur zusätzlichen Blockierung der Nebennierenandrogene können Antiandrogene verwendet werden.^{26,139,152,158} Erhöhte kardiovaskuläre Komplikationen infolge hoher Östrogengaben führten in den letzten Jahren zu einer Ablehnung dieser Therapieform¹⁵⁸ Hauptnebenwirkungen der anderen Formen des Androgenentzugs sind Impotenz, Hitzewallungen und Gynäkomastie (Antiandrogene). Bislang steht der Beweis aus, ob und inwieweit die

endokrine Therapie das Leben der betroffenen Patienten verlängern kann oder ob sie lediglich die Lebensqualität verbessert bzw. das symptomfreie Intervall verlängert.¹⁵⁸

(2) Chemotherapie

Bei hormonrefraktären Tumoren (Hormoninsensitivität des Karzinoms) steht die Lebensqualität des Patienten und die Rückbildung von Knochenschmerzen, Harnstauungsniere und Lymphödemen der unteren Extremitäten im Vordergrund. Eine zumindest kurzfristige subjektive Besserung durch Chemotherapie lässt sich bei der Mehrzahl der Patienten erzielen. Ein kurativer Effekt der Zytostase ist allerdings nicht zu erwarten.¹⁵⁸

(3) Therapiekombinationen

Insbesondere für die Behandlung des lokal fortgeschrittenen Karzinoms (pT3 N0 M0) und das Karzinom mit regionären Lymphknotenmetastasen werden verschiedene Therapiekombinationen bzw. adjuvante Therapieformen diskutiert, welche jedoch nicht als kurative Behandlungskonzepte angesehen werden können. Daten aus prospektiven RCTs mit einer Gegenüberstellung von kombinierten und einzelnen Behandlungsmethoden liegen bislang nicht vor.¹⁵⁸

C.2.4 Beschreibung der Technologie "PSA-Bestimmung"

C.2.4.1 Entwicklungsstand der Technologie

Die Einführung der Bestimmung von PSA in die Palette klinischer Laboruntersuchungen im Jahre 1984 hat die Diagnostik und Überwachung von Patienten mit Prostatakarzinom grundlegend verändert.¹³⁰ Die Bestimmung von PSA im Serum ist für sich genommen ein relativ kostengünstiger Bluttest, der von den Patienten gut akzeptiert wird. Das PSA diente zunächst nur der Verlaufskontrolle bereits erkannter Prostatakarzinome, der Metastasendiagnostik und der Nachsorge nach radikaler Prostatektomie. In den letzten Jahren wird es jedoch zunehmend auch zur Diagnose von Prostatakarzinom herangezogen und ist zur Zeit der meist gebrauchte Parameter der urologisch-onkologischen Labordiagnostik.¹³⁰

Während früher im Rahmen der Diagnostik und Verlaufskontrolle vorwiegend der PSA-Wert verwendet wurde, werden heute vermehrt abgeleitete Werte wie der PSA-Prostatavolumen-Quotient (PSA-D), die Verlaufsbestimmung mit Ermittlung der PSA-Anstiegsgeschwindigkeit (PSA-V) und neuerdings der Quotient aus freiem und Gesamt-PSA (Q-PSA) als alleinige oder Zusatzparameter verwendet. Ferner führte die Einführung altersabhängiger PSA-Grenzwerte zu vielfältigen Strategien.^{4,26,32,51,59,129,130,139,148,149,157,158} Dies ist bei der zeitabhängigen Beurteilung des jeweiligen Standes der Technologie innerhalb des technologischen Lebenszyklus (life cycle) zu beachten. Ferner ist in der Zukunft mit weiteren Modifikationen der Parameter oder deren Kombination zu rechnen. Zu Einzelheiten bezüglich der genannten Parameter sei auf das Kapitel "Technische Beschreibung" weiter unten verwiesen.

Bezüglich der Beurteilung des life cycle der gesamten Screening-Strategie ist insbesondere auch die Entwicklung der Biopsietechniken zu berücksichtigen. Während früher gezielte Biopsien in auffälligen Areale durchgeführt wurden, hat sich in den letzten Jahren die systematische bzw. randomisierte Biopsie zur repräsentativen Untersuchung der Prostata durchgesetzt.^{26,129,148,158} Zusätzlich werden palpatorisch oder sonographisch verdächtige Areale gesondert biopsiert.^{129,149,158}

Weitere noch anhaltende Entwicklungen betreffen den Einsatz neuer PSA-Varianten wie PSA-D, PSA-V oder Q-PSA innerhalb des Screenings, von denen man sich eine erhöhte Spezifität erhofft.

C.2.4.2 Chemisch-biologische Wissensgrundlage

PSA ist ein Glykoprotein (Serinprotease), welches von normalem, benignem hyperplastischem und malignem Prostatagewebe in die Samenflüssigkeit sezerniert wird.^{26,139} Geringe Mengen gehen auch ins Blut über, wo seine Konzentration mittels verschiedener Immunoassays quantitativ bestimmt werden kann. Der normale PSA-Spiegel steigt mit dem Alter an, was teilweise auf die altersabhängige Zunahme des Prostatavolumens zurückzuführen ist. Pathologisch kommt es insbesondere bei einer Verletzung der Basalmembran der Prostatakapsel zu einer Erhöhung der PSA-Konzentration im Serum. Da malignes Gewebe offenbar vermehrt PSA exprimiert und bei Vorliegen eines Karzinoms eine höhere Wahrscheinlichkeit des Übertritts in den Blutkreislauf besteht, wird PSA als Tumormarker herangezogen. Allerdings ist PSA zwar gewebe-, nicht aber tumorspezifisch.^{4,26,125,129,139,149} Das Hauptproblem

der PSA-Bestimmung liegt in der Differenzierung zwischen BPH und lokalisiertem Prostatakarzinom. Hier besteht eine breite Grauzone sich überschneidender PSA-Konzentrationen, welche unterschiedliche prostatistische Veränderungen anzeigen können und damit insbesondere die Spezifität der PSA-Bestimmung bei der Tumordiagnose einschränken.^{51,129,139,149} Außerdem kann der Serum-PSA-Spiegel auch nach einer akuten Prostatitis, nach akuter Harnsperre, Zystoskopie, TRUS, transrektaler Biopsie oder nach einem operativem Eingriff an der Prostata für mehrere Wochen erhöht sein. Inzwischen scheint belegt, dass nach DRU keine signifikante PSA-Erhöhung auftritt.¹⁴⁹

C.2.4.3 Technische Beschreibung

Um die diagnostische Validität der PSA-Bestimmung für die Erkennung von klinisch relevantem, organbegrenztem Prostatakarzinom zu erhöhen, wurden neben Untersuchungen zur Festlegung eines Referenzbereiches verschiedene Möglichkeiten zur Steigerung der Malignomspezifität des PSA erprobt.^{4,26,32,51,59,129,130,139,148,149,157,158}

- (1) Die Bildung des PSA-Prostatavolumen-Quotienten (PSA-Dichte; PSA-density PSA-D), wobei das Volumen sonographisch bestimmt wird. Basierend auf dem Prostatavolumen lässt sich alternativ ein Vorhersagewert für den PSA-Wert (predicted PSA pPSA) ermitteln.
- (2) Die Einführung altersabhängiger PSA-Grenzwerte.
- (3) Die Verlaufsbestimmung mit Ermittlung der PSA-Anstiegsgeschwindigkeit (PSA im Zeitverlauf; PSA-velocity PSA-V).
- (4) Die Bildung des Quotienten aus freiem und Gesamt-PSA (Q-PSA).

Die Aussagekraft der ersten zwei Methoden ist wegen ihrer geringen Reproduzierbarkeit umstritten. Als aussichtsreichste Methode zur Erhöhung der diagnostischen Aussagekraft der PSA-Bestimmung erscheint die Bildung des Quotienten aus freiem und Gesamt-PSA, doch sind hinsichtlich der Bestimmung von freiem PSA noch einige Fragen offen.¹⁵⁷

Grundlage aller Verfahren ist die PSA-Bestimmung im Serum. Voraussetzung für den sicheren Umgang mit diesen Diagnosemethoden ist demnach die genaue Kenntnis des benutzten PSA-Messverfahren und der zugehörigen Grenzwerte. Ein

Großteil der Studien zur Früherkennung und Diagnose von Prostatakarzinomen wurden in Nordamerika durchgeführt, wobei vornehmlich die folgenden vier PSA-Assays zum Einsatz kamen.^{4,26,129,149} die monoklonalen Tandem-R und Tandem-E PSA-Assays (Assay-Typ: IRMA bzw. ELISA) von Hybritech, der monoklonale IMx PSA-Assay (Assay-Typ: ELISA) von Abbott sowie der polyklonale Pros-Check PSA-Assay (Assay-Typ: RIA) von Yang Laboratories. Die drei monoklonalen Assays scheinen gut zu korrelieren. Der polyklonale Assay zeichnet sich jedoch durch 1.6fach höhere Werte aus.¹³⁰ In den meisten Studien wurden mittels eines Abbott- oder Hybritech-Assay bestimmte PSA-Spiegel bis zu einem Grenzwert von 4ng/ml (ein- oder ausschließlich) als unverdächtig angesehen. Der entsprechende Grenzwert beim Yang-Assay wurde auf 7ng/ml angesetzt.^{4,129,149} Von einigen Untersuchern wurden zur Verbesserung der Sensitivität eine niedrigere Entscheidungsgrenze von 3ng/ml für die monoklonalen Assays gewählt.^{4,129,149} Insbesondere im Bereich von 4ng/ml bis 10ng/ml scheint jedoch eine große Überschneidung der PSA-Konzentrationen zwischen den Gruppen der Patienten mit BPH und lokalisiertem Prostatakarzinom zu bestehen. Der Bestimmung der PSA-D wurde für den Abbott- bzw. Hybritech-Assay in der Regel eine Entscheidungsgrenze von 0.12ng/ml/ccm oder 0.15ng/ml/ccm zugrunde gelegt und ein Anstieg von mehr als 0.75ng/ml/Jahr als verdächtiger PSA-V-Wert angesehen.¹⁴⁸

Eine einheitliche Kalibrierung der Assays gibt es nicht, eine internationale Referenzpräparation ist nicht verfügbar.^{26,130} Dies erschwert die Vergleichbarkeit verschiedener PSA-Assays. Durch die Vielzahl der verfügbaren Bestimmungsverfahren ergeben sich für den Kliniker Probleme bei der Interpretation von PSA-Werten. In Deutschland werden über 50 Verfahren zur PSA-Bestimmung im Serum angeboten.^{130,157} Die meisten der neu auf dem Markt erschienenen Verfahren lehnen sich dem etablierten Referenzbereich (<4ng/ml bzw. ≤4ng/ml) an, ohne dass dieser für alle Assays verifiziert wurde.¹³⁰ Einige Hersteller verzichten gänzlich auf die Angabe eines Referenzbereiches, andere ermittelten ihn anhand sehr kleiner Probandengruppen. Referenzbereiche, die bei jungen Männern oder gar mit einem nicht genannten Anteil an Frauen festgelegt wurden, sind nicht auf 50- bis 80-jährigen Männern übertragbar.¹³⁰ Einige Hersteller empfehlen, ihr Verfahren nicht zur Diagnose, sondern zur Verlaufskontrolle bei bekanntem Prostatakarzinom zu nutzen, eine Warnung, die dem einsendenden Mediziner meist ebenso verborgen bleibt, wie der Name des benutzten Verfahrens. Da der PSA-Wert in Abhängigkeit vom Bestimmungsverfahren in einer Serumprobe um einen Faktor von mehr als zwei differieren kann, ist es für den beurteilenden Arzt dringend erforderlich, das

Verfahren und den zugehörigen Referenzbereich zu kennen. Ohne diese Information verliert die PSA-Bestimmung ihre diagnostische Aussagekraft.¹³⁰

C.2.5 Beschreibung der Intervention "PSA-Screening beim Prostatakarzinom"

C.2.5.1 PSA-Bestimmung als Komponente verschiedener Screening-Technologien

Verbreitete Verfahren zur Früherkennung von Prostatakarzinom sind die digitale rektale Untersuchung (DRU) und die Bestimmung des PSA im Serum. Je nach Befund folgen die diagnostischen Verfahren transrektale Ultraschalluntersuchung (TRUS) und/oder transrektale Biopsie. Allerdings kann die Bestimmung von PSA als Screeningmaßnahme beim Prostatakarzinom nicht isoliert betrachtet werden, da sie mit den obengenannten Verfahren in Kombination eingesetzt wird. Erst die Festlegung der parallelen oder sequentiellen Teststrategie unter Einschluss einer oder mehrerer dieser Verfahren spezifiziert die medizinische Technologie. Somit ist die PSA-Bestimmung lediglich als Komponente unterschiedlicher Screening-Technologien zu verstehen.

Diese verschiedenen Strategien können im Rahmen des Health Technology Assessment einander gegenübergestellt und bezüglich des medizinischen Nutzens, der medizinischen Risiken und der Kosten verglichen werden. Hierbei ist auch die Strategie "Kein Screening" als mögliche Alternative zu berücksichtigen.

Die einzelnen Verfahren werden im folgenden kurz beschrieben und es werden Aspekte aufgezeigt, welche bei der Beurteilung der Effektivität des PSA-Screenings zu beachten sind.

PSA-Bestimmung

Die reine Bestimmung von PSA im Serum ohne Berücksichtigung der Folgerisiken und Folgekosten durch nachgeschaltete abklärende Diagnostik und Therapie ist einfach durchzuführen, kostengünstig, risikoarm und wird von den Patienten gut akzeptiert.

Zur Erhöhung der diagnostischen Wertigkeit der PSA-Bestimmung wurden zwei grundsätzlich verschiedene Ansätze diskutiert und untersucht:

- (1) die oben beschriebenen PSA-bezogenen Methoden wie die Bestimmung der PSA-Dichte, die Verwendung altersabhängiger PSA-Grenzwerte, die Verlaufsbestimmung mit Ermittlung der PSA-Anstiegsgeschwindigkeit und die Bildung des Quotienten aus freiem und Gesamt-PSA.
- (2) Die Kombination mit DRU.

Diese Kombination ist derzeit die von Urologen in Nordamerika wie auch in Europa bevorzugte Strategie zur Früherkennung und Diagnose von Prostatakarzinom. Die Kombination der PSA-Bestimmung mit der DRU erhöht zu einem bestimmten Grad die Sensitivität des Screenings, da Karzinome identifiziert werden, welche bei alleiniger Anwendung jeweils einer der beiden Methoden der Entdeckung entgangen wären.

Digitale rektale Untersuchung

Die DRU ist die herkömmliche diagnostische Methode beim Prostatakarzinom. Viele Tumoren entwickeln sich in den dorsal und lateral gelegenen peripheren Drüsenanteilen, welche der rektalen Palpation theoretisch zugänglich sind. Charakteristisch ist im frühen Stadium eine umschriebene Verhärtung innerhalb der Prostata (Konsistenzveränderung mit Bildung eines harten Karzinomknotens), die jedoch nicht obligat ist. Auch eine vom Palpationsbefund her völlig unverdächtige Prostata kann karzinomatös verändert sein.¹⁵² In fortgeschrittenen Stadien ist die Prostata insgesamt derb infiltriert, später ist die Grenze zur Umgebung nicht mehr sicher feststellbar. Nicht alle Verhärtungen der Prostata entsprechen einem Karzinom, auch die BPH, die granulomatöse Prostatitis, postoperative Veränderungen und Infarkte sowie entzündliche Residuen und Prostatasteine können Verhärtungen der Prostata hervorrufen.^{149,152}

Die Sensitivität der DRU ist beschränkt, da ventral und medial lokalisierte Tumoren palpatorisch kaum zu erreichen sind. Die Wahrscheinlichkeit der Entdeckung von klinisch nicht relevanten Karzinomen ist bei der DRU relativ gering, da insbesondere großvolumige Tumoren tastbar sind (diese Wahrscheinlichkeit erhöht sich allerdings, wenn bei einem suspekten DRU-Befund zusätzlich zur Punktion der verdächtigen Stelle eine systematische Biopsie durchgeführt wird). Allerdings bedeutet dies auch, dass sich ein relativ hoher Anteil der bei der DRU identifizierten Tumoren bereits über die Prostata kapsel ausgedehnt hat, womit die Möglichkeit einer kurativen

Therapie begrenzt ist. Hinsichtlich der Durchführung und Interpretation der DRU wird auf die hohe Interobserver-Variabilität hingewiesen.¹⁴⁹

Transrektale Ultraschalluntersuchung

Die Prostata kann heutzutage mittels einer rektal eingeführten Schallsonde bildlich dargestellt werden. Dabei bietet dieses transrektale Verfahren eine bessere Sichtbedingung auf die Prostata als die abdominale Sonographie.¹¹⁶ Die TRUS erlaubt die Darstellung intraprostatischer, nicht palpabler Bezirke, den Nachweis einer Infiltration der Samenblasen und die Schätzung des Prostatavolumens^{26,139}, wobei sich in Abhängigkeit des Berechnungsverfahrens Variationen ergeben können.¹⁵⁷ In der Regel werden echoarme (hypodense) Areale und/oder Asymmetrien der Drüse bzw. Unregelmäßigkeiten des glandulären Aufbaus als verdächtig angesehen.^{26,139} Allerdings werden damit nicht nur Karzinome, sondern auch benigne Vergrößerungen oder Entzündungen erfasst.¹³⁹ Aufgrund der geringen Eindringtiefe werden Karzinome in der peripheren Drüse besser erkannt als solche im Bereich der Transitionalzone oder am Apex. Außerdem scheinen insbesondere Tumoren der Transitionalzone ein heterogenes Echomuster aufzuweisen.¹⁴⁹ US ist ein relativ zeitaufwendiges, invasives und teures Verfahren, welches besonderer Qualifikation bedarf. Hier kann hinsichtlich der Durchführung und Interpretation, wie für die DRU, das Problem der Interobserver-Variabilität angeführt werden. Teils wurde die TRUS zur diagnostischen Abklärung eines suspekten DRU-Befundes und/oder PSA-Wertes eingesetzt, insbesondere aber zur Durchführung von gezielten wie auch systematischen Biopsien, allerdings hat sich inzwischen international die Meinung durchgesetzt, dass sich TRUS zur diagnostischen Abklärung eines pathologischen DRU- und/oder PSA-Befundes nicht eignet.^{4,26,129,139,149}

Transrektale Biopsie

Zur Sicherung der Diagnose ist die Biopsie unverzichtbar. Sie kann entweder in Form einer transrektalen bzw. perinealen Stanzbiopsie oder in Form der transrektalen Feinnadel-Aspirationsbiopsie erfolgen.^{26,59,129,139,148,158} Die durch transrektale Feinnadelaspiration unter Ultraschallkontrolle und/oder digitaler Führung gewonnene Zytologie wird vor allem in den skandinavischen Ländern als gleichwertige Alternative zur transrektalen Stanzbiopsie angesehen, die

Entdeckungswahrscheinlichkeit für ein Karzinom scheint in erfahrenen Händen für beide Methoden gleich hoch zu sein.^{152,158} Für Therapieentscheidungen wird allerdings meistens eine histologische Sicherung des Biopsiebefundes gefordert, für welche die Zytologie nicht herangezogen werden kann.

Methode der Wahl ist heute die transrektale Stanzbiopsie (core biopsy). Sie wird normalerweise unter Ultraschallkontrolle mit einer 18 Gauge-Nadel auf einer automatischen Biopsiepistole durchgeführt.^{26,59} Um die Prostata möglichst repräsentativ zu untersuchen und das Volumen eines möglichen Tumors abschätzen zu können, hat sich dabei in den letzten Jahren die systematische oder randomisierte Biopsie durchgesetzt. Dabei werden aus beiden Prostatalappen jeweils drei (Basis, Mitte und Apex: »Sextanten«-Biopsie), manchmal auch nur zwei (»Quadranten«-Biopsie) Stanzzyylinder entnommen, was einem mapping der Prostata gleichkommt.^{26,129,148,158} Zusätzlich werden palpatorisch oder sonographisch verdächtige Areale gesondert biopsiert.^{129,149,158} Die Biopsie kann ambulant ohne Anästhesie durchgeführt werden. Allerdings besteht ein beträchtliches Infektionsrisiko, so dass eine Antibiotikaprophylaxe angezeigt ist. Je nach Studie werden allerdings trotz Antibiotikaprophylaxe Komplikationsraten (z.B. Harnwegsinfektion, Urosepsis) von 0,7% bis zu 13,5% berichtet, die auch ein gewisses Mortalitätsrisiko mit sich bringen.¹²⁹ Auch die Biopsie ist, wie die TRUS, ein ressourcenaufwendiges und invasives Verfahren und wird deshalb in der Regel lediglich zur diagnostischen Abklärung eingesetzt. In Studien zur Validität diagnostischer Methoden wird sie zumeist als Referenzverfahren bzw. Goldstandard herangezogen.^{139,149}

Diagnostische Validität der Früherkennungsverfahren

Aufgrund ethischer, logistischer wie auch ökonomischer Gesichtspunkte scheiden die TRUS und die Biopsie als primäre Screening Tests aus.^{129,139} Die meisten Studien konzentrieren sich demnach auch auf die DRU und die Bestimmung von PSA als eigentliche Screening-Verfahren und verwenden TRUS und Biopsie als abklärende Diagnoseverfahren.¹⁴⁹

Die einzelnen Studien differieren hinsichtlich der Kombinationen der verwendeten Screening-Tests, hinsichtlich der Strategien der diagnostischen Abklärung sowie hinsichtlich der Entscheidungskriterien bei den einzelnen Methoden.^{129,149} Die wirkliche Validität der verschiedenen Teststrategien konnte in der Regel nicht genau

bestimmt werden, da der Einsatz des Goldstandards aus ethischen Gründen auf screening-positive Studienteilnehmer beschränkt werden musste.^{4,51,129,149} Daher beziehen sich die Angaben zu Gütekriterien entweder auf die selektierte Untergruppe der Probanden mit Biopsie, oder aber auf die gesamte Studienpopulation unter der (zumeist impliziten) Annahme, dass alle Karzinome in der betreffenden Population mittels Screening-Test und abklärender Biopsie entdeckt wurden und Probanden mit negativem Screening-Testergebnis als nicht erkrankt einzustufen sind. Im ersteren Fall führt der work-up bias zu einer Über- oder Unterschätzung der Sensitivität und Spezifität des Screening-Verfahrens. Im letzterem Fall wird die Sensitivität überschätzt. Zusätzlich fehlen oft Angaben zum Tumorstadium. Dies lässt Zahlenwerte zur diagnostischen Güte der eingesetzten Screening-Verfahren praktisch wertlos erscheinen, da die Sensitivität und Spezifität der Tests für potentiell kurierbare Tumorstadien im Mittelpunkt des Interesses steht.^{52,154} Bei Ermittlung der Gütekriterien in einem urologischen Versorgungsrahmen ist mit Überschätzung der Sensitivität durch selection bias zu rechnen, da Urologen zumeist erst bei manifesten Beschwerden konsultiert werden.^{129,149} Das Design und die Analyse von Studien zur Ermittlung des diagnostischen Werts von Früherkennungsmaßnahmen wird zudem dadurch erschwert, dass kleinvolumige, hochdifferenzierte Karzinome als "Nichterkrankung" betrachtet werden können, und dass die Entdeckung fortgeschrittener Karzinomen keinen therapeutisch Nutzen bringt.^{139,149}

Angaben zum PPV der Screening-Tests oder -Testkombinationen sollten aufgrund eines besseren work-up zuverlässiger sein, hängen aber stark von der Tumorprävalenz in der jeweiligen Studienpopulation und damit auch von den Rahmenbedingungen der Studie ab.¹⁵⁴ In diesem Zusammenhang seien der Einfluss der Länge des Intervalls zwischen zwei Früherkennungsuntersuchungen, der Versorgungsstrukturen, der Entscheidungskriterien der einzelnen Screening-Tests und des Alters der Studienteilnehmer auf den PPV genannt.¹³⁹ Auch die Intensität der diagnostischen Abklärung bei positivem Screening-Testergebnis ist ein den PPV beeinflussender Faktor. So kann durch die Durchführung systematischer Biopsien die Karzinomentdeckungsrate auf einen verdächtigen Screening-Test hin erhöht und ein relativ hoher PPV erzielt werden.⁹⁶ Allerdings kann die Karzinomdiagnose dabei auf Zufall beruhen (serendipity), wenn der Test nicht wegen der Tumoreigenschaften, sondern aufgrund anderer Anomalien positiv ausgefallen ist und damit eigentlich als falschpositiv betrachtet werden müsste.⁹⁶ Der PPV liegt in diesem Falle fälschlicherweise hoch. Studien legen nahe, dass dieses Phänomen auch bei der Früherkennung von Prostatakarzinom zu beachten ist.⁹⁶ Für sich

genommen, ist der PPV daher ein ungenügendes Maß für die diagnostische Validität.

Eindeutig geht aus den derzeit vorliegenden Daten zur diagnostischen Validität der Methoden zur Früherkennung des Prostatakarzinoms lediglich die hohe Falschpositiv-Rate aller Tests oder Teststrategien hervor.^{4,32,51,129,139,149}

Prätherapeutisches Staging bei Patienten mit nachgewiesenem Prostatakarzinom

Für die Primärdiagnostik (T-Kategorie) werden bei positiver Biopsie in der Regel zusätzlich eine abdominale Ultraschalluntersuchung und/oder ein Ausscheidungsurogramm (AUG) durchgeführt. Die abdominale Ultraschalluntersuchung von Nieren und Blase schließt die Harnstauung aus und bestimmt den Restharn, während mittels AUG Erkrankungen der oberen Harnwege und Auswirkungen des Tumors auf den Harntrakt erkannt werden können.^{125,152}

Die folgenden Diagnoseverfahren dienen der Ausbreitungsdiagnostik (N- oder M-Kategorie) bei nachgewiesenem Prostatakarzinom. Die pelvine Lymphadenektomie kann vor einer Operation oder Radiotherapie zur Beurteilung des regionären Lymphknotenstatus herangezogen werden. Offene und laparoskopische Operationstechniken scheinen hinsichtlich der diagnostischen Sicherheit gleichwertig zu sein.¹⁵⁸ Vor einer geplanten Radikaloperation wird in gleicher Narkose oft eine Staging-Lymphadenektomie mit intraoperativem Schnellschnitt durchgeführt. Der therapeutische Wert der radikalen Operation bei positiven Lymphknoten ist fraglich.^{152,158}

Computertomographie (CT) und Kernspinresonanztomographie (NMR oder MRT) werden öfters zur Erhebung des Lymphknotenstatus herangezogen. Über ihren diesbezüglichen Wert beim Prostatakarzinom gibt es in der Literatur keinen Beleg. Im präoperativen Staging sind sie entbehrlich.^{152,156,158}

Eine Röntgenaufnahme der Lunge (Röntgen-Thorax) gibt Auskunft über das Vorliegen von Lungenmetastasen und gehört zur Routinediagnostik.^{148,152,158} Früher als durch Röntgenaufnahmen lassen sich Knochenmetastasen durch die Skelettszintigraphie feststellen. Die Untersuchung ist zwar sehr sensitiv, aber keineswegs metastasenspezifisch.¹⁵⁶ Falsch positive Befunde können z.B. durch sonstige entzündlich-degenerative Veränderungen und durch Traumen vorgetäuscht

werden. Zur Differenzierung muss gegebenenfalls eine zusätzliche Röntgendiagnostik erfolgen.¹⁵²

C.2.5.2 Effektivität des Prostatakarzinom-Screenings

Für keines der Screening-Verfahren, konnte bislang eine Senkung der krankheitsspezifischen Mortalität zweifelsfrei nachgewiesen werden. Das Hauptproblem der bislang durchgeführten Studien liegt im Fehlen nichtgescreenter Kontrollgruppen oder zu kurzer Nachverfolgung der Studienteilnehmer.

Inzwischen wurden sowohl in Nordamerika (Prostate cancer screening in the prostate, lung, colorectal and ovarian cancer screening trial of the National Cancer Institute (PLCO Trial)⁴⁷ wie auch in Europa (European Randomized Study of Screening for Prostate Cancer - ERSPC, Studienleitung: Schröder, Rotterdam)¹³⁵ große randomisierte kontrollierte Studien zur Früherkennung wie auch zur Behandlung des Prostatakarzinoms in die Wege geleitet. Allerdings werden die ersten Ergebnisse erst in einigen Jahren vorliegen.^{4,26,32,51,129,139,149} Selbst dann ist allerdings mit erheblichem Bias in diesen Studien zu rechnen, so z.B. durch die Kontamination der Kontrollgruppen durch Screening-Teilnahme. Was zum jetzigen Zeitpunkt vorliegt, sind Daten aus unkontrollierten Screening-Studien, die zumindest für Nordamerika eine in den letzten Jahren erfolgte Verschiebung der Stadienverteilung zugunsten weniger fortgeschrittener Stadien nahelegen.¹⁴⁹ Die Interpretation dieser Daten hinsichtlich einer Senkung der langfristigen Mortalität wird durch die Möglichkeit von lead time bias, length bias und stage migration jedoch erschwert.

Dem ungewissen Nutzen des Screenings steht zudem das sog. Eisbergphänomen⁶⁷ des Screenings gegenüber, das sich aus dem unsicheren Progressionsverhalten der Prostatatumoren ergibt. Obwohl Prostatakarzinome mit zunehmendem Alter zwar stark zunehmen, bleiben sie jedoch meist klinisch stumm und beeinträchtigen die Lebenserwartung nicht. Daraus folgt, dass Männer eher mit als an Prostatakarzinom sterben.^{51,129} Da jedoch durch den Einsatz neuerer diagnostischer Verfahren und Vorgehensweisen ein Teil der klinisch unbedeutsamen Karzinome entdeckt wird, ergeben sich Patientengruppen, die unnötigerweise mit der Diagnose Krebs konfrontiert und versorgt werden müssen.⁶⁷ Da derzeit eine Unterscheidung zwischen klinisch relevanten und latent bleibenden Prostatakarzinomen nicht möglich ist, ergibt sich das Dilemma der Unter- oder Überversorgung. Die ethische

Problematik liegt darin, dass die aggressive Behandlung von Prostatakarzinom ihrerseits mit erheblicher Morbidität und auch Mortalität verbunden ist. Die Beunruhigung und Angst, welche bei einem positiven Früherkennungstest aufkommen, sind weitere negative Folgen eines Screenings. Bei Männern mit falschpositivem Testergebnis wird dabei die Abklärungsdiagnostik einschließlich Biopsie unnötigerweise durchgeführt. Oft führt ein verdächtiges Screening-Testergebnis mit negativem Biopsiebefund zu einer Intensivierung der Überwachung bzw. Überprüfung mit zusätzlichen Tests und Biopsien.

C.2.5.3 Empfehlungen zum Prostatakarzinom-Screening und gegenwärtige Praxis

Nationale und internationale Empfehlungen

Die nachfolgenden Zitate verschiedener internationaler und nationaler Einrichtungen lassen die unterschiedlichen Gesichtspunkte und Haltungen zur Früherkennung beim Prostatakarzinom erkennen.

International Union Against Cancer (1991)⁶⁶

"Not only is an acceptable and valid screening test yet to be developed, but an acceptable and effective treatment for the preclinical lesions found as a result of screening has to be developed. There is no question that even when these problems are solved, it will be necessary to establish the effectiveness of screening programs for prostatic cancer by well-designed, randomised trials before a recommendation on public health policy could be developed. We must be at least a decade away from such a recommendation."

National Cancer Institute (1993)¹⁰³

"Given the possibility of unnecessary morbidity associated with diagnosis and treatment of many such lesions, careful evaluation of prostate cancer screening is desirable. There is insufficient evidence to recommend transrectal ultrasound and serum tumor markers for routine screening in asymptomatic men."

Canadian Task Force on the Periodic Health Examination (1991)²³

"...fair evidence to exclude PSA from the periodic health examination of asymptomatic men over 40 years of age."

United States Preventive Services Task Force (1989)¹⁵⁰

"There is insufficient evidence to recommend for or against routine digital rectal examinations (DRE) as an effective screening test for prostate cancer in asymptomatic men. Transrectal ultrasound and serum tumor markers are not recommended for routine screening in asymptomatic men."

Australian Cancer Society (1995)³

"There is insufficient evidence that men's health will be advanced through prostate screening of asymptomatic men. Prostate cancer is common and does, to a lesser extent, cause death. No prospective data are available to demonstrate quantifiable and reproducible reductions in death from prostate cancer due to screening of asymptomatic men by PSA or DRE or both, at any interval. In contrast, detection of asymptomatic prostate cancer engenders of itself deleterious iatrogenic sequelae without conclusive net gain in absolute life span or quality of life."

Demgegenüber empfiehlt die American Cancer Society (1993)⁹⁹ eine jährliche Früherkennungsuntersuchung mit DRU und PSA-Bestimmung ab einem Alter von 50 Jahren (bei Familienanamnese von Prostatakarzinom und Afroamerikanern bereits ab dem 40. Lebensjahr) und stellt fest, dass

"Generally, men with a life expectancy of at least 10 years after detection may benefit from examination.... It is important to emphasize that the guidelines are not recommendation for public programs of mass screening. They are intended to help individual physicians and patients select the best early detection protocol for their personal needs. The purpose is to promote discovery of otherwise unsuspected early cancer in asymptomatic persons on an individual basis."

1997 wurde diese Empfehlung mit folgendem Zusatz revidiert:¹⁵³

"Information should be provided to men regarding the potential risks and benefits of screening."

Die American Urological Association wie auch das American College of Radiology schließen sich der Empfehlung zur jährlichen Früherkennungsuntersuchung mit DRU und PSA-Bestimmung bei Männern im Alter von 50 bis 75 Jahren an.¹⁴⁹

Die Canadian Urological Association (1992)²⁴ empfiehlt die jährliche Früherkennungsuntersuchung mit DRU und PSA-Bestimmung bei Männern im Alter von 50 bis 70 Jahren:

"...(the association) strongly believes that to provide adequate patient care, this test (PSA) must be universally available and that funding should be provided to cover the costs."

Das Tumorzentrum München (1997)¹⁴⁸ legt in Anlehnung an die Empfehlungen der American Cancer Society, der American Urological Association und des American College of Radiology dar, dass

"Zusätzlich zu der in Deutschland anerkannten jährlichen rektalen Untersuchung empfehlen wir - bis zum Vorliegen der randomisierten Studien - die jährliche Durchführung einer PSA-Bestimmung im Serum. Sollte der altersspezifische Grenzwert überschritten sein, ist zunächst eine transrektale Sonographie mit Bestimmung der PSA-D durchzuführen. Bei pathologischem Ergebnis oder bei suspektem Sonographiebefund ist eine Prostatabiopsie indiziert. Diese sollte stets als Randombiopsie (drei Stanzzyylinder aus jeder Seite) unter antibiotischer Prophylaxe durchgeführt werden. Auch eine transrektale Feinnadelbiopsie kommt in Betracht, wenn ein geübter Zytologe zur Verfügung steht. Bei nur gering erhöhtem PSA-Wert (etwa <200% des Grenzwertes) ist auch der Quotient von freiem und gesamtem PSA (Q-PSA) wertvoll für die Indikationsstellung zu einer Biopsie."

Die im Auftrag der Deutschen Krebsgesellschaft, der Deutschen Krebshilfe und der Deutschen Gesellschaft für Urologie herausgegebenen diagnostischen und therapeutischen Standards in der Urologischen Onkologie (1998)¹⁵⁸ legen folgendes nahe:

"Screening-Untersuchungen sind nur in der Altersgruppe der 50- bis 70jährigen zu erwägen, da die Lebenserwartung mindestens zehn Jahre betragen sollte, damit der Patient von einer lokalen Therapie einen Nutzen hat. Der Einfluß eines Screenings auf die Mortalität ist bis heute nicht bewiesen. Eine jährliche Untersuchung ist sinnvoll. Bei familiärer Prostatakarzinom-Belastung sollte das Screening früher beginnen (ab 45 Jahren) und zweimal jährlich durchgeführt werden."

Gegenwärtige Praxis

Aufgrund von Bedenken hinsichtlich einer Abnahme des Quotienten aus potentielltem Nutzen und Risiko ist von unterschiedlicher Seite zur Vorsicht beim Screening und bei der Behandlung von älteren Männern mit Prostatakarzinom geraten worden.¹⁴⁹ Eine verbreitete Empfehlung geht dahin, die Früherkennung und aggressive Behandlung auf Männer mit einer Lebenserwartung von mehr als zehn Jahren zu begrenzen. In den Vereinigten Staaten käme dies einer oberen Altersgrenze von etwa 73 Jahren gleich.¹⁴⁹ Trotzdem war zwischen 1984 und 1990 bei Männern im Alter von 75 bis 79 Jahren ein Anstieg der Raten für die radikale Prostatektomie um mehr als das Vierfache festzustellen.¹⁴⁹ Die altersstandardisierte Mortalität beim Prostatakarzinom ist in den Vereinigten Staaten zwischen 1960 und 1988 von ungefähr 21 pro 100 000 auf 25 pro 100 000 angestiegen.¹⁴⁹ Gleichzeitig wurde ein sehr viel bedeutender Anstieg der Inzidenz von Prostatakarzinom verzeichnet, der auf die vermehrte Durchführung von TUR-P bei Verdacht auf BPH und auf eine Intensivierung der Früherkennungsbemühungen zurückzuführen sein dürfte.^{70,117} Auch in Deutschland zeichnet sich eine solche Entwicklung ab.² Die Trends spiegeln sich in einer tendenziellen Zunahme der Diagnose von Karzinomen in weniger fortgeschrittenen Stadien sowie in verbesserten stadiumsspezifischen 5-Jahres-Überlebensraten wider.¹⁴⁹ Allerdings können aus Daten der Stadienverteilung nicht ohne weiteres Folgerungen hinsichtlich der krankheitsspezifischen Mortalität gezogen werden.

Unterschiede bestehen bezüglich der Art und Weise, wie nach einer ersten negativen Biopsie bei Karzinomverdacht weiterverfahren wird. Einige Empfehlungen gehen dahin, systematische Biopsien (insbesondere bei PSA-Werten über 10ng/ml pro ccm) mindestens einmal zu wiederholen. So wird in den im Auftrag der Deutschen Krebsgesellschaft, der Deutschen Krebshilfe und der Deutschen Gesellschaft für Urologie herausgegebenen *Diagnostischen und therapeutischen Standards in der Urologischen Onkologie* (1998)¹⁵⁸ folgender diagnostische Algorithmus als sinnvoll vorgeschlagen:

"Liegt ein positiver rektaler Tastbefund vor, so erfolgt unabhängig vom PSA-Wert eine weitere Abklärung. Ergibt die TRUS keinen weiteren Aufschluß über den Palpationsbefund (Verkalkung?), wird biopsiert. Bei

unauffälligem Palpationsbefund und einem PSA von 4-10ng/ml (Hybritech-Assay) erfolgt eine Laborkontrolle (nach 4 bis 6 Wochen), bei Bestätigung des Wertes eine Biopsie. Bei einem PSA >10ng/ml wird biopsiert.... Ist bei Karzinomverdacht die erste Biopsie negativ, so rechtfertigt das eine zweite Biopsie.... Bei negativer DRU wird bei der zweiten Biopsie zusätzlich die Transitionalzone punktiert...."

Vor 1994 gab es in Deutschland für die Bestimmung des PSA keine Gebührensätze im EBM; die Verrechnung kann unter anderen Leistungspositionen erfolgt sein. Die Frequenzstatistiken für die PSA-Bestimmung liegen demnach erst seit 1994 vor; andere EBM-Positionen im Zusammenhang mit medizinischen Leistungen bei der Prostata - wie die Sonographie, die DRU oder die Biopsie - umfassen Komplexeleistungen, sind demnach nicht prostataspezifisch und ohne die Versorgungsdaten der Leistungserbringer ohne größeren Aussagewert. Für die EBM-Position 4208 (Prostataspezifisches Antigen) ist eine Zunahme der Leistungshäufigkeit von 922 600 im Jahre 1994 auf 1 632 600 im Jahre 1995 und 2 030 100 im Jahre 1996 zu verzeichnen (BMAE und EGO zusammen).⁷⁵ Aus der Frequenzstatistik ist allerdings nicht ersichtlich, für welche Indikation (de facto-Screening, Hilfsmittel bei der abklärenden Erstdiagnose, Verlaufskontrolle, Metastasendiagnostik) die PSA-Bestimmung eingesetzt wurde, so dass es schwierig ist, sie ohne Verbindung zu Morbiditätsdaten zu interpretieren. Die Volumenzunahme der PSA-Bestimmung legt jedoch nahe, dass der PSA-Assay nicht nur zu Diagnose- bzw. Überwachungszwecken eingesetzt wird.

C.3 Forschungsfragen

Im vorliegenden Bericht sollen auf der Basis der medizinischen, gesundheitsökonomischen und methodischen Literatur folgende Forschungsfragen beantwortet werden:

1. Welche Beiträge aus der Literatur liegen zur gesundheitsökonomischen Komponente des Themas "PSA-Screening beim Prostatakarzinom" vor?
2. Welche Kriterien sind an eine gesundheitsökonomische Studie anzulegen, um die notwendigen inhaltlichen und methodischen Voraussetzungen für einen validen Beitrag im Rahmen des Entscheidungsprozesses überprüfen zu können?
3. Welchen methodischen Problemen steht man beim Bewerten gesundheitsökonomischer Studien gegenüber und welche fachliche Expertise ist für diesen Prozess erforderlich?
4. Wie ist die Transparenz und inhaltlich-methodische Qualität der identifizierten gesundheitsökonomischen Beiträge zum PSA-Screening beim Prostatakarzinom zu bewerten?
5. Wie ist die gesundheitsökonomische Evidenz zum PSA-Screening beim Prostatakarzinom zu beurteilen? Neben der Frage nach der Effektivität der Technologie sind hierbei insbesondere die Kosten und die Relation zwischen Kosten und Effektivität zu berücksichtigen.
6. Welche Folgerungen bzw. Einschränkungen ergeben sich im Hinblick auf die Übertragbarkeit der vorhandenen Ergebnisse nach Deutschland.
7. In welchen Bereichen liegt der größte weitere Forschungsbedarf?

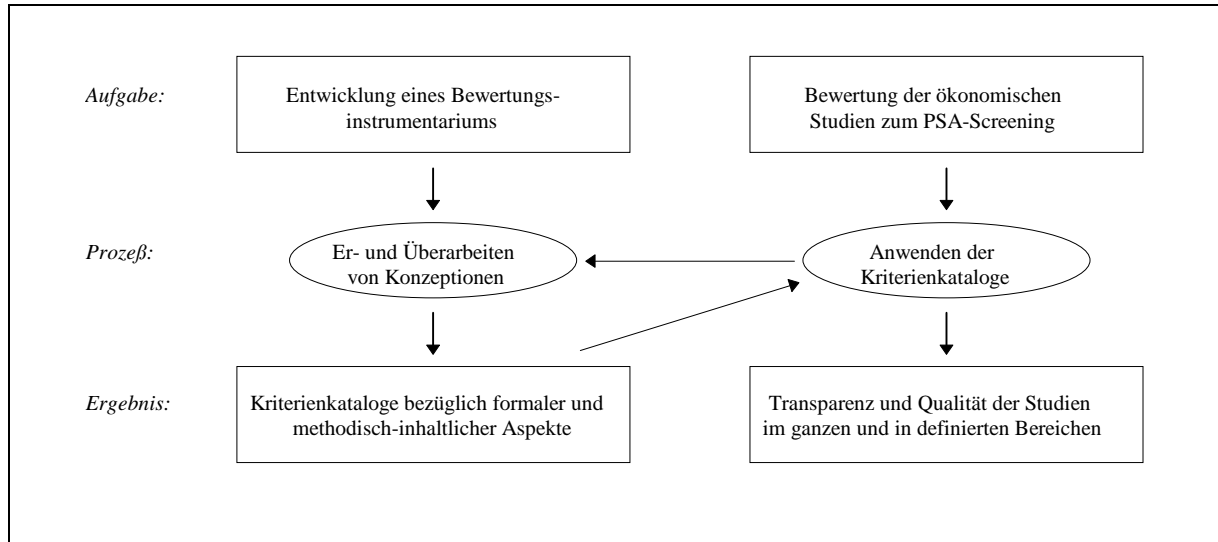
C.4 Methodik

Gemäß der Aufgabenstellung des BMG-Projekts gliedert sich die vorliegende Arbeit in zwei Teilbereiche:

- Ein Teil der Aufgabenstellung war die Entwicklung von Kriterienkatalogen zur standardisierenden Bewertung gesundheitsökonomischer Studien, welche im Rahmen der geplanten HTA-Datenbank einzusetzen sind.
- Der zweite Teil des Auftrags bestand darin, die entwickelten Kriterienkataloge auf Studien zum PSA-Screening anzuwenden und so zu einer methodisch abgesicherten gesundheitsökonomischen Technologiefolgenabschätzung zu gelangen.

Entwicklung und Anwendung des Beurteilungsinstrumentariums waren eng verzahnt, wobei die praktische Anwendung zur Überprüfung der Praktikabilität und Vollständigkeit der Kriterienkataloge genutzt wurde, was rückkoppelnd wiederum Einfluss auf deren Form und Inhalt hatte. Zusätzlich zur internen Qualitätssicherung fanden regelmäßige Konsensustreffen zwischen den ökonomischen Gruppen München, Hannover und Ulm statt, da die in dieser Arbeit entwickelten Bewertungsinstrumente in allen gesundheitsökonomischen Untersuchungen des BMG-Projektes angewendet werden sollten.

Abbildung 1: Verzahnung der Entwicklung des Bewertungsinstrumentariums mit dessen praktischer Anwendung bei der Bewertung der gesundheitsökonomischen Studien zum PSA-Screening beim Prostatakarzinom



C.4.1 Entwicklung des Bewertungsinstrumentariums für gesundheitsökonomische Studien

Die Entwicklung der Kriterienkataloge (Deskriptionskatalog, Transparenzkatalog und Qualitätskatalog) fand gemeinsam mit den beiden gesundheitsökonomischen Projektgruppen aus Hannover und Ulm statt. Dabei lag die Federführung bei der Münchner Projektgruppe. Insgesamt umfasste das Panel zur Entwicklung des Instrumentariums für die Bewertung gesundheitsökonomischer Studien folgende Autoren:

Tabelle 5: Panel zur Entwicklung eines Instrumentariums für die Bewertung von gesundheitsökonomischer Studien im Rahmen von Health Technology Assessment

München (Federführung)	Hannover	Ulm
Prof. Dr. J. Wasem ¹ U. Siebert, M.P.H. ¹ Dr. N. Mühlberger, M.P.H. ² Dipl.-Ökotroph. C. Behrend ¹	Prof. Dr. J.-M. Graf von der Schulenburg Dipl.-Ök. Dr. W. Greiner	Prof. Dr. R. Leidl R. Welte, M.P.H., MSc
¹ Postgradualer Studiengang Öffentliche Gesundheit und Epidemiologie Institut für Medizinische Informationsverarbeitung, Biometrie und Epidemiologie Ludwig-Maximilians-Universität München	Institut für Versicherungs- betriebslehre Universität Hannover	Abteilung Gesundheitsökonomie Universität Ulm
² GSF Forschungszentrum für Umwelt und Gesundheit Institut für Medizinische Informatik und Systemforschung, Neuherberg		

Da in der bisherigen Entwicklungsphase vom BMG noch keine endgültigen bzw. abschließenden Festlegungen hinsichtlich der Adressaten der Datenbank und der Höhe des für ihren Aufbau und Betrieb geplanten Ressourceneinsatzes gemacht werden sollen und können, wurde bei der Entwicklung des Beurteilungsinstrumentariums versucht, unterschiedlichen Erwartungen und Bedürfnissen gerecht zu werden. Dabei sollte die aus der Kombination medizinischer, ökonomischer, epidemiologischer, entscheidungsanalytischer und statistischer Fragestellungen erwachsende Komplexität gesundheitsökonomischer Evaluationsprozesse zum Ausdruck gebracht und die ziel- bzw. zweckabhängig unterschiedlichen fachlichen, personellen und finanziellen Erfordernisse für die

Durchführung von Studienbewertungen offengelegt werden. Zu diesem Zweck wurden drei Kriterienkataloge mit zunehmender Beurteilungstiefe (Deskription, Transparenz, methodisch-inhaltliche Qualität) entwickelt, deren Aneinanderreihung zugleich den stufenweisen Ablauf der vorgenommenen methodischen Evaluation verdeutlicht.

In Anlehnung an internationale Datenbanken zur Bewertung von medizinischen Technologien wurde von der Münchner Projektgruppe ein den Spezifika gesundheitsökonomischer Studien gerecht werdendes Raster (standardisierte Berichte) zur standardisierten Berichterstattung gesundheitsökonomischer Studien entwickelt und beispielhaft für das Thema "PSA-Screening beim Prostatakarzinom" eingesetzt. (Vgl. zu den Katalogen auch unten Kapitel C.5.1 und zum Raster der standardisierten Berichte Kapitel C.5.3)

C.4.1.1 Theoretische Grundlagen

Theoretische Grundlage für die Entwicklung des Bewertungsinstrumentariums bildeten hauptsächlich die Monographien "Standards for Socioeconomic Evaluation of Health Care Products von Luce & Elixhauser" (1990)⁹⁴, "Valuing Health Care. Cost, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies" von Sloan (1995)¹³² und "Cost-Effectiveness in Health and Medicine" von Gold, Siegel, Russel & Weinstein (1996).⁴⁸ Die inhaltliche Orientierung erfolgte außerdem an verschiedentlich aufgestellten Prinzipien und Empfehlungen bezüglich der Durchführung und Dokumentation gesundheitsökonomischer Untersuchungen. Als Quellen zu nennen sind hierbei neben den eingangs genannten Standardwerken: eine Checkliste für die Bewertung gesundheitsökonomischer Studien von Stoddart & Drummond (1984b)¹³⁸, der "Guide to the Economic Analysis of Clinical Practices" von Eisenberg (1989)⁴³, die "Guidelines for Economic Evaluation of Pharmaceuticals: Canada" des Canadian Coordinating Office for Health Technology Assessment (1994)²⁴, die "Guidelines for Authors and Peer Reviewers of Economic Submissions to the BMJ" des British Medical Journal Economic Evaluation Working Party (Drummond & Jefferson, 1996)⁴¹, die "Deutschen Empfehlungen zur gesundheitsökonomischen Evaluation - Hannoveraner Konsens" der Konsensgruppe "Gesundheitsökonomie" (1996)⁷⁷ sowie eine Kurzfassung des "Analysis Assessment Scheme" des Department of Health, Economics and Operational Research Division in Großbritannien (Mason & Drummond, 1995).⁹⁵ Außerdem wurde auf die Artikelserien "How To Read Clinical Journals" (1981-1984)^{61,123,124,137,138,145-147}

der McMaster University (Kanada) und "Users' Guides to the Medical Literature" (1993-1997) 38,54-57,62,63,71,72,85,88,104,105,109,111,112,119,120 der Evidence-Based Medicine Working Group, welche ebenfalls an der McMaster University angesiedelt ist, zurückgegriffen. Für deutsche Formulierungen erwies sich ein Beitrag von Leidl (1997)⁸⁷ und die Bücher "Gesundheitsökonomische Evaluationen" von Schöffski, Glaser & von der Schulenburg (1998)¹²⁶ und "Medizinische Ökonomie" von Szucs (1997)¹⁴⁰ als äusserst hilfreich.

C.4.1.2 Deskriptionskatalog

Dieser in der ersten Beurteilungsstufe eingesetzte, sehr umfangreiche Katalog (vgl. Tabelle 24 im Anhang 2: Deskriptionskatalog) dient dazu, die inhaltlichen und methodischen Aspekte gesundheitsökonomischer Studien in standardisierter Weise beschreibend zu erfassen und weniger geschulten Lesern die Komplexität gesundheitsökonomischer Evaluationsverfahren vor Augen zu führen.

Die inhaltliche und strukturelle Konzeption des Kriterienkatalogs basiert zum einen auf Empfehlungen, die bezüglich der Methodik von Durchführung und Dokumentation gesundheitsökonomischer Studien konsistent aus verschiedenen Literaturquellen hervorgehen, zum anderen auf Vorstellungen sowie Abgrenzungen und Definitionen hinsichtlich der verwendeten Terminologie, die in projektinternen und -externen Gesprächsrunden festgelegt wurden. Darüber hinaus veranlassten die Studien und HTA-Berichte zum PSA-Screening zu Überlegungen bezüglich weiterer, bei der Beurteilung zu beachtender Kriterien, die z.T. aufgegriffen und in den Kriterienkatalog integriert wurden.

C.4.1.3 Standardisierte Berichte

Um einen systematischen Zugriff auf die Information in den einzelnen Studien zu gewährleisten, ist die standardisierte Dokumentation und eine methodisch und inhaltlich sinnvolle Zusammenfassung der wesentlichen Inhalte aus den gesundheitsökonomischen Publikationen notwendig.

Die meisten Publikationen zu gesundheitsökonomischen Studien entsprechen in ihrem Aufbau nicht dem üblichen Schema der Publikationen von Studien zur medizinischen Effektivität. Zu den Spezifika gesundheitsökonomischer Publikationen gehört, dass sie nach einer kurzen Darstellung des Evaluationsrahmens meist

getrennt die gesundheitsbezogenen Parameter (Effektivität) und die ökonomischen Parameter (Kosten) beschreiben. Diese werden anschließend in einem weiteren Teil zueinander in Beziehung gesetzt bzw. synthetisiert. Insbesondere bei den entscheidungsanalytischen Modellierungsstudien spielt der Bereich Sensitivitätsanalysen zur Behandlung von Unsicherheiten eine wichtige Rolle, so dass hierzu oft ein eigener Abschnitt folgt.

Um diesen Spezifika der gesundheitsökonomischen Studien gerecht zu werden und im Sinne einer Datenbank systematisch den Inhalt dieser Publikationen zu erfassen und zu berichten, wurde ein Raster zur Erfassung der relevanten Information aus gesundheitsökonomischen Studien entwickelt (vgl. das Raster der standardisierten Berichte in Kapitel C.5.3).

Der Aufbau der standardisierten Berichte basiert auf den Informationsaufbereitungsstrukturen der Datenbanken DARE und NEED, welche in Anlehnung an die Themenbereiche der Kriterienkataloge zweckmäßig modifiziert und adaptiert wurden. Ziel dieses Arbeitsschrittes war eine übersichtliche, standardisierte Erfassung und Wiedergabe des Inhalts der einzelnen Studien in folgenden Blöcken:

- Evaluationsrahmen
- Studiendesign und Studien- bzw. Zielpopulation
- Bestimmung der gesundheitsbezogenen Parameter
- Bestimmung der ökonomischen Parameter
- Diskontierung
- Synthese von Kosten und Effekten
- Behandlung von Unsicherheiten
- Diskussion und Schlussfolgerungen der Autoren

Neben diesen für die rein objektiv-deskriptive Extraktion von Daten angelegten Bereichen wurde zusätzlich unter der Rubrik "Kommentare" Freiraum für den kritischen Vermerk von Auffälligkeiten bezüglich des Studiendesigns, der Studiendurchführung, -auswertung und -dokumentation sowie ggf. für bewertende Kommentare bezüglich der internen und externen Validität gelassen.

C.4.1.4 Transparenzkatalog

Im Gegensatz zum vorher genannten rein deskriptiven Katalog, erlaubt der für die zweite Beurteilungsstufe entwickelte Transparenzkatalog bereits eine bewertende Beurteilung gesundheitsökonomischer Studien (vgl. den Transparenzkatalog in Kapitel C.5.1.2). Bewertet wird dabei allerdings nicht die inhaltlich-methodische Qualität der Studien, was das höchste Niveau von Anforderungen an die Qualifikation der Beurteiler stellen würde, sondern lediglich deren formale Vollständigkeit in der Dokumentation der einzelnen Bereiche und der inhaltlichen und methodischen Nachvollziehbarkeit.

Auch dieser Katalog ist das Produkt eingehender Konsensusgespräche, bei denen die für die Beurteilung der Transparenz relevanten Items aus dem Kriterienpool des deskriptiven Katalogs ausgewählt und zu beurteilungsrelevanten Themenbereichen zusammengefasst wurden. Bei Anwendung des Katalogs kann jedem Einzelkriterium ein Punktwert zugewiesen werden, der sowohl isoliert, als auch als Teil eines themenbereichsspezifische Summenscores diskutiert werden kann. Mögliche Punktwertangaben sind:

1 = Kriterium erfüllt

1/2 = Kriterium teilweise erfüllt

0 = Kriterium nicht erfüllt

Ist ein Kriterium nicht relevant, d.h. ist die dazugehörige Frage für die gerade zu beurteilende Studie ohne Belang, wird kein Punktwert sondern die Kennzeichnung "nr" (nicht relevant) vergeben. Mit "nicht relevant" gekennzeichnete Antwortkategorien werden bei der Berechnung relativer Summenscores weder im Zähler noch im Nenner berücksichtigt.

Es sei darauf hingewiesen, dass die Gewichtungen der einzelnen Kriterien nicht auf biometrischen Modellen beruhen und damit die Skaleneigenschaften des Summenscores nicht bekannt sind. Dem Summscore liegt vielmehr ein heuristischer Ansatz zugrunde, welcher lediglich der Deskription einzelner Bereiche bzw. der gesamten Studie dient. Es bleibt den einzelnen evaluierenden Forschern überlassen, Summenscores anzuwenden oder nicht.

C.4.1.5 Qualitätskatalog

Zielsetzung der dritten Beurteilungsstufe war es, die methodisch-inhaltliche Qualität der gesundheitsökonomischen Studien in dem von ihnen vorgegebenen Kontext bzw. Rahmen aus Fragestellung, Evaluationsansatz und Studiendesign in standardisierter Weise zu bewerten. Die Entwicklung des hierzu verwendbaren Qualitätskatalogs erfolgte analog zur Entwicklung des vorher beschriebenen Transparenzkataloges (vgl. den Qualitätskatalog in Kapitel C.5.1.3). Die Themenbereiche und Einzelkriterien der beiden Kataloge zeigen starke Übereinstimmung. Im Gegensatz zum Transparenzkatalog verlangt der Qualitätskatalog jedoch vom Anwender ein Urteil über die Angemessenheit der Methodik. Dieser Ansatz erfordert daher ein Maximum an medizinischer, ökonomischer, epidemiologischer und entscheidungsanalytischer Fachkompetenz. Bezüglich der Eigenschaften der Summenscores gelten beim Qualitätskatalog dieselben Einschränkungen wie beim Transparenzkatalog.

C.4.2 Zielpopulation und Ergebnisparameter

Das vom BMG geförderte Vorhaben "Bestandsaufnahme, Bewertung und Vorbereitung der Implementation einer Datensammlung 'Evaluation medizinischer Verfahren und Technologien' in der Bundesrepublik Deutschland" richtete sich primär auf Fragen der medizinischen Effektivität.

Zielsetzung des vorliegenden Berichtes war die Ergänzung dieses Vorhabens um eine ökonomische Komponente. Konkret wurden folgende Teilzielsetzungen formuliert:

- Entwicklung eines Bewertungsinstrumentariums zur entscheidungsorientierten Dokumentation gesundheitsökonomischer Evaluationsstudien einschließlich eines Kriterienkataloges zur Bewertung gesundheitsökonomischer Studien.
- Exemplarische Anwendung dieses Bewertungsinstrumentariums am Themenbereich "Ökonomischen Evaluation des PSA-Screenings beim Prostatakarzinom".
- Ergänzende Überlegungen zu Kosten und Kosten-Effektivität der medizinischen Technologie "PSA-Screening beim Prostatakarzinom".

Die Projektvorgabe bestand darin, den Pool an zu evaluierenden Zieldokumenten und Ergebnisparametern aus den Forschungsfragen abzuleiten und die Literaturrecherche anhand dieser Dokumenten- und Parametertypen einzugrenzen (s. Kapitel 7, Band Bestandsaufnahme, Bewertung und Vorbereitung der Implementation einer Datensammlung "Evaluation medizinischer Verfahren und Technologien" in der Bundesrepublik¹⁴). Dieses Verfahren erschien insbesondere bei der Evaluation gesundheitsökonomischer Studien sinnvoll, da sich hier der Typ aussagekräftiger Dokumente und die relevanten Parameter grundsätzlich von denen der medizinischen Effektivitätsbeurteilung unterscheiden. Insbesondere ist zu vermerken, dass in Praxisrichtlinien und Konsensdokumenten bisher systematische Aussagen zur Ökonomie der Technologie fehlen und auch in Sekundärprodukten wie HTA-Berichten im Gegensatz zur medizinischen Evidenz die gesundheitsökonomische Komponente, wenn überhaupt, meist nur fragmentarisch und unsystematisch dargestellt wird. Um präzise und nachvollziehbare Aussagen machen zu können und extrahierbare Daten zu erhalten, ist grundsätzlich auf die Primärstudien und Reviews mit zusätzlichen Primäranalysen zurückzugreifen. Da die meisten Studien mit entscheidungsanalytischem Modellierungsansatz per Studiendesign auf Parameter zurückgreifen, die mehr oder weniger systematisch aus der Literatur extrahiert und synthetisiert werden, stellt eine gesundheitsökonomische Primärstudie mit einem entscheidungsanalytischen Modell in vielen Fällen das ökonomische Komplement zur Betrachtung der medizinischen Effektivität innerhalb der HTA-Berichte dar. Da in diesen Modellierungsstudien jedoch im Gegensatz zu den meisten HTA-Berichten eigene quantitative Analysen durchgeführt werden und teilweise Primärdaten zur Effektivität, zur Lebensqualität oder zu den Kosten einfließen, wurden die Modellierungsstudien im folgenden als Primärstudien bezeichnet.

Aus den genannten Gründen wurde die systematische Literaturrecherche auf Primärstudien und Reviews mit zusätzlichen Primäranalysen eingegrenzt. Dies war auch kompatibel mit dem Ziel, für diese Studien einen Kriterienkatalog zur qualitativen Beurteilung zu entwickeln.

Abgeleitet aus diesen Zielsetzungen wurden für die Evaluation alle Primärstudien und Reviews eingeschlossen, die eigene gesundheitsökonomische Evaluationen durchgeführt haben. Neben den klassischen gesundheitsökonomischen Evaluationsstudien zur simultanen Betrachtung von Kosten und Konsequenzen wurden auch reine Kostenbetrachtungen eingeschlossen (s. Tabelle 6). HTA-Berichte, Reviews und andere Sekundärprodukte wurden als Hintergrunddokumente bzw. technische Berichte eingeschlossen, falls dort detailliertere, über die

Primärstudien hinausgehende Kosten- oder Kosten-Effektivitäts-Betrachtungen angestellt wurden.

Die relevanten verfügbaren Studien wurden identifiziert und nach Studientyp kategorisiert. Es erfolgte eine systematische Darstellung der Inhalte in Form von standardisierten Berichten und eine Bewertung der Transparenz und der methodischen und inhaltlichen Studienqualität der identifizierten Studien.

Tabelle 6: In die Evaluation eingeschlossene gesundheitsökonomische Studientypen

Untersuchte Faktoren	Gesundheitsökonomischer Evaluationstyp
Kosten	Kostenbeschreibung (cost description, economic dossier)
	Krankheitskosten-Studie (cost-of-illness analysis)
	Vergleichende Kosten-Analyse (cost analysis)
Kosten und Konsequenzen	Kosten-Konsequenzen-Beschreibung (cost outcome description)
	Kosten-Konsequenzen-Analyse (cost-consequence analysis)
	Kosten-Minimierungs-Analyse (cost-minimisation analysis)
	Kosten-Effektivitäts-Analyse (cost-effectiveness analysis)
	Kosten-Nutzwert-Analyse (cost-utility analysis)
	Kosten-Nutzen-Analyse (cost-benefit analysis)

Damit ergaben sich als Ergebnisparameter für die Kosten und die Relation von Kosten und Konsequenzen folgende Maßzahlen:

- Kosten pro Screening-Teilnehmer
- Kosten pro identifiziertes Prostatakarzinom
- Kosten pro gerettetes Leben
- Kosten pro gewonnenes Lebensjahr
- Kosten pro gewonnenes qualitätsadjustiertes Lebensjahr (QALY)

Da in gesundheitsökonomischen Studien die Kosten bzw. die Kosten-Effektivität in Zusammenhang mit der Effektivität zu interpretieren sind, wurden für die berücksichtigten Studien auch die Effektivitätsparameter wie z.B. Entdeckungsrate an Prostatakarzinomfällen, gewonnene Leben, gewonnene Lebensjahre und gewonnene QALYs extrahiert und systematisch dargestellt. Wo möglich, wurden durchschnittliche und inkrementelle Maßzahlen extrahiert.

Zur detaillierten Beschreibung der Effektivität sei auf den Bericht von Pientka¹¹⁶ verwiesen.

Die Übertragbarkeit nach Deutschland wurde qualitativ in der Diskussion berücksichtigt. Eine methodische Bearbeitung dieses Themenbereiches fand übergreifend durch die gesundheitsökonomische Projektgruppe in Ulm statt, so dass hierauf verwiesen werden kann.¹⁵⁹

C.4.3 Datenquellen und Recherche

Zur Erschließung von gesundheitsökonomischem bzw. HTA-relevantem Datenmaterial zum PSA-Screening beim Prostatakarzinom wurden verschiedene Literaturdatenbanken, bereits veröffentlichte HTA-Berichte, einschlägige Fachzeitschriften, und Informationen verschiedener Institutionen genutzt.

C.4.3.1 Literaturdatenbanken und Fachzeitschriften

Die genutzten Datenbanken sind in Doku 2 im Anhang aufgeführt. Als Zeitrahmen für die Literaturrecherche wurden die Jahrgänge ab 1986 gewählt, da PSA-Tests erst in den späten 1980er Jahren weithin zur Verfügung standen. Die Recherche wurde bis einschließlich des Monats August 1998 durchgeführt.

Zur Maximierung des Suchergebnisses wurde eine sensitive Suchstrategie angewendet, was sich bei der zu erwarteten Seltenheit ökonomischer Studien zum PSA-Screening als praktikabel erwies. In diesem Sinne wurde auf die Verwendung der MeSH-Terminologie u.ä. oder eine Eingrenzung der Suche auf bestimmte Datenbankfelder bewusst verzichtet und eine reine Freitextsuche nach Schlagwörtern und deren Verknüpfung mit Boole-Operatoren durchgeführt. Für die Recherche in den von WebSPIRS und OVID angebotenen oben genannten Datenbanken wurde die im Anhang in Doku 2a aufgeführte Strategie eingesetzt, welche u.a. die Begriffe "cost" oder "costs" einschloss. Mit "cost" wurden auch die interessierenden Begriffe cost(-)effectiveness, cost(-)minimisation bzw. cost(-)minimization, cost(-)utility, cost(-)benefit, cost(-)of(-)illness, cost(-)consequence und cost(-)comparison erfasst.

Ferner erfolgte eine Durchsicht der Referenzlisten aller zusammengetragenen Publikationen einschließlich der HTA-Berichte zum Themengebiet und relevanter

Fachzeitschriften. Die verwendeten Fachzeitschriften sind ebenfalls in Doku 2 im Anhang aufgelistet.

C.4.3.2 Information einschlägiger Einrichtungen und Organisationen

Die Publikations- und Projektlisten (Stand bis 8/1998) von HTA vornehmenden bzw. unterstützenden Einrichtungen wurden im Internet nach relevanter (Hintergrund-)Information durchsucht. Die Internet-Seiten der folgenden Institutionen wurden hierfür aufgerufen.

Tabelle 7: Internet-Seiten der berücksichtigten HTA vornehmenden bzw. unterstützenden Institutionen

Agency for Health Care Policy and Research (AHCPR)	http://www.ahcpr.gov
Australian Health Technology Advisory Committee (AHTAC)	http://aesculapius.health.gov.au/hfs/nhmrc/nhmrc.htm
British Columbia Office of Health Technology Assessment (BCOHTA)	http://www.chspr.ubc.ca/bcohta
Canadian Coordinating Office for Health Technology Assessment (CCOHTA)	http://www.ccohta.ca
Danish Institute for Health Technology Assessment	http://www.dsi.dk
International Society of Technology in Health Care (ISTAHC)	http://www.istahc.org
McMaster University, Centre for Health Economics & Policy Analysis (CHEPA)	http://hiru.mcmaster.ca/ceb/programmes/chepa/default.html
Ministère de la Santé et des Services sociaux du Québec	http://www.msss.gouv.gc.ca
National Coordinating Centre for Technology Assessment (NCCHTA)	http://www.soton.ac.uk/%7ehta/hta-pubs.htm
Office of Technology Assessment, Congress of the United States (OTA)	http://www.wws.princeton.edu:80/~ota
The Swedish Council on Technology Assessment in Health Care (SBU)	http://www.sbu.se

Das Emergency Care Research Institute (ECRI) wurde per E-Mail (ecri@hslc.org) kontaktiert.

Parallel dazu wurden verschiedene Einrichtungen und Organisationen in Deutschland angeschrieben, um Informationen zur Situation in Deutschland zu folgenden Themen zu erhalten:

- Epidemiologie des Prostatakarzinoms (s. Doku 3 im Anhang)

- Nationale Leitlinien, Empfehlungen oder Stellungnahmen zur (PSA-gestützten) Früherkennung von Prostatakarzinom und zu therapeutischen Behandlungsoptionen (s. Doku 1 im Anhang)
- Häufigkeit von präventiven, diagnostischen und therapeutischen Interventionen beim Prostatakarzinom (s. Doku 3 im Anhang)

Folgende Experten haben durch ihre kritischen Reviews zu dem Bericht beigetragen:

- Prof. Dr. D. Hölzel
Leiter des Tumorregisters am Tumorzentrum München,
Institut für Medizinische Informationsverarbeitung, Biometrie und
Epidemiologie, Klinikum Großhadern, Ludwig-Maximilians-Universität München
- Dr. med. habil. M. Hegemann
Facharzt für Urologie, Urologische Praxis, München

C.4.3.3 Manuelle Selektion

Die mit Hilfe dieser Suchstrategien aufgefundenen Referenzen wurden anhand von Titel und Abstrakt manuell selektiert. Dabei galten folgende Einschlusskriterien:

- Primärstudien und Reviews, in welchen eigene Kostenberechnungen zur PSA-Bestimmung im Rahmen von Screening- bzw. Früherkennungsprogrammen aufgestellt wurden
- Der Volltext der Publikationen musste in englischer, deutscher oder französischer Sprache verfasst sein.

Ausgeschlossen wurden:

- Leitartikel, Briefe, "personal opinions".
- Konferenz-Abstracts, für die unter den angegebenen Autoren in einer zusätzlichen Recherche (MEDLINE EXPRESS, EMBASE) keine entsprechende Volltext-Publikation gefunden werden konnte.
- HTA-Berichte, sofern sie nicht als Hintergrundmaterial für Primärpublikationen mit eigenen Kostenberechnungen herangezogen wurden.

In den seltenen Fällen, in denen aus Titel oder Abstrakt nicht eindeutig hervorging, ob tatsächlich eine eigene Kostenberechnung durchgeführt worden war, wurde anhand des Volltextes der Studie über deren Berücksichtigung bzw. Nichtberücksichtigung entschieden. Studien, zu welchen mehrere Veröffentlichungen mit gleichem oder leicht abgeändertem Inhalt vorlagen, wurden unter dem Autor und Titel der erst erschienenen Veröffentlichung gruppiert und als Ganzes behandelt.

C.4.4 Bewertung der Information/Studienqualität

An die systematische Erfassung der Information in den standardisierten Berichten schloss sich die Bewertung der Transparenz und der inhaltlich-methodischen Qualität der Studien anhand der hierfür entwickelten Kriterienkataloge an. Für den Bewertungsprozess wurde ein dreistufiges Verfahren gewählt. Zunächst wurden alle Studien von jeweils zwei Projektmitarbeitern unabhängig beurteilt. Anschließend wurden die Ergebnisse verglichen und eventuelle Diskrepanzen im gemeinsamen Gespräch ausgeräumt. Konnte kein Konsens erzielt werden, wurde ein dritter Bewerter hinzugezogen.

Die auf diese Weise für alle Katalogkriterien erhaltenen Punktwerte boten die Basis zur Beurteilung der Studienqualität. Anhand einzelner Punktwerte konnte die Erfüllung einzelner Kriterien, anhand von Summenscores das Abschneiden in bewertungsrelevanten Teilbereichen sowie die Gesamtqualität beurteilt werden. In der am stärksten aggregierten Form der Ergebnisdarstellung wurden die Punktwerte nach Studie und Beurteilungskriterium kreuztabelliert, wodurch ein anschaulicher Überblick über die Qualitätsunterschiede der Studien vermittelt wird. Infolge dieser übersichtlichen Informationsaufbereitung war es zudem möglich, die Empfehlungen der Studien im Bezug auf das PSA-Screenings in Abhängigkeit von deren methodischer Qualität zu diskutieren.

C.4.5 Datenextraktion ökonomischer Parameter

Prinzipiell wurden, wo angegeben, aus allen berücksichtigten Studien die relevanten qualitativen Merkmale extrahiert und im Sinne einer Datenbank standardisiert wiedergegeben. Die qualitativen Merkmale umfassten die Identifikation (Autor, Jahr, Land), den Evaluationstyp und die Perspektive der gesundheitsökonomischen Evaluation, das Alter der Zielpopulation, das Setting, die zu vergleichenden

Technologien, die Wirkungsdimension bzw. den Zeithorizont, den Ursprung der klinischen Effektdaten (Originalstudie, Literaturstudie, mit bzw. ohne entscheidungsanalytische Modellierung), die Berücksichtigung von Nutzwerten (z.B. QALYs), die berücksichtigten Kostenarten, Art und Erhebung des Mengengerüsts, Art und Quelle der Preise, Währung und Bezugsjahr, Diskontierungsraten, Behandlung von Unsicherheiten (Variablen der Sensitivitätsanalysen) und die Autorenschlussfolgerungen.

Als quantitative Ergebnisparameter wurden systematisch extrahiert und berichtet: Kosten (individuell oder populationsaggregiert), Effektivität und Kosten-Effekt-Relation. Je nach Bericht in der Publikation wurden durchschnittliche und/oder inkrementelle Werte für jede der in den einzelnen Studien untersuchten Vergleichstechnologien angegeben.

Ferner wurden die in den Publikationen diskutierten oder im Rahmen der Bewertung der Studienqualität aufgefallenen Bias-Typen systematisch dokumentiert.

Die qualitativen Studienmerkmale und die quantitativen Ergebnisparameter wurden systematisch in Tabellenform zusammengestellt (Aufbau s. Tabelle 15: Extrahierte Merkmale und Ergebnisparameter aus den berücksichtigten gesundheitsökonomischen Untersuchungen zum Screening des Prostatakarzinoms).

Um einen Überblick über die Kosten der einzelnen Screening-Strategien und deren Kosteneffektivität zu erhalten, wurden zum einen detailliertere Angaben zu den Kostenarten zusammengestellt, die in den evaluierten Studien in die Analyse einbezogen wurden. Zum anderen wurden die wichtigsten Kosten-Effektivitäts-Parameter extrahiert und aufgelistet, die über alle evaluierten Studien hinweg vergleichbar waren. Damit sollte aufgezeigt werden, inwieweit die Studien gleiche oder verschiedene Kostenarten berücksichtigten und ob bzw. wie sich dies in den Ergebnissen zur Kosten-Effektivität widerspiegelt. Gleichzeitig dient der angegebene Katalog der Kostenarten als Vorschlag für einen Standard bei Kosten-Effektivitäts-Analysen von Screening-Technologien bei Tumorerkrankungen. Bei der Erstellung des Kostenartenkataloges wurden die Vorschläge von Krahn & Naglie⁸¹ berücksichtigt.

C.4.5.1 Berücksichtigte Kostenarten

Die Kostenarten wurden in direkte medizinische Kosten und direkte nichtmedizinische Kosten unterteilt. Als weitere Kategorie wurden

Produktivitätsausfallskosten definiert. Der in diesem Zusammenhang umstrittenen Begriff der indirekten Kosten wurde dadurch vermieden.⁴⁸ Bei den meisten Kosten handelt es sich um variable Kosten, unter den direkten Kosten wurden als Fixkosten im weiteren Sinne die Kosten für die Programmimplementierung und Organisation berücksichtigt und als Fixkosten im engeren Sinne Kosten für Personal, Equipment und Räume. Tabelle 8 zeigt die insgesamt unterschiedenen Kostenarten.

Tabelle 8: Kostenarten in Screening-Programmen für Tumorerkrankungen

Direkte medizinische Kosten
Programmimplementierung, Organisation
Rekrutierung
Screening-Untersuchung
Konfirmatorische Diagnostik
Weiterverfolgung falschpositiver, schwachpositiver oder zweifelhafter Testergebnisse
Behandlung identifizierter Karzinome
Behandlung von fortgeschrittener oder metastasierender Erkrankung
Behandlung von Behandlungskomplikationen
Regelmäßige posttherapeutische Nachsorge
Kosten für die Behandlung anderer Erkrankungen aufgrund zusätzlicher Lebensjahre
Fixkosten (Personal, Equipment, Räume)
Direkte nichtmedizinische Kosten
Transport zur Screening-Untersuchung
Transport zur Behandlung
Haushaltshilfe / Pflege
Produktivitätsausfallskosten
durch Screening-Untersuchung (Reisen, Warten)
durch Behandlung / Langzeitmorbidity / Frühmortality

C.4.5.2 Kosten und Kosten-Effektivität

Zur Darstellung der Kosten bzw. Kosten-Effektivität wurden folgende Maßzahlen für alle evaluierten Studien extrahiert:

- die Kosten pro Screening-Teilnehmer (Kosten/ST)
- die Kosten pro identifizierter Krebsfall (Kosten/Ca)

- die Kosten pro gerettetes Lebensjahr (Kosten/LJ)
- die Kosten pro gewonnenes qualitätsadjustiertes Lebensjahr (Kosten/QALY)

Der Vergleich dient als rein deskriptiver Vergleich, da in den einzelnen gesundheitsökonomischen Studien zum Teil unterschiedliche Strategien oder Strategievarianten bzw. PSA-Grenzwerte analysiert wurden. Ferner unterschieden sich die Untersuchungen hinsichtlich weiterer wichtiger Faktoren, wie Perspektive, Zeithorizont, Altersgruppen, dem Jahr der Durchführung und insbesondere der berücksichtigten Kostenarten.

Wo mehrere Strategien gleichzeitig analysiert wurden oder Subgruppenanalysen (z.B. für verschiedene Altersgruppen) durchgeführt wurden, wurden die Spannweiten der Ergebnisse berichtet. Beim multiplen Vergleich von Strategien wurden teils Technologien eingesetzt, die kein PSA verwendeten, so dass auch solche Technologien in den Spannweiten mit enthalten sind. Zunächst wurden die rohen Ergebnisse - ohne Preisanpassungen, Berücksichtigung von Inflation/Deflation oder Währungsumrechnungen - dargestellt. Bei allen Angaben wurde zwischen durchschnittlichen und inkrementellen Maßzahlen unterschieden.

C.4.5.3 Währungskonversion und Inflationsbereinigung

Da die Ergebnisse in unterschiedlichen Währungen vorlagen, wurden Währungskonversionen durchgeführt. Zur Umrechnung in DM wurden die von der OECD veröffentlichten Kaufpreisparitäten des Jahres 1993 herangezogen (Quelle: OECD Health Data 97). Auf zusätzliche Deflationierung wurde verzichtet, da in Deutschland derzeit keine zuverlässigen sektoralen Inflationsraten für Gesundheitsgüter zur Verfügung stehen. Alle bereinigten Angaben wurden sowohl für die durchschnittlichen als auch für die inkrementellen Maßzahlen durchgeführt.

C.5 Ergebnisse

C.5.1 Bewertungsinstrumentarium für gesundheitsökonomische Studien

Die im folgenden vorgestellten drei Kriterienkataloge (Deskriptionskatalog, Transparenzkatalog und Qualitätskatalog) wurden gemeinsam entwickelt von den drei gesundheitsökonomischen Projektgruppen München (Federführung), Hannover und Ulm (vgl. zu den Mitgliedern des Panels, welche die Kataloge erarbeitet haben Tabelle 5).

Um unterschiedlichen Ansprüchen und Möglichkeiten der Beurteilung gesundheitsökonomischer Studien gerecht zu werden, wurden drei Kataloge (Deskriptionskatalog, Transparenzkatalog, Qualitätskatalog) mit unterschiedlicher Bewertungstiefe entwickelt (vgl. dazu auch oben Kapitel C.4.1).

Vor allem der Transparenz- und der Qualitätskatalog fokussieren vollständige ökonomische Evaluationen, welche sich durch den Vergleich von zwei oder mehr Technologien und die gleichzeitige Berücksichtigung von Kosten und Effekten auszeichnen. Für eine Optimierung der Ressourcenallokation im Gesundheitswesen sind generell vergleichende Studien erforderlich. Trotzdem können für bestimmte Fragestellungen auch nicht vergleichende Studien - d.h. Kostenbeschreibungen sowie Krankheitskosten-Studien - berechtigt sein. Nichtvergleichende Studien gelten als partielle ökonomische Evaluationen, welche die relevanten Transparenz- bzw. Qualitätskriterien gleichermaßen erfüllen sollten.

C.5.1.1 Deskriptionskatalog

Der Deskriptionskatalog ist in Tabelle 24 im Anhang 2: Deskriptionskatalog dargestellt. Dieser Katalog ist eine detaillierte Zusammenstellung der in gesundheitsökonomischen Evaluationen möglichen Varianten der Studienanlage, -form, -durchführung und -auswertung. In Form eines rasterartigen Leitfadens dient er dazu, aufzuzeigen, was an methodisch Möglichem in der jeweils vorliegenden Studie tatsächlich realisiert wurde. Er unterstützt damit den grundlegenden Schritt der Informationsidentifikation bzw. -erfassung und eignet sich gleichzeitig als Schulungsinstrument.

Es würde den Rahmen dieses Berichts sprengen, alle im Deskriptionskatalog spezifizierten Evaluationsmerkmale und deren konkret zulässigen bzw. gewünschten Ausprägungen einzeln aufzugreifen und zu erörtern. Deshalb werden im Anschluss an die Darstellung des Deskriptionskataloges lediglich dessen übergeordnete Themenblöcke - "Publikation", "Fragestellung und Evaluationsrahmen", "Gesundheitseffekte (gesundheitsbezogenes Outcome)", "Kosten (ökonomisches Outcome)", "Diskontierung", "Ergebnispräsentation" und "Behandlung von Unsicherheiten" - kurz kommentiert.

Im Anhang 3: Erläuterungen zu den gesundheitsökonomischen Kriterienkatalogen befinden sich Ausführungen zu den Teilbereichen des Deskriptionskataloges, die ferner für die praktische Anwendung aller drei gesundheitsökonomischen Kriterienkataloge (Deskriptionskatalog, Transparenzkatalog und Qualitätskatalog) hilfreich sind.

C.5.1.2 Transparenzkatalog

Die Anforderungen an die Transparenz von gesundheitsökonomischen Untersuchungen lassen sich in 10 Themenbereiche - "Fragestellung", "Evaluationsrahmen", "Analysemethoden und Modellierung", "Gesundheitseffekte", "Kosten", "Diskontierung", "Ergebnispräsentation", "Behandlung von Unsicherheiten", "Diskussion" und "Schlussfolgerungen" - mit insgesamt 56 Fragen organisieren, die sich auf die Prüfung der formalen Beachtung von Standards der gesundheitsökonomischen Evaluation beziehen. So gilt es etwa festzustellen, inwieweit der Problemkontext in der Studie dargestellt wurde, inwieweit der Zeithorizont für Kosten und Effekte angegeben wurde oder inwieweit die Datenquellen für Kosten und Effekte genannt wurden.

Eine solche Beurteilung der Transparenz und Nachvollziehbarkeit von Studien ermöglicht die Identifizierung von Studien, welche formalen Anforderungen nicht genügen. Jedoch kommt dabei nicht zum Ausdruck, ob die gewählten Ansätze, Methoden, Studiengrößen und Parameter im situativen Kontext der Studie auch angemessen gewählt bzw. korrekt umgesetzt wurden.

Zu Erklärungen und Hintergrundinformationen bezüglich der im Transparenzkatalog abgefragten Themenbereiche sei auf die Ausführungen im Anhang 3: Erläuterungen zu den gesundheitsökonomischen Kriterienkatalogen und das Kapitel "Methoden gesundheitsökonomischer Evaluation" verwiesen.

Tabelle 9: Transparenzkatalog. Kriterienkatalog zur Bewertung der formalen Transparenz und Nachvollziehbarkeit gesundheitsökonomischer Studien (entwickelt im Konsensusverfahren von den gesundheitsökonomischen Projektgruppen München, Hannover, Ulm)

Transparenzkatalog		
Autoren, Titel und Publikationsorgan:	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung		
1. Wurde die Fragestellung formuliert? 2. Wurde der medizinische und ökonomische Problemkontext dargestellt?		
Evaluationsrahmen		
3. Wurden alle in die Studie einbezogenen Technologien beschrieben? 4. Werden mindestens zwei alternative Technologien explizit verglichen? 5. Wurde die Auswahl der Vergleichstechnologien begründet? 6. Wurde die Zielpopulation beschrieben? 7. Wurde der Zeithorizont für Kosten und Gesundheitseffekte angegeben? 8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt? 9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht? 10. Wurde die Perspektive der Untersuchung explizit genannt?		
Analysemethoden und Modellierung		
11. Wurden die statistischen Tests/Modelle zur Analyse der Daten beschrieben? 12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter dokumentiert (in der Publikation bzw. einem technischen Report)? 13. Wurden Annahmen explizit formuliert? 14. Wurden in entscheidungsanalytischen Modellen die Datenquellen für die Pfadwahrscheinlichkeiten angegeben?		

<p>Gesundheitseffekte</p> <p>15. Wurden die in der Studie untersuchten Gesundheitszustände aufgeführt?</p> <p>16. Wurden die Quellen für die Gesundheitseffektdaten angegeben?</p> <p>17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden beschrieben und wurden die Ergebnisse dargestellt? (falls auf einer einzelnen Studie basierend)</p> <p>18. Wurden die Methoden zur Identifikation, Extraktion und Synthese der Effektparameter beschrieben? (falls auf einer Informationssynthese basierend)</p> <p>19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und die Methoden und Messinstrumente dieser Bewertung angegeben?</p> <p>20. Wurden die Quellen der Bewertungsdaten für die Gesundheitszustände angegeben?</p> <p>21. Wurden Aussagen zur Evidenz der Gesundheitseffekte gemacht?</p>	
<p>Kosten</p> <p>22. Wurden die den Kosten zugrunde liegenden Mengengerüste angegeben?</p> <p>23. Wurden die Quellen und Methoden zur Ermittlung der Mengengerüste angegeben?</p> <p>24. Wurde die den Kosten zugrunde liegenden Preisgerüste angegeben?</p> <p>25. Wurden die Quellen und Methoden zur Ermittlung der Preise angegeben?</p> <p>26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes begründet?</p> <p>27. Wurden Daten zu Produktivitätsausfallskosten (falls berücksichtigt) getrennt aufgeführt?</p> <p>28. Wurde die Währung genannt?</p> <p>29. Wurden Angaben zu Währungskonversionen gemacht?</p> <p>30. Wurden Angaben zu Preisanpassungen bei Inflation oder Deflation gemacht?</p>	
<p>Diskontierung</p> <p>31. Wurden zukünftige Gesundheitseffekte <u>und</u> Kosten diskontiert?</p> <p>32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?</p> <p>33. Wurden die Diskontraten angegeben?</p> <p>34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung begründet?</p>	

<p>Ergebnispräsentation</p> <p>35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?</p> <p>36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?</p> <p>37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?</p> <p>38. Wurde eine Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?</p> <p>39. Wurden reine (nicht lebensqualitätsadjustierte) klinische Effekte berichtet?</p> <p>40. Wurden Ergebnisse in disaggregierter Form dargestellt?</p> <p>41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?</p>	
<p>Behandlung von Unsicherheiten</p> <p>42. Wurden univariate Sensitivitätsanalysen durchgeführt?</p> <p>43. Wurden multivariate Sensitivitätsanalysen durchgeführt?</p> <p>44. Wurde strukturelle Sensitivitätsanalysen durchgeführt?</p> <p>45. Wurden die in den Sensitivitätsanalysen berücksichtigten Werte oder Wertebereiche bzw. Strukturvarianten angegeben?</p> <p>46. Wurden die Ergebnisse der Sensitivitätsanalysen dargestellt?</p> <p>47. Wurden statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?</p>	
<p>Diskussion</p> <p>48. Wurde die Datenqualität diskutiert?</p> <p>49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis diskutiert?</p> <p>50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis diskutiert?</p> <p>51. Wurden die Einschränkungen und Schwächen der Studie diskutiert?</p> <p>52. Wurden Angaben zur Generalisierbarkeit der Ergebnisse gemacht?</p> <p>53. Wurden ethische und Verteilungsfragen diskutiert?</p> <p>54. Wurde das Ergebnis im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?</p>	
<p>Schlussfolgerungen</p> <p>55. Wurden Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?</p> <p>56. Wurde eine Antwort auf die Fragestellung gegeben?</p>	

C.5.1.3 Qualitätskatalog

Die Fragen des anschließend wiedergegebenen Qualitätskataloges beziehen sich auf die gleichen Themenbereiche und Einzelitems der ökonomischen Evaluationen wie die des Transparenzkataloges. Allerdings gilt es, zusätzlich die Angemessenheit bzw. korrekte Umsetzung der gewählten Ansätze, Methoden, Studiengrößen und Parameter im situativen Kontext der Studie zu beurteilen. Damit werden auch methodische und inhaltliche Zusammenhänge in dem durch die Studie vorgegebenen Rahmen aus Fragestellung, Evaluationsansatz und Studiendesign kritisch untersucht. So stellt sich hier etwa die Frage, inwieweit der Problemkontext ausreichend dargestellt wurde, inwieweit ein adäquater Zeithorizont für Kosten und Effekte gewählt wurde oder inwieweit geeignete Datenquellen für Kosten und Effekte verwendet wurden. Bei den Kriterien, bei denen eine Bewertung hinsichtlich der Angemessenheit bzw. korrekten Durchführung entfällt - beispielsweise bei der Forderung nach expliziter Nennung des Typs der gesundheitsökonomischen Evaluation -, zielen auch die Fragen des Qualitätskataloges auf die Prüfung der Transparenz ab.

Zu Erklärungen und Hintergrundinformationen bezüglich der im Qualitätskatalog abgefragten Themenbereiche sei auf die Ausführungen im Anhang 3: Erläuterungen zu den gesundheitsökonomischen Kriterienkatalogen und das Kapitel "Methoden gesundheitsökonomischer Evaluation" verwiesen.

Tabelle 10: Qualitätskatalog. Kriterienkatalog zur Bewertung der inhaltlichen und methodischen Studienqualität gesundheitsökonomischer Studien (entwickelt im Konsensusverfahren von den gesundheitsökonomischen Projektgruppen München, Hannover, Ulm)

Qualitätskatalog		
Autoren, Titel und Publikationsorgan:	1 = Kriterium erfüllt 1/2 = Kriterium teilweise erfüllt 0 = Kriterium nicht erfüllt nr = nicht relevant	1, ½, 0, nr
Fragestellung 1. Wurde die Fragestellung präzise formuliert? 2. Wurde der medizinische und ökonomische Problemkontext ausreichend dargestellt?		
Evaluationsrahmen 3. Wurden alle in die Studie einbezogenen Technologien hinreichend detailliert beschrieben? 4. Wurden alle im Rahmen der Fragestellung relevanten Technologien verglichen? 5. Wurde die Auswahl der Vergleichstechnologien schlüssig begründet? 6. Wurde die Zielpopulation klar beschrieben? 7. Wurde ein für die Fragestellung angemessener Zeithorizont für Kosten und Gesundheitseffekte gewählt und angegeben? 8. Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt? 9. Wurden sowohl Kosten als auch Gesundheitseffekte untersucht? 10. Wurde die Perspektive der Untersuchung eindeutig gewählt und explizit genannt?		
Analysemethoden und Modellierung 11. Wurden adäquate statistische Tests/Modelle zur Analyse der Daten gewählt und hinreichend gründlich beschrieben? 12. Wurden in entscheidungsanalytischen Modellen die Modellstruktur und alle Parameter vollständig und nachvollziehbar dokumentiert (in der Publikation bzw. einem technischen Report)? 13. Wurden die relevanten Annahmen explizit formuliert? 14. Wurden in entscheidungsanalytischen Modellen adäquate Datenquellen für die Pfadwahrscheinlichkeiten gewählt und eindeutig genannt?		

<p>Gesundheitseffekte</p> <p>15. Wurden alle für die gewählte Perspektive und den gewählten Zeithorizont relevanten Gesundheitszustände berücksichtigt und explizit aufgeführt?</p> <p>16. Wurden adäquate Quellen für die Gesundheitseffektdaten gewählt und eindeutig genannt?</p> <p>17. Wurden das epidemiologische Studiendesign und die Auswertungsmethoden adäquat gewählt und beschrieben und wurden die Ergebnisse detailliert dargestellt? (falls auf einer einzelnen Studie basierend)</p> <p>18. Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben? (falls auf einer Informationssynthese basierend)</p> <p>19. Wurden die verschiedenen Gesundheitszustände mit Präferenzen bewertet und dafür geeignete Methoden und Messinstrumente gewählt und angegeben?</p> <p>20. Wurden adäquate Quellen der Bewertungsdaten für die Gesundheitszustände gewählt und eindeutig genannt?</p> <p>21. Wurde die Evidenz der Gesundheitseffekte ausreichend belegt? (s. ggf. entsprechende Kontextdokumente)</p>	
<p>Kosten</p> <p>22. Wurden die den Kosten zugrunde liegenden Mengengerüste hinreichend gründlich dargestellt?</p> <p>23. Wurden adäquate Quellen und Methoden zur Ermittlung der Mengengerüste gewählt und eindeutig genannt?</p> <p>24. Wurden die den Kosten zugrunde liegenden Preisgerüste hinreichend gründlich beschrieben?</p> <p>25. Wurden adäquate Quellen und Methoden zur Ermittlung der Preise gewählt und eindeutig genannt?</p> <p>26. Wurden die einbezogenen Kosten anhand der gewählten Perspektive und des gewählten Zeithorizontes schlüssig begründet und wurden alle relevanten Kosten berücksichtigt?</p> <p>27. Wurden Daten zu Produktivitätsausfallkosten (falls berücksichtigt) getrennt aufgeführt und methodisch korrekt in die Analyse einbezogen?</p> <p>28. Wurde die Währung genannt?</p> <p>29. Wurden Währungskonversionen adäquat durchgeführt?</p> <p>30. Wurden Preisanpassungen bei Inflation oder Deflation adäquat durchgeführt?</p>	
<p>Diskontierung</p> <p>31. Wurden zukünftige Gesundheitseffekte <u>und</u> Kosten adäquat diskontiert?</p> <p>32. Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die Kosten?</p> <p>33. Wurden die Diskontraten angegeben?</p> <p>34. Wurde die Wahl der Diskontraten bzw. der Verzicht auf eine Diskontierung plausibel begründet?</p>	

<p>Ergebnispräsentation</p> <p>35. Wurden Maßnahmen zur Modellvalidierung ergriffen und beschrieben?</p> <p>36. Wurden absolute Gesundheitseffekte und absolute Kosten jeweils pro Kopf bestimmt und dargestellt?</p> <p>37. Wurden inkrementelle Gesundheitseffekte und inkrementelle Kosten jeweils pro Kopf bestimmt und dargestellt?</p> <p>38. Wurde eine für den Typ der gesundheitsökonomischen Evaluation sinnvolle Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben?</p> <p>39. Wurden reine (nicht lebensqualitätsadjustierte) klinische Effekte berichtet?</p> <p>40. Wurden die relevanten Ergebnisse in disaggregierter Form dargestellt?</p> <p>41. Wurden populationsaggregierte Kosten und Gesundheitseffekte dargestellt?</p>	
<p>Behandlung von Unsicherheiten</p> <p>42. Wurden univariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?</p> <p>43. Wurden multivariate Sensitivitätsanalysen für die relevanten Parameter durchgeführt?</p> <p>44. Wurden Sensitivitätsanalysen für die relevanten strukturellen Elemente durchgeführt?</p> <p>45. Wurden in den Sensitivitätsanalysen realistische Werte oder Wertebereiche bzw. Strukturvarianten berücksichtigt und angegeben?</p> <p>46. Wurden die Ergebnisse der Sensitivitätsanalysen hinreichend dokumentiert?</p> <p>47. Wurden adäquate statistische Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten eingesetzt und die Ergebnisse berichtet?</p>	
<p>Diskussion</p> <p>48. Wurde die Datenqualität kritisch beurteilt?</p> <p>49. Wurden Richtung und Größe des Einflusses unsicherer oder verzerrter Parameterschätzung auf das Ergebnis konsistent diskutiert?</p> <p>50. Wurde Richtung und Größe des Einflusses struktureller Modellannahmen auf das Ergebnis konsistent diskutiert?</p> <p>51. Wurden die wesentlichen Einschränkungen und Schwächen der Studie diskutiert?</p> <p>52. Wurden plausible Angaben zur Generalisierbarkeit der Ergebnisse gemacht?</p> <p>53. Wurden wichtige ethische und Verteilungsfragen diskutiert?</p> <p>54. Wurde das Ergebnis sinnvoll im Kontext mit unabhängigen Gesundheitsprogrammen diskutiert?</p>	
<p>Schlussfolgerungen</p> <p>55. Wurden in konsistenter Weise Schlussfolgerungen aus den berichteten Daten/Ergebnissen abgeleitet?</p> <p>56. Wurde eine auf Wissensstand und Studienergebnissen basierende Antwort auf die Fragestellung gegeben?</p>	

C.5.2 Ergebnis der Literaturrecherche

C.5.2.1 Dokumentation des Selektionsprozesses

Mittels der im Methodenteil beschriebenen Vorgehensweise bei der Literaturrecherche lagen am Ende des Vorauswahlprozesses 36 Dokumente vor, davon 25 Primärstudien, sieben HTA-Berichte und vier Reviews. Diese wurden wie folgt zusammengetragen.

Literaturdatenbanken

Die letzte Aktualisierung der Literaturrecherche fand am 15. September 1998 statt. Unter Zugrundelegen der genannten Suchstrategie fanden sich in den Datenbanken EMBASE und MEDLINE EXPRESS zusammen 197 Eintragungen (EMBASE 109; MEDLINE EXPRESS 88 bei explizitem Ausschluss von EMBASE-Referenzen). Bei der manuellen Vorauswahl wurden 171 Eintragungen aus folgenden Gründen ausgeschlossen:

Tabelle 11: Ausschluss von Referenzen bei der manuellen Vorauswahl (Datenbanken EMBASE und MEDLINE EXPRESS). Stand 15.9.1998

Grund des Ausschlusses	Anzahl ausgeschlossener Referenzen
Zweifachnennung in den beiden Datenbanken (trotz expliziten Ausschlusses von EMBASE-Referenzen bei der Recherche in MEDLINE EXPRESS)	50
Ausführungen zum Screening ohne Kostenermittlung	42
Empfehlung zur Durchführung von Kosten-Effektivitäts-Untersuchungen bzw. zur Ermittlung von Kostenschätzern als Schlussfolgerung	21
Annahme der Kosten-Effektivität aufgrund einer alleinigen Effektivitätsermittlung	16
Äusserung von Kostenbedenken im Zusammenhang mit Screening, Beeinflussung der Haltung gegenüber dem Screening durch Kostenaspekte	16
Leitartikel, Briefe, personal opinions, Kommentare	7
Ermittlung der Kosten der PSA-Bestimmung bei BPH oder zur Verlaufskontrolle	6
Kostenschätzungen, Kostenannahmen ohne genaue Kostenermittlung	5
Ermittlung der Kosten-Effektivität von Interventionen infolge von PSA-Bestimmungen	4
"Kosten" im Zusammenhang mit "auf Kosten von"	3
"PSA" im Zusammenhang mit "potentiometric stripping analysis"	1

Drei der 26 bei der Vorauswahl in den Datenbanken EMBASE und MEDLINE EXPRESS berücksichtigten Referenzen waren HTA-Berichte (MEDLINE EXPRESS); die anderen 23 Referenzen waren Primärstudien und Reviews (je fünf in einer der beiden Datenbanken, 13 sowohl in EMBASE als auch in MEDLINE EXPRESS). Bei der Abfrage in den anderen genannten Datenbanken konnte eine zusätzlich in der Vorauswahl berücksichtigte Referenz (Datenbank HealthSTAR) geortet werden.

Einschlägige Veröffentlichungen sowie Fachzeitschriften

Bei der Durchsicht der Referenzlisten der zusammengetragenen Literatur konnten fünf weitere in der Vorauswahl berücksichtigte Publikationen identifiziert werden. Die im Publikationsorgan von ECRI für 1997 angekündigte Veröffentlichung eines HTA-Berichts hat sich verzögert und sollte erst in der zweiten Hälfte des Jahres 1998 erfolgen (schriftliche Mitteilung vom 06.04.1998, D.C. Cummins, ECRI).

Einrichtungen und Organisationen

Über die Internet-Seiten der im Methodenteil aufgeführten Einrichtungen konnten neben den drei bereits über die Datenbank MEDLINE EXPRESS identifizierten HTA-Berichten vier weitere aufgefunden werden.

C.5.2.2 Berücksichtigte Publikationen

Von den 36 im Rahmen der Vorauswahl identifizierten Arbeiten erfüllten insgesamt 16 Publikationen die im Methodenteil definierten Einschlusskriterien. Davon waren 14 Primärstudien bzw. -studiengruppen und zwei Reviews.

Der Veröffentlichung von Barry et al. (1995) im Rahmen einer vierteiligen Publikationsserie^{6-8,31} liegt der HTA-Bericht des OTA (1995) zugrunde, welcher als technischer Bericht ausgewiesen wird. Eine spätere Studie von Coley et al. (1997)³⁰ basiert auf den gleichen Annahmen, wurde jedoch für unterschiedliche Altersgruppen angefertigt. Beide Studien werden zusammen aufgeführt und mit dem OTA-Bericht als Ganzes bewertet.

Krahn et al.⁸¹ wiesen in ihrer Publikation auf einen technischen Report hin, der als Hintergrunddokument angefordert wurde.

Das von der Autorengruppe um Littrup veröffentlichte Zahlenmaterial zweier Studien (1993⁹¹ und 1994c⁹³) wurde in anderen Publikationen (1994a⁹⁰ und 1994b⁹² bzw. 1995⁸⁹) erneut dargestellt. Die später erschienenen Publikationen werden deshalb zusammen mit der entsprechenden Erstpublikation aufgelistet.

Der Studie von Ciatto et al. in italienischer Sprache (1994)²⁸ wurde in einer anderen Zeitschrift in englischer Übersetzung (1995)²⁹ veröffentlicht. Beide Publikationen werden zusammen aufgeführt.

Die für die Studienbewertung selektierten Primärstudien, -studiengruppen und Reviews werden im weiteren durchgehend als "Studien" (in den standardisierten Berichten wegen der eigenen Kostenberechnung auch als "Primärstudien") bezeichnet. Es handelt sich dabei um die folgenden 16 Studien und 7 Hintergrunddokumente.

Tabelle 12: In die gesundheitsökonomische Evaluation eingeschlossenen Publikationen

Studien	Ähnliche Publikationen, Hintergrunddokumente, technische Berichte
1. Abramson et al. (1994) ¹	
2. Barry et al. (1995) ⁷	- HTA-Bericht des OTA (1995) ¹⁴⁹ - Coley et al. (1997) ³⁰
3. Benoit & Naslund (1994) ¹¹	
4. Chadwick et al. (1995) ²⁵	
5. Ciatto et al. (1995) ²⁹	- Ciatto et al. (1994) ²⁸
6. de Biasi et al. (1996) ³⁴	
7. Dorr et al. (1993) ³⁷	
8. Gottlieb et al. (1996) ⁴⁹	
9. Gustafsson et al. (1995) ⁵²	
10. Kantrowitz et al. (1995) ⁷⁴	
11. Krahn et al. (1994) ⁸¹	- Technical Report (1994) ⁸⁰
12. Kramer et al. (1993) ⁸³	
13. Launois (1992) ⁸⁴	
14. Littrup et al. (1993) ⁹¹	- Littrup et al. (1994a ⁹⁰ und 1994b ⁹²)
15. Littrup et al. (1994c) ⁹³	- Littrup et al. (1995) ⁸⁹
16. Optenberg & Thompson (1990) ¹¹⁰	

C.5.2.3 Nicht berücksichtigte Publikationen

Die folgende Tabelle zeigt die nicht berücksichtigten gesundheitsökonomischen Publikationen mit dem jeweiligen Ausschlussgrund.

Tabelle 13: In der gesundheitsökonomischen Evaluation nicht berücksichtigte Publikationen mit Angabe des Ausschlussgrundes

Publikation	Ausschlussgrund
1. Benoit & Naslund (1996) ¹²	Konferenzabstract ohne Volltextpublikation
2. Simpson & Brown (1993) ¹³¹	Konferenzabstract ohne Volltextpublikation
3. Kovacs et al. (ohne Angaben bezüglich des Jahrgangs) ⁷⁹	Konferenzabstract ohne Volltextpublikation
4. Rosen (1995) ¹²¹	Konferenzabstract ohne Volltextpublikation
5. Nakagawa et al. (1997) ¹⁰²	nur japanischer Volltext
6. Grasso Leanza et al. (1997) ⁵⁰	nur italienischer Volltext
7. Benoit & Naslund (1997, 1997) ^{10,13}	Bei Kostenangaben Bezug auf eine angeblich im Druck befindliche Volltext-Publikation, die bis September 1998 in keiner der benutzten Datenbanken vermerkt wurde

C.5.3 Standardisierte Berichte zu berücksichtigten Publikationen und qualitative Informationssynthese

Der Inhalt der 16 gesundheitsökonomischen Studien, die die Einschlusskriterien erfüllen, wird in standardisierter Berichtform wiedergegeben. Diese Berichtform wurde für die gesundheitsökonomischen Studien entwickelt, um den Spezifika der ökonomischen Komponenten in den Publikationen gerecht zu werden. Um eine qualitative Informationssynthese zu ermöglichen, wurden anschließend die wichtigsten Merkmale und Ergebnisparameter aller berücksichtigten Studien tabellarisch zusammengefasst.

In der folgenden Übersicht ist die Dokumentationsstruktur der standardisierten Berichte wiedergegeben. Zu beachten ist die unterschiedliche Struktur für die Erfassung des Teils Gesundheitseffekte bei (1) Screening-/Diagnosestudien mit Originaldaten und bei (2) Synthesestudien, die sich auf bereits veröffentlichte Effektivitätsdaten beziehen.

Tabelle 14: Standardisierte Berichte. Dokumentationsstruktur für Original- und Synthesestudien

1 Fragestellung und Evaluationsrahmen	
1.1 Technologie	
1.2 Fragestellung	
1.3 Perspektive	
1.4 Zeithorizont	
1.5 Art der ökonomischen Evaluation	
2 Studiendesign und Studien- bzw. Zielpopulation	
2.1 Studientyp	
2.2 Datierung der zugrundeliegenden Daten	
2.3 Studienpopulation / Zielpopulation	
2.4 Setting	
2.5 Spezifikation der Technologie	
3 Gesundheitseffekte <i>(1) Screening-/Diagnosestudie</i>	3 Gesundheitseffekte <i>(2) Synthese bereits veröffentlichter Studien</i>
3.1 Untersuchte Zielgrößen	3.1 In der Synthese untersuchte klinische Parameter
3.2 Ein-/Ausschlusskriterien	3.2 Annahmen
3.3 Rekrutierungsmodus	3.3 Berücksichtigung von Primärstudien: Studiendesigns und Ein-/Ausschlusskriterien
3.4 Teilnehmerate	3.4 Quellen und Suchstrategie bei der Literaturrecherche
3.5 Reproduzierbarkeit der Testergebnisse	3.5 Validitäts- bzw. Qualitätskriterien bei der Bewertung der Primärstudien
3.6 Auswertung der Testergebnisse	3.6 Methoden der Bewertung von Relevanz und Validität bzw. Qualität der Primärstudien
3.7 Drop-Outs	3.7 Methoden der Extraktion von Daten aus den Primärstudien
3.8 Ergebnisse der Screening-/Diagnosestudie	3.8 Anzahl berücksichtigter Primärstudien
3.9 Effektmaße für die ökonomische Analyse	3.9 Methode der Synthese der gesundheitsbezogenen Parameter
	3.10 Untersuchung von Heterogenität der gesundheitsbezogenen Parameter
	3.11 Ergebnisse der Synthese
	3.12 Effektmaße für die ökonomische Analyse
4 Kosten	
4.1 Berücksichtigte Ressourcenveränderungen	
4.2 Beschreibung des Mengengerüsts	
4.3 Monetäre Bewertung des Mengengerüsts	
4.4 Währung	
5 Diskontierung	
6 Ergebnisse	
6.1 Ermittelte Gesundheitseffekte	
6.2 Ermittelte Kosten	
6.3 Synthese von Kosten und Effekten	
7 Behandlung von Unsicherheiten	
8 Diskussion und Schlussfolgerungen der Autoren	
8.1 Bemerkungen hinsichtlich Einschränkungen/Schwächen/Bias der Analyse	
8.2 Bemerkungen hinsichtlich der Generalisierbarkeit der Ergebnisse (externe Validität)	
8.3 Schlussfolgerungen	
9 Kommentar	
10 Ähnliche Publikationen / Originalpublikationen / Technische Reports (wenn vorhanden)	

Die standardisierten Berichte der berücksichtigten Publikationen befinden sich in alphabetischer Reihenfolge im Anhang.

Die Studien können in acht Untersuchungen mit eigener Erhebung der gesundheitsbezogenen Effekte ("Originaldaten")^{1,25,29,34,52,74,91,93} und acht Untersuchungen mit Rückgriff auf Literaturdaten bezüglich der Gesundheitseffekte ("Synthesestudien")^{8,11,37,49,81,84,110} unterteilt werden. Dabei lassen sich zwei übergeordnete Fragestellungen erkennen:

Acht Studien^{1,11,25,29,34,52,74,84} stellen hauptsächlich die Diagnose von Prostatakarzinom im asymptomatischen Stadium in den Vordergrund, bei sechs Studien^{8,37,49,81,83,110} interessieren mehr die therapeutischen und gesundheitsbezogenen Konsequenzen einer Diagnose. Zwei Studien^{91,93} nehmen eine Zwitterstellung ein.

In vier Studien^{8,81,84,110} wird explizit, in zwei Studien^{37,83} implizit der Status quo bzw. die Behandlung bei auftretender Symptomatik als Vergleichstechnologie zum Screening gewählt, in einer Studie⁴⁹ ist dies die Alternative "Keine Biopsie". In fünf Studien^{11,29,52,91,93} werden unterschiedliche Screening-Strategien ohne Einbeziehung des Status quo als Vergleichstechnologie gegenübergestellt. In vier Studien^{1,25,34,74} erfolgt keine Gegenüberstellung von Technologien.

Alle Studien fokussieren das Prävalenz-Screening. Auf das Inzidenz-Screening wird nur in einer Untersuchung⁹¹ näher eingegangen. Der gewählte Zeithorizont umfasst in vier Untersuchungen^{8,49,81,84} die Restlebensdauer. In einer Publikation¹¹⁰ wird die 5-Jahres-Überlebensrate in Betracht gezogen. Die restlichen elf Studien gehen über die diagnostische Abklärung (mit oder ohne Therapie) nicht hinaus.

In zwölf Studien^{1,8,25,29,34,37,52,74,83,84,110} wird als hauptsächlich gesundheitsbezogenes Outcome ein rein klinischer, morbiditäts- oder mortalitätsbezogener Effekt-Parameter untersucht. In vier Studien werden zusätzlich Lebensqualitätsaspekte^{49,81,91,93} berücksichtigt. Als gesundheitsökonomischer Studientyp kommen elf Kosten-Effektivitäts-Analysen,^{1,8,11,25,29,34,37,52,74,84,110} eine Kosten-Konsequenzen-Analyse⁸³ und zwei Kosten-Nutzwert-Analysen^{49,81} vor. Zwei Studien^{91,93} führten sowohl eine Kosten-Effektivitäts- als auch eine Kosten-Nutzen-Analyse durch.

Modellierungen wurden in zehn Studien eingesetzt: Zwei Studien^{8,81} setzten Markov-Kohorten ein. In den anderen acht Studien^{37,49,52,83,84,91,93,110} wurde auf

Entscheidungsbäume zurückgegriffen, wobei zusätzlich in einer Studie Sterbetafel¹¹⁰ und in zwei weiteren DEALE^{49,84} zur Generierung langfristiger Outcome-Parameter verwendet wurden.

Drei Studien^{8,81,84} wählen explizit die Perspektive des Kostenträgers, in einer⁴⁹ wird explizit der gesellschaftliche Standpunkt angegeben. In den anderen zwölf Studien wird keine Perspektive genannt, der implizit eingenommene Standpunkt lässt sich jedoch als der eines Kostenträgers auslegen.

In allen Studien werden direkte variable Kosten der medizinischen Versorgung erfasst. Dabei werden Kosten für das Screening und die diagnostische Abklärung in allen Studien untersucht, während Therapiekosten ohne Berücksichtigung von Komplikationen in zwei Studien^{1,49} und Therapiekosten mit Berücksichtigung von Komplikationen in sieben Studien^{8,37,81,83,91,93,110} ermittelt werden. Lediglich in einer Untersuchung⁵² wird direkt beim Patienten anfallender Ressourcenverbrauch, (d.h. Transport- und Zeitkosten) erfasst. Rekrutierungs- und Organisationskosten werden in zwei Studien^{29,74} in Betracht gezogen. Produktivitätskosten und intangible Kosten werden in zwei Studien^{91,93} explizit genannt, ansonsten nicht oder aber bei der Ermittlung der Lebensqualität miteingefasst.^{49,81}

In der Rubrik "Kommentare" (Ziffer 9 des Berichtsrasters) wurden Bemerkungen zu systematischen Fehlern und Schwächen der Studien festgehalten. Nur in einer Studie⁵² werden Maßnahmen zur Kontrolle eines selection bias dokumentiert, mit dem entweder aufgrund des Fehlens einer randomisierten Rekrutierung von Studienteilnehmern in den beschriebenen Querschnitts- bzw. Longitudinalstudien oder aufgrund fehlender Angaben zur Methodik der Datenidentifikation und -bewertung bei Literatursynthesen zu rechnen ist. Alle Studien unterliegen einem work-up bias, der lediglich in einer Untersuchung⁸⁴ im Rahmen einer Sensitivitätsanalyse berücksichtigt wurde. Bemerkenswert ist außerdem, dass in zehn von 16 Studien^{1,8,11,29,34,49,74,91,93,110} Unzulänglichkeiten der Studierendokumentation wie Inkongruenzen zwischen Text und Tabellen bzw. Abbildungen, eingeschränkte Nachvollziehbarkeit der Effekt- bzw. Kostenermittlung oder fehlerhafte Kostenberechnungen festzustellen waren.

Tabelle 15 dient der komprimierten und im Sinne einer Datenbank standardisierten Darstellung von Fragestellung (verglichene Screening-Strategien), Evaluationsrahmen, Zielpopulation, Methodik, Ergebnissen zu Effektivität, Kosten und Kosteneffektivität, Diskontierungsraten, Behandlung von Unsicherheiten,

Autorenschlussfolgerungen und Limitierungen bezüglich interner Validität, Generalisierbarkeit und Übertragbarkeit.

Tabelle 15: Extrahierte Merkmale und Ergebnisparameter aus den berücksichtigten gesundheitsökonomischen Untersuchungen zum Screening des Prostatakarzinoms

Autor, Jahr, Land	Abramson et al. (1994) ¹ : USA	Barry et al. (1995) ⁸ / Coley et al. (1997) ³⁰ : USA
Evaluationstyp	Kosten-Effektivitäts-Analyse	Kosten-Effektivitäts-Analyse
Perspektive	kA	Kostenträger
Alter Zielpopulation (in Jahren)	≥40	Barry:65, 70, 75 Coley: 50-59, 60-69, 70-79
Setting	Prävalenz-Screening Gemeindebezogen / Rekrutierung über Lokalzeitung: n=564	Prävalenz-Screening
Verglichene Alternativen (mit Biopsie-Kriterien)	A: DRU,PSA10 (bzw. 2xPSA4) B: --	A) PSA4,DRU B) Behandlung bei Symptomatik
Wirkungsdimensionen/ Zeithorizont	Anzahl diagnostizierter PCa	Restlebenserwartung
Klinische Effektdaten	Originaldaten (Querschnittsstudie)	Literatursynthese mit Modellierung (Markov-Kohorten)
Nutzwerte/Nutzenwerte	--	--
Kosten - Beschreibung	Direkte variable Kosten für Screening, Diagnose und primäre Therapie	Direkte variable Kosten für Screening, Diagnose, primäre und sekundäre Therapie
Mengengerüst - Beschreibg. Erhebung	Aufwand der Leistungserstellung, erstellte (Zwischen-)Produkte Originaldaten (retrospektiv)	Erstellte (Zwischen-)Produkte: CPT-4 / DRG Modellierung
Preise - Art Quelle	Unklar Institutionseigene Daten	Administrative Preise Medicare fee schedule, basierend auf RBRVS
Währung, Bezugsjahr	US-\$, 1992	US-\$, 1992
Diskontierung	--	5%
Ergebnisse		Im folgenden Wertetripel für versch. Altersgr.:
Kosten: Durchschnittlich	US-\$ 231 /ST ohne Therapie US-\$ 520 /ST mit Therapie	
<i>Inkrementell</i>		Barry et al.: US-\$ 293/365/414 /ST ^a Coley et al.: US-\$ 216/387/532 /ST ^a
Effekte: Durchschnittlich	0.03PCa /ST (1PCa A, 14PCa B, 3PCa C)	
<i>Inkrementell</i>		Barry et al.: 7.5/5.3/2.9d/ST bzw. 16/10/5d ^{oD} /ST 1.52/0.85/0.43y ^{oD} /PCa-RP Coley et al.: 6 ^{8f} /7/3d/ST 17/17/6d ^{oD} /ST
Kosten-Effekt-Relation: Durchschnittlich	US-\$ 7,240 /PCa ohne Therapie US-\$ 16,300 /PCa mit Therapie	
<i>Inkrementell</i>		Barry et al.: US-\$ 15,123/25,289/51,267 /y ^a Coley et al.: US-\$ 12,491/18,769 /65,909 /y ^a
Behandlung Unsicherheiten	--	Mortalität bei Metastas., Heilungsrate PCa-RP, Kosten term. Versorg., SV: PCa-ST vs. PCa-RP
Autorenschlussfolgerung	Weiterer Klärungsbedarf beim Screening	Weiterer Klärungsbedarf beim Screening
Bias/methodische Schwächen	1,3,4,5,10,11,12	3,4,6,7,9,11

(Fortsetzung)

Autor, Jahr, Land	Benoit & Naslund (1994)¹¹ : USA	Chadwick et al. (1991)²⁵ : Großbritannien
Evaluationstyp	Kosten-Effektivitäts-Analyse	Kosten-Effektivitäts-Analyse
Perspektive	kA	kA
Alter Zielpopulation (in Jahren)	50-70	55-69
Setting	Prävalenz-Screening	Prävalenz-Screening Ambulante Praxis-Versorgung / Verschiedene Rekrutierungsmodi: n=472 (TR 58%)
Verglichene Alternativen (mit Biopsie-Kriterien)	A: 1) PSA4,DRU 2) PSA4→DRU,TRUS	A: PSA4→TRUS,DRU→TRUS
	B: --	B: --
Wirkungsdimensionen/Zeithorizont	Anzahl diagnostizierter Karzinome	Anzahl diagnostizierter PCa
Klinische Effektdaten	Literatursynthese (2 Longitudinalstudien)	Originaldaten (Querschnittsstudie)
Nutzwerte/Nutzenwerte	--	--
Kosten - Beschreibung	Direkte variable Kosten für Screening und Diagnose	Direkte variable Kosten für Screening und Diagnose
Mengengerüst - Beschreibg.	Erstellte (Zwischen-)Produkte	Aufwand der Leistungserstellung, erstellte (Zwischen-)Produkte
Erhebung	Literatur (2 Longitudinalstudien)	Originaldaten (retrospektiv)
Preise - Art Quelle	Administrative Preise Medicare allowed charges (MAC), basierend auf RBRVS	Unklar kA
Währung, Bezugsjahr	US-\$, 1992	GB-£, kA
Diskontierung	--	--
Ergebnisse		
Kosten: Durchschnittlich	1) US-\$ 109 /ST 2) US-\$ 55 /ST	GB-£ 25 /ST
<i>Inkrementell</i>		
Effekte: Durchschnittlich	1) 0.05PCa /ST 2) 0.03PCa /ST	0.017PCa /ST (5PCa pT2pN0M0, 2PCa T0N0M0)
<i>Inkrementell</i>		
Kosten-Effekt-Relation: Durchschnittlich	1) US-\$ 2,372 /PCa 2) US-\$ 2,205 /PCa	GB-£ 1,654 /PCa
<i>Inkrementell</i>		
Behandlung Unsicherheiten	Überweisungsrate, Kosten der DRU, SV: serielles Screening über 3 Jahre	--
Autorenschlussfolgerung	Screening zu empfehlen	Weiterer Klärungsbedarf beim Screening
Bias/methodische Schwächen	3,4,6,7,9,(11)	1,4,5

(Fortsetzung)

Autor, Jahr, Land	Ciatto et al. (1995)²⁹ : Italien	de Biasi et al. (1996)³⁴ : Italien
Evaluationstyp	Kosten-Effektivitäts-Analyse	Kosten-Effektivitäts-Analyse
Perspektive	kA	kA
Alter Zielpopulation (in Jahren)	60-75	55-74
Setting	Prävalenz-Screening Ambulante Praxis-Versorgung / Rekrutierung über Einladungsschreiben: n=1,421 (TR 33.7%) bei 1) bzw. n=1,315 (TR 66.9%) bei 2)	Prävalenz-Screening Gemeindebezogen / Rekrutierung im Rahmen einer laufenden Longitudinalstudie (zu KHK): n=472 (TR 91%)
Verglichene Alternativen (mit Biopsie-Kriterien)	A: 1) DRU,TRUS 2) PSA4→DRU,TRUS,PSA10 ^{bei PSA-D>0.15} B: --	A: PSA3→DRU→TRUS B: --
Wirkungsdimensionen/ Zeithorizont	Anzahl diagnostizierter PCa	Anzahl diagnostizierter PCa
Klinische Effektdaten	Originaldaten (2 Querschnittsstudien)	Originaldaten (Querschnittsstudie)
Nutzwerte/Nutzenwerte	--	--
Kosten - Beschreibung	Direkte variable Kosten für Screening, Diagnose, Organisation und Datenauswertung	Direkte variable Kosten für Screening und Diagnose
Mengengerüst - Beschreibg. Erhebung	Aufwand der Leistungserstellung, erstellte (Zwischen-)Produkte Originaldaten (retrospektiv)	Erstellte (Zwischen-)Produkte Originaldaten (retrospektiv)
Preise - Art Quelle	Unklar kA	Administrative Preise kA
Währung, Bezugsjahr	Lit., kA	Lit., kA
Diskontierung	--	--
Ergebnisse		
Kosten: Durchschnittlich	1) Lit. 33,750 /ST 2) Lit. 30,400 /ST	Lit. 22,000 /ST
<i>Inkrementell</i>		
Effekte: Durchschnittlich	1) 0.018PCa /ST (9PCa G-Sc5, 5PCa G-Sc6, 10PCa G-Sc7, 1PCa G-Sc9) 2) 0.016PCa /ST (5PCa G-Sc5, 6PCa G-Sc6, 2PCa G-Sc7, 6PCa G-Sc 8, 3PCa G-Sc9)	0.023PCa /ST (2PCa G-Sc2, 6PCa G-Sc3, 1PCa G-Sc4, 2PCa G-Sc5)
<i>Inkrementell</i>		
Kosten-Effekt-Relation: Durchschnittlich	1) Lit. 1,854,000 /PCa 2) Lit. 1,817,000 /PCa	Lit. 932,000 /PCa
<i>Inkrementell</i>		
Behandlung Unsicherheiten	SV: PSA-Grenzwert (PSA9)	SV: PSA-Grenzwert (PSA4)
Autorenschlussfolgerung	Früherkennung durch Screening realisierbar	PSA-Screening zu empfehlen
Bias/methodische Schwächen	1,4,5,12	1,4,5,(10)

(Fortsetzung)

Autor, Jahr, Land	Dorr et al. (1993)³⁷: USA	Gottlieb et al. (1994)⁴⁹: USA
Evaluationstyp	Kosten-Konsequenzen-Analyse Kosten-Effektivitäts-Analyse	Kosten-Nutzwert-Analyse
Perspektive	kA	Gesellschaftliche Perspektive
Alter Zielpopulation (in Jahren)	50-75	≥50
Setting	Prävalenz-Screening Nationbezogen (n=21.3Mio.)	Prävalenz-Screening und Inzidenz-Screening
Verglichene Alternativen (mit Biopsie-Kriterien)	A) PSA4 B) Status quo (<i>implizit</i>)	A: Biopsie in Abhängigkeit vom Exzess-PSA ^e (<i>Biopsie in Abhängigkeit von PSA-V im 2. Jahr</i>) B: Keine Biopsie
Wirkungsdimensionen/Zeithorizont	Anzahl diagnostizierter PCa, Anzahl verhinderter PCa-Todesfälle	Qualitätsadjustierte Restlebenserwartung
Klinische Effektdaten	Literatursynthese (Narrativer Review) mit Modellierung (Entscheidungsbaumanalyse)	Literatursynthese mit Modellierung (Entscheidungsbaumanalyse und DEALE)
Nutzwerte/Nutzenwerte	--	Literatur
Kosten - Beschreibung	Direkte variable Kosten für Screening, Diagnose, primäre und sekundäre Therapie	Direkte variable Kosten für Diagnose und primäre Therapie in 2 Jahren
Mengengerüst - Beschreibg.	Erstellte (Zwischen-)Produkte	Erstellte (Zwischen-)Produkte
Erhebung	Modellierung	Modellierung
Preise - Art Quelle	Administrative Preise kA ^b	Unklar Institutionseigene Daten
Währung, Bezugsjahr	US-\$, kA	US-\$, kA
Diskontierung	--	--
Ergebnisse		
Kosten: Durchschnittlich	Populationsbetrachtung: A) US-\$ 12.7Mrd. (n=21,266,100) ⇒ US-\$ 597/ST ^{er}	
<i>Inkrementell</i>		
Effekte: Durchschnittlich	Populationsbetrachtung: A) 1.0Mio.PCa mit 5,518TF B) 34,000 verhinderte PCa-Todesfälle	
<i>Inkrementell</i>		
Kosten-Effekt-Relation:		
<i>Durchschnittlich</i>	US-\$ 1,924 /PCa ^c US-\$ 197,647 verhindertem PCa-Tod ^d (Annahme: Mortalitätssenkung um 30%)	
<i>Inkrementell</i>		60J: US-\$ 13,558 /QALY - US-\$ 8,974 /QALY bei Exzess-PSA0 ^e bis Exzess-PSA20 ^e 50J: Biopsie kostengünstiger und effektiver 70J: Biopsie mit Verlust an QALYs
Behandlung Unsicherheiten	--	Alle Hauptvariablen
Autorenschlussfolgerung	Weiterer Klärungsbedarf beim Screening	rein PSA-gestützte Biopsie nicht zu empfehlen
Bias/methodische Schwächen	3,4,(6),7,8,9	2,4,6,7,8,9,11,12

(Fortsetzung)

Autor, Jahr, Land	Gustafsson et al. (1995) ⁵² : Schweden	Kantrowitz et al. (1995) ⁷⁴ : USA	
Evaluationstyp	Kosten-Effektivitäts-Analyse	Kosten-Effektivitäts-Analyse	
Perspektive	kA	kA	
Alter Zielpopulation (in Jahren)	55-70	50-65	
Setting	Prävalenz-Screening Ambulante KH-Versorgung / Randomisierte Stichprobenziehung: n=1,782 (TR 74%)	Prävalenz-Screening Arbeitsplatzbezogen / Rekrutierung über Kampagne und persönliche Aufforderung: n=1,219 (TR 54%)	
Verglichene Alternativen (mit Biopsie-Kriterien)	A: 1) DRU 2) TRUS 3) DRU,TRUS,PSA76) 4) PSA4,DRU 5) PSA4,TRUS DRU,PSA4→TRUS	DRU,PSA4→TRUS,PSA10	
	B: --	B: --	
Wirkungsdimensionen/Zeithorizont	Anzahl diagnostizierter PCa, Anzahl diagnostizierter PCa≤T2a, Anzahl PCa-T	Anzahl diagnostizierter PCa	
Klinische Effektdaten	Originaldaten (Querschnittsstudie) mit Modellierung (Entscheidungsbaumanalyse)	Originaldaten (Querschnittsstudie)	
Nutzwerte/Nutzenwerte	--	--	
Kosten - Beschreibung	Direkte variable Kost. für Rekrut., Screening u. Diagnose, Fixkost., Pat.kost. (Transp., Zeit)	Direkte variable Kosten für Rekrutierung, Screening und Diagnose	
Mengengerüst - Beschreibg.	Aufwand der Leistungserstellung, erstellte (Zwischen-)Produkte Erhebung Modellierung; 1,099 Probandenfragebögen	Aufwand der Leistungserstellung, erstellte (Zwischen-)Produkte Originaldaten (retrospektiv)	
Preise - Art Quelle	Pauschales Verfahren Institutionseigene Daten	Unklar kA	
Währung, Bezugsjahr	US-\$, 1990 (Wechselkurs aus skr)	US-\$, kA	
Diskontierung	--	--	
Ergebnisse			
Kosten: Durchschnittlich	1) US-\$ 75 /ST 2) US-\$ 98 /ST 3) US-\$ 161 /ST	4) US-\$ 71 /ST 5) US-\$ 83 /ST 6) US-\$ 116 /ST	US-\$ 59 /ST
<i>Inkrementell</i>			
Effekte: Durchschnittlich	1) 0.024PCa, 0.006PCa ^{≤T2a} , 0.015PCa-T /ST 2) 0.033PCa, 0.010PCa ^{≤T2a} , 0.020PCa-T /ST 3) 0.036PCa, 0.012PCa ^{≤T2a} , 0.023PCa-T /ST 4) 0.020PCa, 0.004PCa ^{≤T2a} , 0.012PCa-T /ST 5) 0.026PCa, 0.006PCa ^{≤T2a} , 0.018PCa-T /ST 6) 0.032PCa, 0.009PCa ^{≤T2a} , 0.021PCa-T /ST		0.01PCa /ST
<i>Inkrementell</i>			
Kosten-Effekt-Relation:			
<i>Durchschnittlich</i>	1) US-\$ 3,100 /PCa 2) US-\$ 2,950 /PCa 3) US-\$ 4,470 /PCa	4) US-\$ 3,560 /PCa 5) US-\$ 3,180 /PCa 6) US-\$ 3,630 /PCa	US-\$ 6,012 /PCa
<i>Inkrementell</i>	1)vs4) US-\$ 1,100 /PCa-T 5)vs1) US-\$ 2,700 /PCa-T 2)vs5) US-\$ 7,450 /PCa-T 6)vs2) US-\$ 18,600 /PCa-T 3)vs6) US-\$ 22,400 /PCa-T		
Behandlung Unsicherheiten	--	--	
Autorenschlussfolgerung	Kosten-Effektivität der Strategien beachten	Screening im Betriebsrahmen realisierbar	
Bias/methodische Schwächen	4	1,3,4,5,10,12	

(Fortsetzung)

Autor, Jahr, Land	Krahn et al. (1994) ⁸¹ : Kanada	Kramer et al. (1993) ⁸³ : USA
Evaluationstyp	Kosten-Nutzwert-Analyse	Kosten-Konsequenzen-Analyse
Perspektive	Kostenträger	kA
Alter Zielpopulation (in Jahren)	kA (Ergebnisse: 50, 60, 70)	50-74
Setting	Prävalenz-Screening	Prävalenz-Screening Nationbezogen (n=21.2Mio.)
Verglichene Alternativen (mit Biopsie-Kriterien)	A: 1) DRU→TRUS 2) PSA4→DRU,TRUS 3) DRU,PSA4→TRUS 4) PSA4,DRU,TRUS B) Behandlung bei Symptomatik	A) PSA4,DRU B) Status quo (implizit)
Wirkungsdimensionen/Zeithorizont	Qualitätsadjustierte Restlebenserwartung	Anzahl diagnostizierter PCa, Anzahl behandlungsbedingter Todesfälle bzw. PCa-Tote
Klinische Effektdaten	Literatursynthese mit Modellierung (Markov-Kohorten)	Literatursynthese (Narrativer Review) mit Modellierung (Entscheidungsbaumanalyse)
Nutzwerte/Nutzenwerte	time trade-off-Verfahren bei 10 Medizinern	--
Kosten - Beschreibung	Direkte var. Kosten für Screening, Diagnose, primäre und sekundäre Therapie, Nachsorge	Direkte variable Kosten für Screening, Diagnose, primäre und sekundäre Therapie
Mengengerüst - Beschreibg.	Erstellte (Zwischen-)Produkte	Erstellte (Zwischen-)Produkte
Erhebung	Modellierung	Modellierung
Preise - Art Quelle	Selbstkostenrechnung Clinical Cost Manager des New England Medical Center	Administrative Preise kA ^b
Währung, Bezugsjahr	US-\$, 1992	US-\$, kA
Diskontierung	5%	--
Ergebnisse	Im folgenden Wertetripel für 50/60/70-Jährige: B) US-\$ 82/184/396 /ST ^{od}	Populationsbetrachtung: A) US-\$ 11.9Mrd. (n=21,187,000) ⇒ US-\$ 562 ^{er}
<i>Inkrementell</i>	1)US-\$ 64/149/371 /ST ^{od} 3)US-\$ 104/233/572 /ST ^{od} 2)US-\$ 77/179/448 /ST ^{od} 4)US-\$ 253/423/874 /ST ^{od}	
Effekte: Durchschnittlich	B) 27.8-12.7y ^{pd} bzw. 26.9-11.7QALY ^{od} /ST	Populationsbetrachtung: A) 0.65Mio.PCa mit 5,510TF B) 0.06Mio.PCa
<i>Inkrementell</i>	1) 0.0-0.2d ^{pd} /ST bzw. -1.8- -7.1QALY ^{od} /ST 2) 0.6-1.6d ^{pd} /ST bzw. -1.5- -7.2QALY ^{od} /ST 3) 0.6-1.7d ^{pd} /ST bzw. -2.1- -9.5 QALY ^{od} /ST 4) 0.8-2.2d ^{pd} /ST bzw. -2.9- -12.8QALY ^{od} /ST	
Kosten-Effekt-Relation: Durchschnittlich		
<i>Inkrementell</i>	2) US-\$ 113,000/127,000/189,000 /y 4) US-\$ 729,000/475,000/466,000 /y 1) und 3) durch 2) bzw. 4) dominiert Qualitätsadjustiert A) durch B) dominiert	
Behandlung Unsicherheiten	Alle Hauptvariablen einschließlich Wirksamkeit der PCa-RP, SV: PCa-ST vs. PCa-RP	PSA-Grenzwerte
Autorenschlussfolgerung	Screening nicht zu empfehlen	Weiterer Klärungsbedarf beim Screening
Bias/methodische Schwächen	2,3,4,6,7,9	3,4,(6),7,8,9

(Fortsetzung)

Autor, Jahr, Land	Launois (1992)⁸⁴ : Frankreich	Optenberg & Thompson (1990)¹¹⁰ : USA	
Evaluationstyp	Kosten-Effektivitäts-Analyse	Kosten-Konsequenzen-Analyse Kosten-Effektivitäts-Analyse	
Perspektive	Kostenträger	kA	
Alter Zielpopulation (in Jahren)	kA (Ergebnisse: ≥ 50)	50-70	
Setting	Prävalenz-Screening	Prävalenz-Screening Nationbezogen (n=17.5Mio.)	
Verglichene Alternativen (mit Biopsie-Kriterien)	A: 1) DRU→US,PSA ^f 5) DRU,PSA ^f →US 2) DRU,US 6) PSA ^f 3) DRU,PSA ^f 4) PSA ^f →DRU→US,PSA ^f B) Kein Screening bzw. keine PCa-RP	A: 1) DRU 2) TRUS 3) PSA4 4) PSA10 B) Status quo	
Wirkungsdimensionen/Zeithorizont	Anzahl diagnostizierter PCa; Restlebenserwartung bei PCa-RP	Anzahl diagnostizierter PCa; Anzahl Todesfälle in 5 Jahren; Überlebensjahre in 5 Jahren	
Klinische Effektdaten	Literatursynthese mit Modellierung (Entscheidungsbaumanalyse mit DEALE)	Literatursynthese mit Modellierung (Entscheidungsbaumanalyse mit aml. Sterbetafel)	
Nutzwerte/Nutzenwerte	--	--	
Kosten - Beschreibung	Direkte variable Kosten für Screening und Diagnose	Direkte variable Kosten für Screening, Diagnose, primäre und sekundäre Therapie	
Mengengerüst - Beschreibg.	Erstellte (Zwischen-)Produkte	Erstellte (Zwischen-)Produkte: CPT-4 / DRG	
Erhebung	Modellierung	Modellierung	
Preise - Art Quelle	Administrative Preise kA	Administrative Preise Civilian Health and Medical Program for the Uniform Services and Health Care Financing Administration (ProPSB)	
Währung, Bezugsjahr	FF, kA	US-\$, 1988	
Diskontierung	--	--	
Ergebnisse			
Kosten: Durchschnittlich	1) FF 166 /ST 2) FF 320 /ST 3) FF 301 /ST	4) FF 394 /ST 5) FF 424 /ST 6) FF 835 /ST	Pop.betrachtg., Nation in 5 J., n=17,496,288 (Angabe der Kosten/ST ^{erf} in US-\$) A: 1) US-\$ 3.8Mrd. 3) US-\$ 28.0Mrd. 2) US-\$ 23.6Mrd. 4) US-\$ 11.3Mrd. B) US-\$ 255Mio. ⇒ US-\$/ST ^{erf} : 1)219; 2)1347; 3)1598; 4)648
Inkrementell Effekte: Durchschnittlich	1) 0.01PCa /ST 2) 0.01PCa /ST 3) 0.01PCa /ST	4) 0.02PCa /ST 5) 0.02PCa /ST 6) 0.03PCa /ST	Populationsbetrachtung (Nation in 5 Jahren): A: 1) 0.2Mio.PCa-T; 0.08Mio.TF; 98.1Mio.y 2) 1.3Mio.PCa-T; 0.3Mio.TF; 97.3Mio.y 3) 1.6Mio.PCa-T; 0.4Mio.TF; 97.1Mio.y 4) 0.8Mio.PCa-T; 0.3Mio.TF; 97.5Mio.y B) 0.02Mio.PCa-T; 0.04Mio.TF; 98.1Mio.y
Inkrementell	0.18y /PCa-RP		
Kosten-Effekt-Relation: Durchschnittlich	1) FF 18,360 /PCa 2) FF 30,883 /PCa 3) FF 23,848 /PCa	4) FF 19,753 /PCa 5) FF 20,708 /PCa 6) FF 29,387 /PCa	A: 1) US-\$ 39 /y 3) US-\$ 288 /y 2) US-\$ 242 /y 4) US-\$ 116 /y B) US-\$ 2.6 /y
Inkrementell	1)vsB) FF 18,360 /PCa bzw. FF 102,230 /y 4)vs1) FF 20,905 /PCa bzw. FF 115,970 /y 6)vs4) FF 52,104 /PCa bzw. FF 288,110 /y 2), 3) und 5) dominiert durch 3), 1) bzw. 4)		
Behandlung Unsicherheiten	Prävalenz, Sensitivität, Spezifität, Korrektur des work-up bias: 6) dominiert	Sensitivität und Spezifität, 5-Jahres-Überlebensraten	
Autorenschlussfolgerung	Invasive Diagnostik bei hoher PCa-P denkbar	Screening nicht zu empfehlen	
Bias/methodische Schwächen	2,3,(4),6,7,8,9	3,4,6,7,9,10,11,12	

(Fortsetzung)

Autor, Jahr, Land	Littrup et al. (1993) ⁹¹ : USA	Littrup et al. (1994c) ⁹³ : USA
Evaluationstyp	Kosten-Effektivitäts-Analyse (KEA) Kosten-Nutzen-Analyse (KNA)	Kosten-Effektivitäts-Analyse (KEA) Kosten-Nutzen-Analyse (KNA)
Perspektive	kA	kA
Alter Zielpopulation (in Jahren)	55-70	55-70
Setting	Prävalenz- und Inzidenz-Screening Ambulante KH-Versorgung / Rekrutierung über öffentliche Kampagne: n=2,425	Prävalenz-Screening Ambulante KH-Versorgung / Rekrutierung über öffentliche Kampagne: n=2,900
Verglichene Alternativen (mit Biopsie-Kriterien)	A: 1) (PSA4),DRU,TRUS 2) TRUS 6) PSA4,DRU 3) DRU 7) PSA2 4) PSA2,DRU 8) PSA3 5) PSA3,DRU 9) PSA4 B: --	A: 1) PSA4,DRU 6) PSA ^{Alter^h} 2) PSA4,DRU,PSA-D 3) DRU,PSA-D 7) PSA ^{Alter^h} ,PSA-D 4) PSA3,PSA-D 8) PSA ^{Alter^h} ,DRU 5) PSA4,PSA-D 9) PSA ^{Alter^h} ,DRU,PSA-D B: --
Wirkungsdimensionen/Zeithorizont	Anzahl diagnostizierter PCa pro Jahr und in 3 Jahren	Anzahl Biopsien und Anzahl diagnostizierter PCa in bis zu 5 Jahren
Klinische Effektdaten	Originaldaten (Longitudinalstudie ^g) mit Modellierung (Entscheidungsbaumanalyse)	Originaldaten (Longitudinalstudie ^g) mit Modellierung (Entscheidungsbaumanalyse)
Nutzwerte/Nutzenwerte	Nutzenwerte der KNA: Literatur, Annahmen	Nutzenwerte der KNA: Literatur, Annahmen
Kosten - Beschreibung	KEA: Direkte variable Kosten für Screening und Diagnose / KNA: Direkte Kosten für Screening, Diagnose und Therapie, Produktivitätskosten, intangible Kosten	KEA: Direkte variable Kosten für Screening und Diagnose / KNA: Direkte Kosten für Screening, Diagnose und Therapie, Produktivitätskosten, intangible Kosten
Mengengerüst - Beschreibg.	Erstellte (Zwischen-)Produkte	Erstellte (Zwischen-)Produkte
Erhebung	Modellierung, Annahmen	Modellierung, Annahmen
Preise - Art Quelle	Administrative Preise, Annahmen Health Care Finance Administration, Literatur, Annahmen	Administrative Preise, Annahmen Health Care Finance Administration, Literatur, Annahmen
Währung, Bezugsjahr	US-\$, kA	US-\$, kA
Diskontierung	--	--
Ergebnisse		
Kosten: Durchschnittlich		KEA: US-\$ 101 /ST bei 7) bis US-\$ 219 /ST bei 1)
<i>Inkrementell</i>		
Effekte: Durchschnittlich	KEA: 3-J.Durchschnitt: 0.041 PCa/ST bei 3) bis 0.068 PCA/ST bei 1)	KEA: 0.025 PCa/ST bei 3) bis 0.046 PCa/ST bei 1)
<i>Inkrementell</i>	KEA: 1. Jahr: 0.027PCa /ST bei 3) u. 9) bis 0.042PCa /ST bei 1) 2. Jahr: 0.013PCa /ST bei 3) bis 0.020PCa /ST bei 1) und 4) 3. Jahr: 0.002PCa /ST bei 3) bis 0.006PCa /ST bei 1),4),5),6)	
Kosten-Effekt-Relation: Durchschnittlich	KEA: 3-J.Durchschnitt: US-\$ 3,415 /PCa bei 3) bis US-\$ 11,135 /PCa bei 1) KNA: 3) US-\$ -474 (=Nettokosten) 4) US-\$ -5,544 7) US-\$ -5,216 5) US-\$ -2,936 8) US-\$ -2,473 6) US-\$ -1,646 9) US-\$ -1,126	KEA: US-\$ 3,144 /PCa bei 7) bis US-\$ 4,752 /PCa bei 1) KNA: 1) US-\$ -2,663 (=Nettokosten) 2) US-\$ -1,386 6) US-\$ -1,141 3) nicht untersucht 7) US-\$ -1,093 4) US-\$ -1,149 8) US-\$ -2,259 5) US-\$ -1,037 9) US-\$ -952
<i>Inkrementell</i>	KEA: 1. Jahr: US-\$ 2,178 /PCa bei 3) bis US-\$ 6,556 /PCa bei 1) 2. Jahr: US-\$ 3,693 /PCa bei 3) bis US-\$ 12,562 /PCa bei 1) 3. Jahr: US-\$ 12,079 /PCa bei 9) bis US-\$ 37,588 /PCa bei 1)	
Behandlung Unsicherheiten	Durchschnittspreise, Spezifität, Nutzen der Frühbehandlung, Krankheitsprävalenz	Preise für DRU, PSA-Assay und Biopsie
Autorenschlussfolgerung	Kosten-effektive Strategien möglich	Kosten-effektive Strategien möglich
Bias/methodische Schwächen	1,(2),(3),4,5,11,12	1,(3),4,5,10,11,12

(Fortsetzung)

Legende zu Bias/methodische Schwächen:

- 1 Nicht randomisierte Rekrutierung (selection bias)
- 2 Keine explizite Definition der Studien- bzw. Zielpopulation
- 3 Keine hinreichende Definition von anormalen Testkriterien
- 4 Unvollständige Anwendung des Referenz-(Gold-)Standards auf Studienpopulation (work-up bias)
- 5 Keine angemessene Verblindung bei Testauswertung bzw. Diagnosestellung (diagnostic suspicion bias, expectation bias)
- 6 Fehlende Angaben zu Methoden der Datenidentifikation
- 7 Fehlende Angaben zu Methoden der Datenextraktion einschließlich der Bewertung von Relevanz und Validität bzw. Qualität
- 8 Fehlende Angaben zu Methoden der Informationssynthese
- 9 Keine hinreichende Beschreibung der herangezogenen Literaturdaten hinsichtlich der Studienpopulation und Intervention
- 10 Inkongruenzen zwischen Text und Tabellen bzw. Abbildungen
- 11 Nachvollziehbarkeit der Effektermittlung (teilweise) nicht gegeben
- 12 Nachvollziehbarkeit der Kostenermittlung (teilweise) nicht gegeben bzw. fehlerhafte Kostenberechnung

(): zum Teil erfüllt (bei Behandlung von Unsicherheiten oder durch Verweis auf Originalpublikationen) bzw. bei Behandlung von Unsicherheiten auftretend

A	untersuchte Technologien
B	Vergleichstechnologie
CPT-4	Current Procedural Terminology, 4 th edition
d	Tage Restlebenserwartung
DEALE	Declining Exponential Approximation of Life Expectancy Method
DRG	Diagnosis-related Groups
er	errechnet aus den Angaben in der Publikation
G-Sc	Gleason-Score (teils mit Angabe des Scorewertes)
J	Jahre
kA	keine Angaben
KEA	Kosten-Effektivitäts-Analyse
KHK	koronare Herzerkrankung
KNA	Kosten-Nutzen-Analyse
n	Anzahl Screening-Teilnehmer
oD	ohne Diskontierung
PCa	Prostatakarzinom
PCa-P	Prävalenz von Prostatakarzinom
PCa-RP	radikale Prostatektomie bei Prostatakarzinom
PCa-ST	Strahlentherapie bei Prostatakarzinom
PCa-T	Therapie bei Prostatakarzinom
ProPSB	Prospective Payment System Bill
PSA4	Entscheidungsgrenze bei einem PSA-Wert von 4ng/ml
PSA-D	PSA-Dichte
RBRVS	Resource Based Relative Value Score (Medicare)
ST	Screening-Teilnehmer
SV	Strukturvariante
TF	Todesfälle (behandlungsbedingt)
y	Jahre Restlebenserwartung

^a Mittlere Preiskategorie

^b In Anlehnung an die Studie von Optenberg & Thompson (1990)

^c Screening und Diagnose

^d Ohne behandlungsbedingte Todesfälle

^e Exzess-PSA-Wert = berechneter volumenadjustierter PSA-Wert (ng/ml) - gemessener PSA-Wert (ng/ml)

Volumenadjustierter PSA-Wert = 0.12ng/ml/ccm x Prostatavolumen (ccm)

Exzess-PSA10 = gemessener PSA-Wert bei 10ng/ml über berechnetem volumenadjustiertem PSA-Wert

^f Entscheidungsgrenze für PSA-Wert unklar: 5ng/ml oder 10ng/ml

^g American Cancer Society National Prostate Cancer Detection Project (ACS-NPCDP): Strategie 1) ursprünglicher Interventionspfad

^h Altersabhängiger PSA-Grenzwert = 2.5ng/ml bei 40- bis 49jährigen, 3.5ng/ml bei 50- bis 59jährigen, 4.5ng/ml bei 60- bis 69jährigen, 6.5ng/ml bei ≥70jährigen

C.5.4 Bewertung der Information/Studienqualität

C.5.4.1 Deskription

Die für jede Studie ausgefüllten Deskriptionskataloge können als rasterartige Abbildung der systematisch erfassten Einzelinformationen gesehen werden. Wegen des Umfangs dieses Erfassungsinstruments wurde darauf verzichtet, die ausgefüllten Kataloge im Anhang dieser Arbeit wiederzugeben; sie können bei den Autoren dieses Berichtes angefordert werden. Für einen Überblick über die wesentlichen deskriptiven Charakteristika der einzelnen Studien, wie sie u.a. mit Unterstützung des Deskriptionskataloges als objektive Informationsgrundlage für den Bewertungsprozess identifiziert wurden, sei auf Tabelle 15 verwiesen.

C.5.4.2 Transparenz

Tabelle 16 beschreibt das detaillierte Abschneiden der Studien bei der Transparenzbewertung. In den Spalten wurden die Punktwerte für jedes der 56 Transparenzkriterien (siehe Transparenzkatalog) studienweise aufgelistet. Zeilenweise wurde der durchschnittliche prozentuale Erfüllungsgrad der einzelnen Transparenzkriterien (% der max. Punktzahl) ermittelt. Bei allen Prozentberechnungen wurden nicht relevante Items aus dem Nenner ausgeschlossen.

Tabelle 16: Beurteilung der PSA-Studien nach dem Erfüllungsgrad der Transparenzkriterien

Transparenzkriterien	Abramson et al., 1994	Barry et al., 1995	Benoit & Naslund, 1994	Chadwick et al., 1991	Ciatto et al., 1995	de Biasi et al., 1996	Dorr et al., 1993	Gottlieb et al., 1996	Gustafs-son et al., 1995	Kantrowitz et al., 1995	Krahn et al., 1994	Kramer et al., 1993	Launois, 1992	Littrup et al., 1993	Littrup et al., 1994c	Opten-berg & Thomp-son, 1990	% der max. Punktzahl
Fragestellung																	
1.	1	1	1	1	1	1	1/2	1	1	1	1	1	1	1	1	1	97%
2.	1/2	1	1/2	1	1	0	1/2	1/2	1	1/2	1	1	1	1	1/2	1	75%
Evaluationsrahmen																	
3.	1/2	1	1/2	1	1	1	1/2	1/2	1	1/2	1	1/2	1	1	1	1/2	78%
4.	0	1	1	0	1	0	0	1	1	0	1	0	1	1	1	1	63%
5.	nr	1	0	nr	0	nr	0	1	1	nr	1	0	0	0	1	0	42%
6.	1	1	1	1	1	1	1/2	1/2	1	1	1	1	0	1/2	1	1	84%
7.	1	1/2	1	1/2	1	1	0	1	1/2	1	1/2	1	1	1/2	1	1	78%
8.	0	1	1	0	0	0	0	1	1	0	1	0	1	1	0	0	44%
9.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100%
10.	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	25%
Analysemethoden und Modellierung																	
11.	1/2	1	1/2	1/2	1	1	1	1	1	1/2	1	1	1	1	1	1	88%
12.	1/2	1	1/2	1/2	1	1	1	1/2	1	1/2	1	0	1	1/2	1/2	1	72%
13.	0	1	1	0	0	0	0	1	1	0	1	1	1	1	1/2	1	59%
14.	nr	1	nr	nr	nr	nr	0	1/2	1	nr	1	1/2	1	1	1	1	80%
Gesundheitseffekte																	
15.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100%
16.	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	94%
17.	1/2	nr	nr	1	1	1	nr	nr	1	1/2	nr	nr	nr	0	1	nr	75%
18.	nr	1	1/2	nr	nr	nr	1/2	0	nr	nr	1	0	0	nr	nr	0	38%
19.	0	0	0	0	0	0	0	1/2	0	0	1	0	0	1/2	0	0	13%
20.	nr	nr	nr	nr	nr	nr	nr	1	nr	nr	1	nr	nr	1	nr	nr	100%
21.	0	1	0	0	0	0	0	0	0	0	1	1	0	0	1/2	0	22%
Kosten																	
22.	1/2	1/2	1	1	1/2	1	1/2	1/2	1	1/2	1/2	1/2	1	0	1/2	1	66%
23.	1/2	1	1	1/2	1/2	1/2	1/2	1/2	1	1/2	1/2	1/2	1	0	1/2	1	63%
24.	1/2	1	1	1	1/2	1	1	1/2	1	1/2	1/2	1/2	1	1	0	1	75%
25.	1	1	1	0	0	0	0	1/2	1	1/2	1/2	0	1/2	0	0	1	44%
26.	0	1	1/2	0	0	1/2	0	1	0	0	0	0	0	0	0	0	19%
27.	nr	nr	nr	nr	nr	nr	nr	nr	1/2	nr	nr	nr	nr	1/2	nr	nr	50%
28.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100%
29.	nr	nr	nr	nr	nr	nr	nr	nr	1	nr	nr	nr	nr	nr	nr	nr	100%
30.	nr	0	nr	nr	0	nr	0	0	0	nr	0	nr	nr	0	nr	0	0%

(Fortsetzung)

Fortsetzung

Transparenz- kriterien	Abramson et al., 1994	Barry et al., 1995	Benoit & Naslund, 1994	Chadwick et al., 1991	Ciatto et al., 1995	de Biasi et al., 1996	Dorr et al., 1993	Gottlieb et al., 1996	Gustaf- son et al., 1995	Kantrowitz et al., 1995	Krahn et al., 1994	Kramer et al., 1993	Launois, 1992	Littrup et al., 1993	Littrup et al., 1994c	Opten- berg & Thomp- son, 1990	% der max. Punktzahl
Diskontierung																	
31.	nr	1	nr	nr	nr	nr	0	0	nr	nr	1	nr	0	0	0	0	25%
32.	1	1	1	0	0	0	0	0	1	0	1	0	0	0	0	1	38%
33.	nr	1	nr	nr	nr	nr	nr	nr	nr	nr	1	nr	nr	nr	nr	nr	100%
34.	nr	0	nr	nr	nr	nr	0	0	nr	nr	0	nr	0	0	0	0	0%
Ergebnispräsentation																	
35.	nr	0	nr	nr	nr	nr	0	0	0	nr	0	0	0	0	0	0	0%
36.	1	1/2	1	1/2	1	1	0	1	1	1/2	1	0	1	1	1	0	72%
37.	0	1	0	0	1/2	0	0	1	1	0	1	0	1	1/2	0	0	38%
38.	1	1	1	1	1	1	1	1	1	1	1	nr	1	1	1	1	100%
39.	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	94%
40.	1	1	1	1	1	1	1	0	1	1	1/2	1	1	1	1/2	1	88%
41.	0	1	0	1/2	0	0	1	0	1	0	0	1	0	0	1	1	41%
Behandlung von Unsicherheiten																	
42.	0	1	1	0	0	0	0	1	0	0	1	1	1	1	1/2	1	53%
43.	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6%
44.	0	1	1	0	1	1/2	0	0	0	0	1	1	0	0	0	0	34%
45.	nr	1/2	1	nr	nr	1	nr	1/2	nr	nr	1	1	1	1/2	1	1	85%
46.	nr	1	1	nr	nr	1	nr	1/2	nr	nr	1/2	1	1	1/2	1/2	1	80%
47.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%
Diskussion																	
48.	0	1	0	0	0	0	0	1	0	0	1	1	1	1/2	0	1	41%
49.	0	1/2	1	1/2	1/2	1/2	1	1	0	1/2	1/2	1/2	1/2	1/2	1/2	1	56%
50.	0	1	1	0	1/2	1	0	0	0	1/2	1/2	1/2	0	0	0	1/2	31%
51.	1/2	1	1	1	1/2	1	1	1	1	1/2	1	1	0	1	1	1	84%
52.	0	1/2	1	1/2	0	1	0	0	1	1/2	1/2	1/2	0	1	1	1/2	50%
53.	0	0	0	1/2	0	1	0	0	0	0	0	1/2	0	0	0	0	13%
54.	0	1	1	0	0	1	1	0	0	0	1	1/2	1/2	1	0	0	44%
Schlussfolgerungen																	
55.	1/2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1/2	94%
56.	1/2	1	1	1	1/2	1	1/2	1	1	1	1	1	1	1/2	1	1	88%

Numerierung der Transparenzkriterien entspricht der Fragenfolge des Transparenzkataloges

Die Einzelanalyse der Transparenzkriterien im Querschnitt aller Studien ergibt eine Variationsbreite des Erfüllungsgrades von 0% bis 100%. Konsistent nicht erfüllt wurden die Kriterien bezüglich:

- Angaben zu Preisanpassungen bei Inflation bzw. Deflation (Kriterium 30)
- Begründungen der Wahl der Diskontraten bzw. des Verzichts auf eine Diskontierung (Kriterium 34)
- Angaben zu Maßnahmen der Modellvalidierung (Kriterium 35)
- Angaben zum Einsatz von statistischen Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten bei der Behandlung von Unsicherheiten (Kriterium 47)

Durchschnittlich mangelhaft erfüllt (<20%) wurden die Transparenzkriterien bezüglich:

- der Präferenzbewertung von Gesundheitszuständen und deren methodische Beschreibung (Kriterium 19)
- Begründung der in die Analyse einbezogenen Kosten (Kriterium 26)
- Durchführung multivariater Sensitivitätsanalysen (Kriterium 43)
- Diskussion ethischer Aspekte und Fragen der Verteilungsgerechtigkeit (Kriterium 53)

Vollständig von allen Studien erfüllt wurden dagegen, soweit relevant, die Kriterien bezüglich:

- Untersuchung sowohl von Kosten als auch von Gesundheitseffekten (Kriterium 9)
- Aufführung der in der Studie untersuchten Gesundheitszustände (Kriterium 15)
- Spezifikation der Datenquellen für die Bewertung von Gesundheitszuständen (Kriterium 20)- dabei ist dieses Kriterium nur in drei Fällen relevant
- Angabe der Währung (Kriterium 28)
- Angabe zu Währungskonversionen (Kriterium 29) - dabei ist dieses Kriterium nur in einem Fall relevant

- Angabe der Diskontrate(n) (Kriterium 33) - dabei ist dieses Kriterium nur in zwei Fällen relevant.

Ebenfalls wird in annähernd allen 16 Studien eine Maßzahl für die Relation zwischen Kosten und Gesundheitseffekt angegeben (Kriterium 38), werden in 15 von 16 Studien rohe, nicht qualitätsbereinigte klinische Effekte berichtet (Kriterium 39), werden in allen 16 Studien mehr oder weniger Schlussfolgerungen aus den berichteten Daten bzw. Ergebnissen abgeleitet (Kriterium 55) und werden in 15 von 16 Studien Angaben zu den Quellen für die Effektdaten gemacht (Kriterium 16).

Tabelle 17 aggregiert die Einzelitems der 11 im Transparenzkatalog festgelegten Transparenzdimensionen und fasst diese studienweise (Spalten) zu einem prozentualen Gesamtscore (Transparenzscore) zusammen, bei dem die einzelnen Transparenzdimensionen, ungeachtet der Anzahl ihrer Einzelitems, gleich gewichtet werden. Zusätzlich wurde der prozentuale Erfüllungsgrad der einzelnen Transparenzdimensionen im Mittel aller Studien ermittelt (Zeilen).

Tabelle 17: Beurteilung der PSA-Studien nach dem Erfüllungsgrad der Transparenzdimensionen

Transparenzdimensionen	Abramson et al., 1994	Barry et al., 1995	Benoit & Naslund, 1994	Chadwick et al., 1991	Ciatto et al., 1995	de Biasi et al., 1996	Dorr et al., 1993	Gottlieb et al., 1996	Gustafsson et al., 1995	Kantrowitz et al., 1995	Krahn et al., 1994	Kramer et al., 1993	Launois, 1992	Littrup et al., 1993	Littrup et al., 1994c	Optenberg & Thompson, 1990	
Fragestellung	75%	100%	75%	100%	100%	50%	50%	75%	100%	75%	100%	100%	100%	100%	75%	100%	86%
Evaluationsrahmen	50%	94%	69%	50%	63%	57%	25%	88%	81%	50%	94%	44%	75%	63%	75%	56%	65%
Analysemethoden und Modellierung	33%	100%	67%	33%	67%	67%	50%	75%	100%	33%	100%	63%	100%	88%	75%	100%	72%
Gesundheitseffekte	50%	80%	50%	60%	60%	60%	30%	58%	60%	50%	100%	60%	40%	58%	70%	40%	58%
Kosten	58%	79%	92%	58%	36%	67%	43%	57%	72%	50%	43%	42%	75%	31%	33%	71%	57%
Diskontierung	100%	75%	100%	0%	0%	0%	0%	0%	100%	0%	75%	0%	0%	0%	0%	33%	30%
Ergebnispräsentation	67%	79%	67%	67%	75%	67%	57%	43%	86%	58%	64%	50%	71%	64%	64%	57%	65%
Behandlung von Unsicherheiten	0%	75%	67%	0%	25%	42%	0%	33%	0%	0%	58%	67%	50%	33%	33%	50%	33%
Diskussion	7%	71%	71%	36%	21%	79%	43%	43%	29%	21%	64%	64%	29%	57%	36%	57%	46%
Schlussfolgerungen	50%	100%	100%	100%	75%	100%	75%	100%	100%	100%	100%	100%	100%	75%	100%	75%	91%
Transparenzscore	49%	85%	76%	50%	52%	59%	37%	57%	73%	44%	80%	59%	64%	57%	56%	64%	60%
Rang	14	1	3	13	12	8	16	9	4	15	2	7	6	10	11	5	

Der durchschnittliche Transparenzscore beträgt 60% mit einer Standardabweichung von 12%, wobei sich eine individuelle Variationsbreite von 37% bis 85% feststellen lässt. Den niedrigsten Scorewert erzielen dabei die Studie von Dorr et al. (1993) mit 37%, die Studie von Kantrowitz et al. (1995) mit 44% und diejenigen von Ciatto et al. (1995) mit 52%. Der Transparenzscore ist am höchsten für die Untersuchungen von Barry et al. (1995), von Krahn et al. (1994) und von Benoit & Naslund (1994), er liegt bei 85%, 80% und 76% in der genannten Reihenfolge.

Die Variationsbreite des durchschnittlichen Erfüllungsgrades der Transparenzdimensionen reicht von 33% bis 91%. Deutliche Schwächen hinsichtlich der Transparenz und Nachvollziehbarkeit lassen sich im Durchschnitt der Studien in folgenden Bereichen erkennen (in Klammern der durchschnittliche prozentuale Erfüllungsgrad):

- bei der Behandlung von Unsicherheiten (33%) - die Variationsbreite über die Studien hinweg liegt zwischen 0% und 75%
- bei der Diskontierung (30%) - die Variationsbreite über die Studien hinweg liegt zwischen 0% und 100%
- bei der Diskussion (46%) - die Variationsbreite über die Studien hinweg liegt zwischen 7% und 79%.

Ein deutlich höheres Ausmaß an Transparenz und Nachvollziehbarkeit findet sich dagegen in den Bereichen:

- Schlussfolgerungen (91%) - die Variationsbreite über die Studien hinweg liegt zwischen 50% und 100%
- Analysemethoden und Modellierung (72%) - die Variationsbreite über die Studien hinweg liegt, wenn relevant, zwischen 33% und 100%
- Fragestellung (86%) - die Variationsbreite über die Studien hinweg liegt zwischen 50% und 100%

C.5.4.3 Methodisch-inhaltliche Qualität

Tabelle 18 beschreibt das detaillierte Abschneiden der Studien bei der Qualitätsbewertung. In den Spalten wurden die Punktwerte für jedes der 56 Bewertungskriterien (siehe Qualitätskatalog) studienweise aufgelistet. Zeilenweise

wurde der durchschnittliche prozentuale Erfüllungsgrad der einzelnen Bewertungskriterien (% der max. Punktzahl) ermittelt. Bei allen Prozentberechnungen wurden nicht relevante Items aus dem Nenner ausgeschlossen.

Tabelle 18: Beurteilung der PSA-Studien nach dem Erfüllungsgrad der Qualitätsbewertungskriterien

Qualitätskriterien	Abramson et al., 1994	Barry et al., 1995	Benoit & Naslund, 1994	Chadwick et al., 1991	Ciatto et al., 1995	de Biasi et al., 1996	Dorr et al., 1993	Gottlieb et al., 1996	Gustafsson et al., 1995	Kantrowitz et al., 1995	Krahn et al., 1994	Kramer et al., 1993	Launois, 1992	Littrup et al., 1993	Littrup et al., 1994c	Optenberg & Thompson, 1990	% der max. Punktzahl	
Fragestellung																		
1.	1/2	1/2	1/2	1/2	1/2	1	0	1	1	1	1/2	1/2	1/2	1/2	1/2	1/2	59%	
2.	0	1	1/2	1/2	1/2	0	1/2	1/2	1	0	1	1	1/2	1	1/2	1/2	56%	
Evaluationsrahmen																		
3.	0	1/2	0	1/2	1/2	1/2	0	0	1	0	1/2	0	1/2	0	0	1/2	28%	
4.	0	1	0	1/2	0	1/2	0	1	1	0	1	1/2	1	1/2	1	1/2	53%	
5.	nr	1	0	nr	0	nr	0	1/2	1/2	nr	1	0	0	0	1	0	33%	
6.	1	1	1	1	1	1	1/2	0	1	1	1/2	1	0	1/2	1	1	78%	
7.	0	1/2	0	0	0	1	0	0	0	0	1/2	1/2	1/2	1/2	0	1/2	25%	
8.	0	1	1	0	0	0	0	1	1	0	1	0	1	1	0	0	44%	
9.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100%	
10.	0	1	0	0	0	0	0	1	0	0	1	0	1	0	0	0	25%	
Analysemethoden und Modellierung																		
11.	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1	1/2	1	0	1/2	0	1/2	1	53%	
12.	1/2	1	1/2	1/2	1/2	1/2	1	0	1	1/2	1	0	1/2	1/2	0	1	56%	
13.	0	1	1/2	0	0	0	0	1	1	0	1	1/2	1/2	0	0	1	41%	
14.	nr	1/2	nr	nr	nr	nr	0	0	1/2	nr	1/2	1/2	1/2	1/2	1/2	1/2	40%	
Gesundheitseffekte																		
15.	1/2	1	0	1/2	1/2	1/2	0	0	1/2	1/2	1	1/2	0	0	0	1	41%	
16.	1	1/2	0	1	1	1	0	1/2	1/2	1	1/2	1/2	1/2	1/2	1/2	1/2	59%	
17.	0	nr	nr	0	0	0	nr	nr	1/2	0	nr	nr	nr	0	1/2	nr	13%	
18.	nr	1/2	0	nr	nr	nr	0	0	nr	nr	1/2	0	0	nr	nr	1/2	19%	
19.	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	6%	
20.	nr	nr	nr	nr	nr	nr	nr	0	nr	nr	1/2	nr	nr	0	nr	nr	17%	
21.	0	1/2	0	0	0	0	0	0	0	0	1/2	1/2	0	0	0	0	9%	
Kosten																		
22.	1/2	1/2	1/2	1	1/2	1/2	1/2	1/2	1	1/2	1/2	1/2	1/2	1	0	1/2	1	59%
23.	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	1/2	0	1/2	1	50%
24.	1/2	1/2	1	1	1/2	1	1/2	1/2	1	1/2	1/2	1/2	1	1/2	0	1	66%	
25.	0	1	1/2	0	0	0	0	0	1/2	0	1/2	0	1/2	0	0	1/2	22%	
26.	0	1/2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3%	
27.	nr	nr	nr	nr	nr	nr	nr	nr	1/2	nr	nr	nr	nr	1/2	nr	nr	50%	
28.	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	100%	
29.	nr	nr	nr	nr	nr	nr	nr	nr	1	nr	nr	nr	nr	nr	nr	nr	100%	
30.	nr	0	nr	nr	0	nr	0	0	0	nr	0	nr	nr	0	nr	0	0%	

(Fortsetzung)

Fortsetzung

Bewertungs- kriterien	Abramson et al., 1994	Barry et al., 1995	Benoit & Naslund, 1994	Chadwick et al., 1991	Ciatto et al., 1995	de Biasi et al., 1996	Dorr et al., 1993	Gottlieb et al., 1996	Gustafs- son et al., 1995	Kantrowitz et al., 1995	Krahn et al., 1994	Kramer et al., 1993	Launois, 1992	Littrup et al., 1993	Littrup et al., 1994c	Opten- berg & Thomp- son, 1990	% der max. Punktzahl
Diskontierung																	
31.	nr	1	nr	nr	nr	nr	0	0	nr	nr	1	nr	0	0	0	0	25%
32.	1	1	1	0	0	0	0	0	1	0	1	0	0	0	0	1	38%
33.	nr	1	nr	nr	nr	nr	nr	nr	nr	nr	1	nr	nr	nr	nr	nr	100%
34.	nr	0	nr	nr	nr	nr	0	0	nr	nr	0	nr	0	0	0	0	0%
Ergebnispräsentation																	
35.	nr	0	nr	nr	nr	nr	0	0	0	nr	0	0	0	0	0	0	0%
36.	1	1/2	1	1/2	1	1	0	1/2	1	1/2	1	0	1	1	1	0	69%
37.	0	1	0	0	1/2	0	0	1/2	1	0	1	0	1	1/2	0	0	34%
38.	1	1	0	1	1	1	1/2	1	1	1/2	1	nr	1	1	1	0	80%
39.	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	94%
40.	1/2	1	1	1	1/2	1	1/2	0	1	1/2	1/2	1/2	1	1/2	0	1	66%
41.	0	1	0	1/2	0	0	1	0	1	0	0	1	0	0	1/2	1	38%
Behandlung von Unsicherheiten																	
42.	0	1	1/2	0	0	0	0	1	0	0	1	0	1/2	1	1/2	1/2	38%
43.	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6%
44.	0	1/2	1/2	0	0	1/2	0	0	0	0	1	1/2	0	0	0	0	19%
45.	nr	1/2	1/2	nr	nr	1	nr	1/2	nr	nr	1	1	1	1/2	1	0	70%
46.	nr	1/2	1/2	nr	nr	1	nr	0	nr	nr	1/2	1/2	1/2	1/2	0	1/2	45%
47.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0%
Diskussion																	
48.	0	1	0	0	0	0	0	1/2	0	0	1/2	1/2	1/2	0	0	1/2	22%
49.	0	1/2	1/2	1/2	1/2	1/2	1/2	1/2	0	0	1/2	1/2	1/2	1/2	1/2	1	44%
50.	0	1/2	1/2	0	1/2	1/2	0	0	0	0	1/2	1/2	0	0	0	1/2	22%
51.	0	1	1/2	1/2	1/2	1/2	1/2	1	1/2	0	1	1	0	1/2	1/2	1/2	53%
52.	0	1/2	1/2	0	0	0	0	0	1	1/2	1/2	1/2	0	1/2	1/2	0	28%
53.	0	0	0	1/2	0	0	0	0	0	0	0	1/2	0	0	0	0	6%
54.	0	1	1/2	0	0	0	1/2	0	0	0	1/2	0	0	0	0	0	16%
Schlussfolgerungen																	
55.	1/2	1	0	1	1	0	1	1	1	1/2	1	1	1/2	1/2	1/2	1/2	69%
56.	1/2	1/2	1/2	1/2	1/2	1/2	0	1	1	1/2	1/2	1/2	1/2	0	1/2	1/2	50%

Numerierung der Bewertungskriterien entspricht der Fragenfolge des Qualitätskatalogs

Die Einzelanalyse der Qualitätskriterien im Querschnitt aller Studien ergibt eine Variationsbreite des Erfüllungsgrades von 0% bis 100%. Konsistent nicht erfüllt wurden die Kriterien bezüglich:

- adäquate Preisanpassungen bei Inflation oder Deflation (Kriterium 30)
- plausible Gründe für die Wahl der Diskontraten bzw. den Verzicht auf eine Diskontierung (Kriterium 34)
- Angaben zu Maßnahmen der Modellvalidierung (Kriterium 35)
- der Einsatz von adäquaten statistischen Inferenzmethoden (statistische Tests, Konfidenzintervalle) für stochastische Daten bei der Behandlung von Unsicherheiten (Kriterium 47)

Durchschnittlich mangelhaft erfüllt (<20%) wurden die Qualitätskriterien bezüglich:

- adäquater Methodenwahl bzw. Beschreibung und Ergebnisdarstellung (Kriterien 17-21)
- Vollständigkeit und schlüssige Begründung der einbezogenen Kosten (Kriterium 26)
- Umfang und Art der durchgeführten Sensitivitätsanalysen (Kriterien 43, 44)
- Diskussion wichtiger ethischer Aspekte und Fragen der Verteilungsgerechtigkeit (Kriterium 53)
- sinnvolle Diskussion des Ergebnisses im Kontext mit anderen unabhängigen Gesundheitsprogrammen (Kriterium 54).

Vollständig von allen Studien erfüllt wurden dagegen, soweit relevant, die Bewertungskriterien bezüglich:

- Untersuchung sowohl von Kosten als auch von Gesundheitseffekten (Kriterium 9)
- Angabe der Währung (Kriterium 28)
- Angabe der Diskontrate(n) (Kriterium 33) - dabei ist dieses Kriterium in zwei Fällen relevant.

Tabelle 19 aggregiert die Einzelitems der 11 im Qualitätskatalog festgelegten Bewertungsdimensionen und fasst diese studienweise (Spalten) zu einem prozentualen Qualitätsscore zusammen. Dabei werden alle Bewertungsdimensionen, ungeachtet der Anzahl ihrer Einzelitems, gleich gewichtet. Zusätzlich wurde der prozentuale Erfüllungsgrad der einzelnen Bewertungsdimensionen im Mittel aller Studien ermittelt (Zeilen).

Tabelle 19: Beurteilung der PSA-Studien nach dem Erfüllungsgrad der Bewertungsdimensionen

Qualitätsdimensionen	Abramson et al., 1994	Barry et al., 1995	Benoit & Naslund, 1994	Chadwick et al., 1991	Ciatto et al., 1995	de Biasi et al., 1996	Dorr et al., 1993	Gottlieb et al., 1996	Gustafsson et al., 1995	Kantrowitz et al., 1995	Krahn et al., 1994	Kramer et al., 1993	Launois, 1992	Littrup et al., 1993	Littrup et al., 1994c	Optenberg & Thompson, 1990	
Fragestellung	25%	75%	50%	50%	50%	50%	25%	75%	100%	50%	75%	75%	50%	75%	50%	50%	58%
Evaluationsrahmen	29%	88%	38%	43%	31%	57%	19%	56%	69%	29%	81%	38%	63%	44%	50%	44%	48%
Analysemethoden und Modellierung	33%	75%	50%	33%	33%	33%	38%	38%	88%	33%	88%	25%	50%	25%	25%	88%	47%
Gesundheitseffekte	30%	50%	0%	30%	30%	30%	0%	8%	30%	30%	67%	30%	10%	8%	20%	40%	26%
Kosten	42%	57%	58%	58%	36%	50%	36%	36%	61%	42%	43%	42%	67%	25%	33%	64%	47%
Diskontierung	100%	75%	100%	0%	0%	0%	0%	0%	100%	0%	75%	0%	0%	0%	0%	33%	30%
Ergebnispräsentation	58%	79%	50%	67%	67%	67%	43%	29%	86%	42%	64%	42%	71%	57%	50%	43%	57%
Behandlung von Unsicherheiten	0%	58%	33%	0%	0%	42%	0%	25%	0%	0%	58%	33%	33%	33%	25%	17%	22%
Diskussion	0%	64%	36%	21%	21%	21%	21%	29%	21%	7%	50%	50%	14%	21%	21%	36%	27%
Schlussfolgerungen	50%	75%	25%	75%	75%	25%	50%	100%	100%	50%	75%	75%	50%	25%	50%	50%	59%
Qualitätsscore	37%	70%	44%	38%	34%	38%	23%	39%	65%	28%	68%	41%	41%	31%	32%	46%	42%
Rang	11	1	5	9	12	10	16	8	3	15	2	6	7	14	13	4	

Der durchschnittliche Qualitätsscore beträgt 42% mit einer Standardabweichung von 13%, wobei sich eine individuelle Variationsbreite von 23% bis 70% feststellen lässt. Den niedrigsten Gesamtscore erzielen dabei die Studie von Dorr et al. (1993) mit 23%, die Studie von Kantrowitz et al. (1995) mit 28% und die Studie von Littrup et al. (1993) mit 31%. Der Gesamtscore der hinsichtlich der methodisch-inhaltlichen Qualität am besten abschneidenden Untersuchungen von Barry et al. (1995), von Krahn et al. (1994) und von Gustafsson et al. (1995) liegt bei 70%, 68% und 65%.

Die Variationsbreite des durchschnittlichen Erfüllungsgrades der Qualitätsdimensionen reicht von 22% bis 59%. Deutliche Schwächen hinsichtlich der methodisch-inhaltlichen Qualität lassen sich im Durchschnitt der Studien in folgenden Bereichen erkennen (in Klammern der durchschnittliche prozentuale Erfüllungsgrad):

- Behandlung von Unsicherheiten (22%) - die Variationsbreite über die Studien hinweg liegt zwischen 0% und 58%
- Gesundheitseffekte (26%) - die Variationsbreite über die Studien hinweg liegt zwischen 0% und 67%
- Diskussion (27%) - die Variationsbreite über die Studien hinweg liegt zwischen 0% und 64%
- Diskontierung (30%) - die Variationsbreite über die Studien hinweg liegt zwischen 0% und 100%.

Ein deutlich höheres Ausmaß an methodisch-inhaltlicher Qualität findet sich dagegen in den Bereichen:

- Schlussfolgerungen (59%) - die Variationsbreite über die Studien hinweg liegt zwischen 25% und 100%
- Fragestellung (58%) - die Variationsbreite über die Studien hinweg liegt zwischen 25% und 100%
- Ergebnispräsentation (57%) - die Variationsbreite über die Studien hinweg liegt zwischen 29% und 86%

C.5.5 Qualitative Informationssynthese ökonomischer Parameter

C.5.5.1 Berücksichtigte Kostenarten

In Tabelle 20 sind die in den evaluierten Studien einbezogenen Kosten dargestellt. Die Kostenarten wurden in direkte medizinische Kosten und direkte nichtmedizinische Kosten und Produktivitätsausfallskosten unterteilt. Berücksichtigt wurden sowohl variable als auch Fixkosten.

Tabelle 20: Übersicht über die berücksichtigten Kostenarten in den evaluierten gesundheitsökonomischen Studien

Kosten	Abramson et al. (1994)	Barry et al. (1995)	Benoit & Naslund (1994)	Chadwick et al. (1991)	Ciatto et al. (1995)	de Biasi et al. (1996)	Dorr et al. (1993)	Gottlieb et al. (1996)	Gustafsson et al. (1995)	Kantrowitz et al. (1995)	Krahn et al. (1994)	Kramer et al. (1993)	Launois (1992)	Littrup et al. (1993)	Littrup et al. (1994)	Optenberg & Thompson (1990)
Direkte medizinische Kosten																
Programmimplementierung, Organisation	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Rekrutierung	-	-	-	-	-	-	-	-	X	X	-	-	-	-	-	-
Screening-Untersuchung	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Konfirmatorische Diagnostik	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X	X
Weiterverfolgung falschpositiver, schwachpositiver oder zweifelhafter Testergebnisse	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-
Behandlung identifizierter Karzinome	X	X	-	-	-	-	X	X	-	-	X	X	-	X ¹	X ¹	X
Behandlung von fortgeschrittener oder metastasierender Erkrankung	X	X	-	-	-	-	X	X	-	-	X	X	-	X ¹	X ¹	X
Behandlung von Behandlungskomplikationen	?	X	-	-	-	-	X	-	-	-	X	X	-	X ¹	X ¹	X
Regelmäßige posttherapeutische Nachsorge	-	X	-	-	-	-	-	-	-	-	X	-	-	-	-	-
Kosten für die Behandlung anderer Erkrankungen aufgrund zusätzlicher Lebensjahre	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Fixkosten (Personal, Ausstattung, Räume)	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-
Direkte nichtmedizinische Kosten																
Transport zur Screening-Untersuchung	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-
Transport zur Behandlung	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Haushaltshilfe / Pflege	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Produktivitätsausfallskosten																
durch Screening-Untersuchung (Reisen, Warten)	-	-	-	-	-	-	-	-	X	-	-	-	-	-	-	-
durch Behandlung / Langzeitmorbidity / Frühmortality	-	-	-	-	-	-	-	-	-	-	-	-	-	X ¹	X ¹	-

¹ Die Punkte waren nur in der KNA, nicht in der KEA berücksichtigt

Eine ausführliche Interpretation der in den jeweiligen Studien verwendeten Kostenarten und deren Auswirkung auf die Validität der Studien erfolgt in der Diskussion.

Diese Ergebnisse sind in aggregierter Form auch in der zusammenfassenden Tabelle 15 wiedergegeben. In der Zeile "Kosten - Beschreibung" wurden die Kostenarten unter den Begriffen Rekrutierung, Screening (für Screening-Untersuchung), Diagnose (für konfirmatorische Diagnostik), primäre Therapie (für Behandlung identifizierter Karzinome, Behandlung von fortgeschrittener oder metastasierender Erkrankung), sekundäre Therapie (Behandlung von Behandlungskomplikationen), Nachsorge (regelmäßige posttherapeutische Nachsorge) und Personal-, Sach- und Fixkosten zusammengefasst. Sofern berücksichtigt, wurden direkte medizinische Kosten und Produktivitätsausfallkosten hier ebenfalls genannt. Diese Tabelle erlaubt dem Leser die Herstellung und Interpretation beliebiger Bezüge zwischen den einzelnen Merkmalen der Studien.

C.5.5.2 Kosten und Kosten-Effektivität

Um einen deskriptiven Vergleich der ermittelten Kostenparameter in den einzelnen Studien zu ermöglichen, wurden die ermittelten Gesamtkosten pro Screening-Teilnehmer aus den einzelnen Studien zusammengestellt. Ferner wurden die Kosten pro identifiziertem Krebsfall extrahiert; sie dienen als Maßzahl der Kosten-Effektivitäts-Relation von Screening-Studien, deren Zeithorizont sich auf den reinen Screening-Prozess beschränkt. Für Screening-Studien, deren gesundheitsbezogenes Outcome die Überlebenswahrscheinlichkeit bzw. die (Rest-)Lebenserwartung ist, wurden zusätzlich die Kosten pro gewonnenem Lebensjahr (Kosten/LY) als Maßzahl der Kosten-Effektivitäts-Relation extrahiert. Für Studien mit qualitätsbereinigter Restlebenserwartung wurden die Kosten pro gewonnenem qualitätsadjustiertem Lebensjahr (Kosten/QALY) als Maßzahl der Kosten-Nutzwert-Relation berichtet. Wo Maßzahlen in den Publikationen nicht angegeben waren, wurden diese errechnet, soweit dies aufgrund der berichteten Werte möglich war (Einzelheiten s. Tabelle 15).

Währungskonversionen wurden anhand der von der OECD veröffentlichten Kaufpreisparitäten für Gesundheitsgüter des Jahres 1993 durchgeführt. Von einer Deflationierung wurde abgesehen, da in Deutschland derzeit keine zuverlässigen sektoralen Inflationsraten für Gesundheitsgüter zur Verfügung stehen und die in

diesem Bericht eingeschlossenen Studien (1990-1996) einen maximalen zeitlichen Abstand von drei Jahren nach oben oder unten zum Gültigkeitsjahr 1993 aufweisen.

Tabelle 21 stellt die evaluierten Studien und die extrahierbaren Kostenparameter zusammengefasst für die 16 Studien in der in der Originalstudie verwendeten Währung dar, Tabelle 22 zeigt die entsprechenden Werte nach Währungskonversion auf DM.

In Tabelle 21 und Tabelle 22 sind nur Ergebnisse dargestellt, die sich auf Screening-Strategien beziehen, die eine PSA-Bestimmung einschließen. Ergebnisse zu Strategien ohne PSA-Bestimmung können Tabelle 15 entnommen werden. Durchschnittliche und inkrementelle Analysen wurden als solche gekennzeichnet. Inkrementelle Werte wurden nur berücksichtigt, wenn sie sich auf die Vergleichsstrategie "Kein Screening" bezogen. Inkrementelle Ergebnisse, die sich auf den Vergleich von PSA-Screening-Varianten beziehen sind ebenfalls Tabelle 15 zu entnehmen.

Tabelle 21: Kosten und Kosten-Effektivität der untersuchten PSA-Screening-Verfahren. Ohne Währungskonversion.

Studie	Studien- typ	Kosten pro Screening- Teilnehmer K/ST	Kosten pro identifizierter Krebsfall K/Ca	Kosten pro gerettetes Leben K/L	Kosten pro gewonnenes Lebensjahr K/LJ	Kosten pro gewonnenes QALY K/QALY
Abramson et al. (1994)	KEA	d: US-\$ 520	d: US-\$ 16,300			
Barry et al. (1995)	KEA	Barry: i: 65J: US-\$ 293 70J: US-\$ 365 75J: US-\$ 414 Coley: i: 50-59J: US-\$ 216 60-69J: US-\$ 387 70-79J: US-\$ 532			Barry i: 65J: US-\$ 15,123 70J: US-\$ 25,289 75J: US-\$ 51,267 Coley i: 65J: US-\$ 12,491 70J: US-\$ 18,769 75J: US-\$ 65,909	
Benoit & Naslund (1994)	KEA	d: US-\$ 55-109	d: US-\$ 2,205-2,372			
Chadwick et al. (1991)	KEA	d: GB-£ 25	d: GB-£ 1,654			
Ciatto et al. (1995)	KEA	d: Lit. 30,400	d: Lit. 1,817,000			
de Biasi et al. (1996)	KEA	d: Lit. 22,000	d: Lit. 932,000			
Dorr et al. (1993)	KKA/KEA	d: US-\$ 597	d: US-\$ 1,924 (nur Screening + Diagnose)	d: US-\$ 197,647		
Gottlieb et al. (1996)	KUA	k.A.				i: 50J: "Screening" dominant 60J: US-\$ 8,974-13,558 70J: "Kein Screening" dominant

(Fortsetzung)

Fortsetzung

Studie	Studien- typ	Kosten pro Screening- Teilnehmer K/ST	Kosten pro identifizierter Krebsfall K/Ca	Kosten pro gerettetes Leben K/L	Kosten pro gewonnenes Lebensjahr K/LJ	Kosten pro gewonnenes QALY K/QALY
Gustafsson et al. (1995)	KEA	d: US-\$ 71-116	d: US-\$ 3,180-4,470			
Kantrowitz et al. (1995)	KEA	d: US-\$ 59	d: US-\$ 6,012			
Krahn et al. (1994)	KUA	i: 50J: US-\$ 77-253 60J: US-\$ 179-423 70J: US-\$ 448-874			i: 50J: US-\$ 113,000-729,000 60J: US-\$ 127,000-475,000 70J: US-\$ 189,000-466,000	i: "Kein Screening" dominiert alle Strategien in allen Altersgruppen
Kramer et al. (1993)	KKA	d: US-\$ 562				
Launois (1992)	KEA	d: FF 166 -835	d: FF 18,360-29,387 i: FF 18,360-52,104		i: FF 102,230-288,110	
Littrup et al. (1993)	KEA/KNA	k.A.	d : US-\$ 2,905-6,556 (1. Jahr) d: US-\$ 4,482-11,135 (3-J.-Durchschnitt)			
Littrup et al. (1994c)	KEA/KNA	d: US-\$ 101-219	d: US-\$ 3,144-4,752			
Optenberg & Thompson (1990)	KKA/KEA	d: US-\$ 648-1,598			d: US-\$ 116-288	

Dargestellt sind nur Ergebnisse, die sich auf Screening-Strategien beziehen, die eine PSA-Bestimmung einschließen.

Wenn die in die Kostenberechnungen einbezogenen Kostenarten von der vorigen Kostenarten-Tabelle abweichen, ist dies gekennzeichnet.

k.A.: keine Angaben

d: durchschnittliche Analyse

i: inkrementelle Analyse

Inkrementell nur dargestellt, wenn gegen "Kein Screening" evaluiert wurde; also keine inkrementelle Darstellung für Vergleiche zwischen evaluierten Screening-Programmen unterschiedlicher Modalität.

Tabelle 22: Kosten und Kosten-Effektivität der untersuchten PSA-Screening-Verfahren. Währungskonversion über Kaufpreisparitäten im Gesundheitssektors des Jahres 1993 (Quelle: OECD Health Data 1997).

Studie	Studien- typ	Kosten pro Screening- Teilnehmer	Kosten pro identifizierter Krebsfall	Kosten pro gerettetes Leben	Kosten pro gewonnenes Lebensjahr	Kosten pro gewonnenes QALY
		K/ST	K/Ca	K/L	K/LJ	K/QALY
Abramson et al. (1994)	KEA	d: DM 588	d: DM 18 419			
Barry et al. (1995)	KEA	Barry: i: 65J: DM 331 70J: DM 412 75J: DM 468 Coley: i: 50-59J: DM 244 60-69J: DM 437 70-79J: DM 601			Barry i: 65J: DM 17 089 70J: DM 28 577 75J: DM 57 932 Coley i: 65J: DM 14 115 70J: DM 21 210 75J: DM 74 477	
Benoit & Naslund (1994)	KEA	d: DM 62-123	d: DM 2 492-2 680			
Chadwick et al. (1991)	KEA	d: DM 109	d: DM 7 189			
Ciatto et al. (1995)	KEA	d: DM 39	d: DM 2 344			
de Biasi et al. (1996)	KEA	d: DM 28	d: DM 1 202			
Dorr et al. (1993)	KKA/KEA	d: DM 675	d: DM 2 174 (nur Screening + Diagnose)	d: DM 22 3341		
Gottlieb et al. (1996)	KUA	k.A.				i: 50J: "Screening" dominant 60J: DM 10 141-15 321 70J: "Kein Screening" dominant

(Fortsetzung)

Fortsetzung

Studie	Studien- typ	Kosten pro Screening- Teilnehmer K/ST	Kosten pro identifizierter Krebsfall K/Ca	Kosten pro gerettetes Leben K/L	Kosten pro gewonnenes Lebensjahr K/LJ	Kosten pro gewonnenes QALY K/QALY
Gustafsson et al. (1995)	KEA	d: DM 80-131	d: DM 3 593-5 051			
Kantrowitz et al. (1995)	KEA	d: DM 67	d: DM 6 794			
Krahn et al. (1994)	KUA	i: 50J: DM 87-286 60J: DM 202-478 70J: DM 506-988			i: 50J: DM 127 690-823 770 60J: DM 143 510-536 750 70J: DM 213 570-526 580	i: "Kein Screening" dominiert alle Strategien in allen Altersgruppen
Kramer et al. (1993)	KKA	d: DM 635				
Launois (1992)	KEA	d: DM 59-298	d: DM 6 545-10 476 i: DM 6 545-18 573		i: DM 36 442-102 702	
Littrup et al. (1993)	KEA/KNA	k.A.	d: DM 3 282-7 408 (1. Jahr) d: DM 5 065-12 582 (3-J.-Durchschnitt)			
Littrup et al. (1994)	KEA/KNA	d: DM 114-248	d: DM 3 553-5 370			
Optenberg & Thompson (1990)	KKA/KEA	d: DM 732-1 806			d: DM 131-325	

Dargestellt sind nur Ergebnisse, die sich auf Screening-Strategien beziehen, die eine PSA-Bestimmung einschließen.

Wenn die in die Kostenberechnungen einbezogenen Kostenarten von der vorigen Kostenarten-Tabelle abweichen, ist dies gekennzeichnet.

k.A.: keine Angaben

d: durchschnittliche Analyse

i: inkrementelle Analyse

Inkrementell nur dargestellt, wenn gegen "Kein Screening" evaluiert wurde; also keine inkrementelle Darstellung für Vergleiche zwischen evaluierten Screening-Programmen unterschiedlicher Modalität.

Inkrementelle Kosten und Kosten-Effektivitäts-Ratios werden nur von einem Teil der Studien berichtet. Die zusätzlichen Kosten durch PSA-Screening wurden je nach Alter der Zielpopulation mit 87 DM bis 988 DM pro Screening-Teilnehmer angegeben. Angaben für inkrementelle Kosten pro zusätzlich identifiziertem Krebsfall streuen zwischen 6 545 DM und 18 573 DM. Die zusätzlichen Kosten pro gerettetes Leben wurden in keiner Studie berichtet. Zusätzlichen Kosten pro gewonnenes Lebensjahr belaufen sich je nach Alter der Zielpopulation auf 14 115 DM bis 823 770 DM. Inkrementelle Kosten pro gewonnenes QALY wurden mit 10 141 DM bis 15 321 DM nur für das Screening von 60-Jährigen berechnet, da sich Screening in höheren Altersgruppen sowohl als teurer als auch als weniger wirksam erwies als die Referenzalternative "Kein Screening". Screening bei 50-Jährigen wurde widersprüchlich beurteilt. Während es von Gottlieb et al. (1996) als dominante Strategie in dieser Altersgruppe eingestuft wird, kommen Krahn et al. zu dem Schluss, Screening sei zum einen teurer als "Kein Screening", bringe aber unter der Berücksichtigung der Lebensqualität einen medizinischen Nettoverlust mit sich.

Es sei nochmals darauf hingewiesen, dass die quantitativen Parameter keine Metaanalyse im engeren Sinne zulassen, da in den einzelnen Studien unterschiedliche Rahmenbedingungen herrschten. Es wurden verschiedenste Screening-Strategien verglichen und verschiedene PSA-Grenzwerte verwendet. Der Evaluationsrahmen unterschied sich bezüglich vieler Faktoren, darunter Perspektive, Zeithorizont, untersuchte Altersgruppen, Jahr der Durchführung und insbesondere Jahr der berücksichtigten Kostenarten. Die Aufstellung kann also lediglich einem deskriptiven Vergleich dienen.

C.6 Diskussion

Die Zielsetzung dieses Berichtes war die Entwicklung eines Bewertungsinstrumentariums für gesundheitsökonomische Studien und dessen exemplarische Anwendung am Thema "PSA-Screening beim Prostatakarzinom". Damit sollte der Prozess der medizinischen Technologiefolgenabschätzung um die ökonomische Komponente ergänzt werden. Es folgt eine Diskussion der entwickelten Instrumente unter Berücksichtigung der Voraussetzungen für ihre Anwendbarkeit. Anschließend werden die Ergebnisse der Bewertung der gesundheitsökonomischen Studien zum Thema "PSA-Screening beim Prostatakarzinom" diskutiert und Betrachtungen zur Übertragbarkeit dieser Ergebnisse nach Deutschland angestellt.

C.6.1 Bewertungsinstrumentarium für gesundheitsökonomische Studien

Übergeordnetes Ziel bei Datenbanken, wie sie von den verschiedenen ausländischen mit HTA befassten Einrichtungen oder auch von der Cochrane Collaboration im Falle der evidenz-basierten Medizin (EBM) unterhalten werden, ist die systematische Erschließung, Aufbereitung und Dissemination vorhandener wissenschaftlicher Evidenz. Auf ihr qualitätsgesichertes, auch neueste Erkenntnisse berücksichtigendes Informationsangebot soll bei spezifischen Fragestellungen zur Unterstützung der Entscheidungsfindung zurückgegriffen werden können. Damit der Nutzer einer Datenbank seiner Informationsauswahl eine nachvollziehbare, von subjektiven Einflüssen möglichst unabhängige Interpretation von Studienergebnissen zugrunde legen kann, ist eine systematische Methodik der Sammlung, Bewertung und Synthese von Informationen aus diversen Quellen unabdingbar.⁴⁶ Dieser Umstand setzt eine strukturierte und formalisierte Vorgehensweise voraus, welche sich entscheidend auch nach den spezifischen Zielsetzungen und Aufgaben der jeweiligen Datenbank richtet.¹²⁸

Da in der derzeitigen Entwicklungsphase der vom BMG vorgesehenen Datenbasis die genauen Bearbeitungs- und Aktualisierungsmodalitäten einschließlich einer verstetigten Ressourcenausstattung noch nicht endgültig feststehen, wurden von den drei am BMG-Projekt beteiligten ökonomischen Arbeitsgruppen verschiedene Alternativen zum Vorgehen geprüft. Der im Konsens entstandene Ansatz der Technologiebewertung ist das Produkt zweierlei Bemühungen: Zum einen galt es,

unterschiedlichen Erwartungen und Bedürfnissen potentieller Nutzer einer Datenbank - seien es Gesundheitspolitiker, Verwaltungsleiter, Ärzte, Verantwortliche der für das Gesundheitswesen relevanten Vorleistungsindustrien oder auch Forscher - gerecht zu werden. Zum anderen sollte die Komplexität des Bewertungsproblems bei ökonomischen Evaluationen mit den dabei zu berücksichtigenden medizinischen, ökonomischen, epidemiologischen, biometrischen und entscheidungsanalytischen Fragestellungen zum Ausdruck gebracht werden. So lassen sich auch die in Abhängigkeit von der Zielsetzung einer Datenbank unterschiedlichen fachlichen und personellen sowie damit verbundene finanzielle Erfordernisse für die Durchführung von Technologiebewertungen unter gesundheitsökonomischen Aspekten offen legen.

C.6.1.1 Deskriptionskatalog

Wie auch bei der Bearbeitung des Beispielthemas "PSA zum Screening des Prostatakarzinoms" deutlich wurde, ist der Aufbau von Publikationen, wie sie in den einschlägigen Fachzeitschriften erscheinen, für eine schnelle Informationsbeurteilung und -nutzung meistens nicht geeignet. Die Präsentation ist oft unsystematisch, in vielen Fällen finden sich die bei der Qualitätsbewertung der betreffenden Studie zu integrierenden Informationen verstreut über die Publikation bzw. einzelne Publikationsabschnitte dokumentiert. Zum Teil kommt eine uneinheitliche Anwendung der Terminologie hinzu. Insgesamt ist die Möglichkeit einer systematischen Nutzung der vorhandenen Information zumeist nicht gegeben. Zur Systematisierung der Datenfindung bietet der Deskriptionskatalog die Möglichkeit, auf disaggregierter Ebene die Einzelinformationen isoliert aus den Publikationen abzufragen. Mit seiner Hilfe können die beurteilungsrelevanten Informationen einer Studie strukturiert gesammelt und dokumentiert werden. Gleichzeitig verdeutlicht diese Dokumentationsmaske, welche inhaltlichen und methodischen Zusammenhänge es bei der sich anschließenden eigentlichen Qualitätsbewertung zu bedenken und zu verarbeiten gilt. Er zielt jedoch nicht selbst auf eine Bewertung der Studienqualität ab, sondern dient als Instrument zur Objektivierung der berichteten Information. Eine Bewertung anhand des Transparenz- und/oder des Qualitätskataloges sollte sich deshalb in jedem Falle anschließen.

Die Identifikation und Erfassung beurteilungsrelevanter Informationen sind Grundlage eines jeden informationsaggregierenden und -integrierenden Bewertungsprozesses. Andererseits ist die Anwendung der vielen Einzelitems des

Deskriptionskataloges sehr zeitintensiv und damit abhängig von den zur Verfügung stehenden Ressourcen. Es bestand Konsens unter den drei gesundheitsökonomischen Projektgruppen, dass die Anwendung des Deskriptionskataloges einen wesentlichen Beitrag zur Standardisierung der Terminologie, der spezifischen gesundheitsökonomischen Methodik und des Bewertungsvorganges leistet. Dieses Instrument ist in diesem Sinne als Schulungsinstrument und Strukturierungshilfe zu verstehen. Es sollte auch im Hinblick auf die Anwendung der kürzeren Kriterienkataloge (Transparenz- und Qualitätskatalog) als qualitätssichernde Maßnahme beim gesundheitsökonomischen Bewertungs- und Beurteilungsprozess von den verschiedenen Projektgruppen zumindest für einige Studien exemplarisch angewendet werden, um eine valide und reliable Dokumentation und Bewertung zu gewährleisten. Das Erfassen der deskriptiven Kriterien stellt dabei in jedem Beurteilungsprozess einen kognitiven Zwischenschritt dar. Implizit sollte das Schema jedem Bewertungsprozess zugrunde liegen.

C.6.1.2 Transparenz- und Qualitätskatalog

Die im Konsensverfahren durch das Panel der gesundheitsökonomischen Arbeitsgruppen festgelegten Kriterien zur Beurteilung von gesundheitsökonomischen Evaluationsstudien beruhen auf methodischen Prinzipien der Durchführung und Dokumentation ökonomischer Untersuchungen, über welche zumindest im Grundsatz ein weitgehender internationaler und nationaler Konsens unter Fachwissenschaftlern herrscht.^{17,20,38,40,43,48,53,76,77,122,126,138} An der Einhaltung dieser Normen rein konventioneller Natur kann sich die Bemessung der Qualität einer Studie orientieren. Wie aus dem Kapitel "Methoden gesundheitsökonomischer Evaluation" und dem Kapitel "Bewertungsinstrumentarium für gesundheitsökonomische Studien" hervorgeht, muss dabei allerdings bedacht werden, dass hinsichtlich der Befolgung dieser Prinzipien einige unterschiedliche Ansichten und methodische Ansätze existieren, über die unter Fachwissenschaftlern weiterhin kontrovers diskutiert wird. Einen Überblick über methodische Aspekte der ökonomischen Evaluation, welche ein hohes bzw. ein geringes Maß an Übereinstimmung finden, vermittelt Tabelle 23.

Tabelle 23: Konsens und Dissens zu methodischen Fragen der ökonomischen Evaluation

Methodische Aspekte mit hoher Übereinstimmung	Methodische Aspekte mit geringer Übereinstimmung
Terminologie der ökonomischen Evaluation	Berücksichtigung von indirekten (Produktivitäts-)Kosten
Superiorität der Marginalbetrachtung	Berücksichtigung der Kosten der Gesundheitsversorgung in zusätzlich gewonnenen Lebensjahren
Bedeutung der Gegenüberstellung von Technologiealternativen	Methode der Messung des Gesundheitszustands
Analytische Betrachtungsweise und Berücksichtigung der gesellschaftlichen Aspekte	Höhe der Diskontrate für Gesundheitseffekte
Diskontierung (im Prinzip)	Berücksichtigung von Verteilungsfragen
Bedeutung von Sensitivitätsanalysen	Berücksichtigung intersektoraler Konsequenzen

nach Drummond, Brandt, Luce & Rovira, 1993⁴⁰

Insbesondere im Zusammenhang mit der Nutzenmessung und der monetären Bewertung von Gesundheitszuständen wird es auch in unmittelbarer Zukunft bei einer Situation des Nebeneinanders alternativer Ansätze und Verfahren bleiben. Die nicht abgeschlossene Diskussion um methodische Standards erschwert die Einschätzung der Studien- und Ergebnisqualität und der sich daraus ableitenden Folgerungen. Auch ist keine der verschiedenen Evaluationsformen bisher so ausgereift, dass sie den Anspruch der Überlegenheit bzw. Allgemeingültigkeit erheben kann,⁷³ weshalb bei der Qualitätsbewertung von ökonomischen Studien jeweils unterschiedliche, jedoch zulässige Vorgehensweisen und Verfahrensansätze berücksichtigt werden müssen. Daraus erhebt sich zunächst der Anspruch auf eine durchgängige Dokumentation, Transparenz und Nachvollziehbarkeit der von den Wissenschaftlern hinsichtlich der Studiendurchführung getroffenen methodischen und inhaltlichen Entscheidungen.^{41,48,126} Nur darauf aufbauend kann die methodisch-inhaltliche Qualität, d.h. die Zweckmäßigkeit und Angemessenheit dieser Entscheidungen in dem von der Studie vorgegebenen Bezugsrahmen aus Fragestellung, Evaluationsansatz und Studiendesign kritisch betrachtet und in gegebenenfalls einem weiteren Schritt auf den Urteils- und Entscheidungskontext des jeweiligen Nutzers übertragen werden.

Transparenzkatalog

Grundlegendes Qualitätsmerkmal von ökonomischen Studien ist die Transparenz ihrer methodischen und inhaltlichen Anlage und Vorgehensweise.^{41,126} Das Studienziel, der Evaluationsrahmen, alle Eingabeparameter, Modelle und Annahmen, auf welchen die jeweilige Untersuchung beruht, sollten klar erkennbar sein. Die einzelnen Transparenzdimensionen wurden im Ergebnisteil dargestellt.

Das Ausmaß an Transparenz einer Studie hat nicht nur einen entscheidenden Einfluss auf die Akzeptanz der von den Autoren gewählten Verfahrensansätze. In Abhängigkeit von der Nachvollziehbarkeit der wichtigsten Studienergebnisse steigt oder fällt auch die Glaubwürdigkeit einer Studie. Allerdings ist die transparente und nachvollziehbare Offenlegung von Ausgangsbedingungen, Methoden und Ergebnissen eine notwendige, aber nicht hinreichende Voraussetzung, um Aussagen über ein in sich schlüssiges und solides Vorgehen zu ermöglichen. Ein solches ist für die methodisch-inhaltliche Qualität einer Studie ausschlaggebend und bedarf einer differenzierteren Überprüfung.

Die Beurteilung der formalen Transparenz und Nachvollziehbarkeit einer Studie anhand des Transparenzkataloges ist mit Grundkenntnissen der Methoden auf dem Gebiet der klinischen Epidemiologie, der Gesundheitsökonomie und der Biometrie möglich. Das Hinzuziehen medizinischer Experten für das jeweilige Thema ist für diesen formalen Beurteilungsprozess nicht notwendig. Soweit die notwendigen Grundkenntnisse aus den unterschiedlichen Disziplinen in einer einzigen Person vereint sind, ist es theoretisch vorstellbar, die Transparenzbewertung von einer Person durchführen zu lassen. Das resultierende Transparenzprofil gibt allerdings nur oberflächliche Hinweise auf die methodisch-inhaltliche Qualität der entsprechenden gesundheitsökonomischen Untersuchung.

Qualitätskatalog

Abgesehen von ihrer transparenten und nachvollziehbaren Dokumentation müssen die von den Autoren der jeweiligen Studien gewählten Mess- und Analyseverfahren für eine gute Studienqualität dem situativen Kontext, d.h. der Perspektive und dem Entscheidungskontext, angemessen sein. Um dies beurteilen zu können, sind spezifische Kenntnisse zur interessierenden Erkrankung bzw. Indikation und zu dem mit ihr im Zusammenhang stehenden medizinischen, sozialen, gesundheitsökonomischen und -politischen Umfeld wie auch zu methodisch-analytischen

Verfahren erforderlich. Zur kritischen Prüfung der inhaltlichen und methodischen Studienqualität bedarf es u.a. Kenntnisse

- der Epidemiologie der Erkrankung

Beispielsweise sollten Informationen zu Häufigkeit und Verteilung der Erkrankung innerhalb bestimmter Bevölkerungsgruppen vorliegen sowie zu Faktoren, die Entstehung, Verlauf, Heilung, Rehabilitation und Rückfall der Erkrankung beeinflussen.

- des Versorgungsgebiets sowie der versorgenden Einrichtungen und Personenkreise

Beispielsweise sollte ein guter Überblick über das komplexe Beziehungsgeflecht der Entscheidungseinheiten vorhanden sein, bei welchen bei medizinischen Leistungen Nutzen und Kosten auftreten können.

- der Behandlungs- und Interventionsmethoden

Beispielsweise muss ein klares Verständnis über derzeitige und, falls bereits absehbar, zukünftige Behandlungs- und Interventionsmethoden vorliegen. Kenntnisse über Risikofaktoren, Screening-Methoden, Diagnostik, Behandlungspläne und verfügbare -alternativen sowie Möglichkeiten der Nachsorge sind unerlässlich. Auch sollten bestehende Therapierichtlinien bekannt sein.

- der gesundheitsökonomischen und -politischen Rahmenbedingungen

Für die Bewertung gesundheitsökonomischer Studien sind Fachkenntnisse über die im deutschen Gesundheitssystem angewendeten Abrechnungssysteme notwendig. Auch Wissen um die gesundheitspolitischen Rahmenbedingungen und Anreizsysteme ist erforderlich, um abschätzen zu können, inwieweit der Einfluss der jeweiligen Interventionen auf die Kosten angemessen analysiert wurde.

- der entscheidungsanalytischen Modelle

Beispielsweise sollten Besonderheiten der formalen entscheidungsanalytischen Modellierungsansätze (z.B. Entscheidungsbaumanalyse, Markov-Modellierung) bekannt sein und mit entsprechenden Software-Programmen umgegangen werden können, um Modelle kritisch würdigen zu können. Auch ist zur

Einschätzung der Qualität der in Modelle eingehenden Parameter ein klares Verständnis über systematische Reviews und Meta-Analysen erforderlich.

- der epidemiologischen Methodik in Studienplanung, Studiendurchführung und Analyse

Es sollten fundierte Kenntnisse zu den verschiedenen epidemiologischen Studiendesigns in Reinform und den daraus resultierenden Hybridstudientypen mit ihren Vor- und Nachteilen vorhanden sein. Eine mehrjährige praktische Erfahrung bei der Durchführung und Auswertung epidemiologischer Studien ist erforderlich.

- biometrische Fachkenntnisse

Insbesondere bei komplexeren multivariaten Verfahren und probabilistischen Entscheidungsanalyse-Modellen, wie sie von den Editoren immer mehr gefordert werden, sind fundiertes Fachwissen und eigene praktische Erfahrung Voraussetzung für die Überprüfung adäquater Methoden. Nur ein Statistiker mit entsprechender Expertise in den genannten Gebieten vermag die Angemessenheit der eingesetzten Methoden und deren Aktualität in einem sich rasch fortentwickelnden Bereich zu beurteilen.

- Diese Darstellung weist auf die bei solider Durchführung gegebene Komplexität sowohl von gesundheitsökonomischen Evaluationsstudien als auch von deren Bewertung hin. Bei der Qualitätsbewertung ausländischer Studien kommt hinzu, dass Kenntnisse der Gegebenheiten und Verhältnisse in den betreffenden Ländern erforderlich sind. Gleichzeitig wird deutlich, dass die methodisch-inhaltliche Qualität von gesundheitsökonomischen Untersuchungen immer zweckgebunden und damit relativ ist, da diese sachlich, zeitlich und räumlich nur begrenzte Aussagekraft besitzen. Sollen die Ergebnisse einer ökonomischen Evaluation zur Unterstützung von Allokationsentscheidungen in einem anderen Kontext - etwa in einem anderen Land oder bei einer anderen Zielpopulation - herangezogen werden, muss des weiteren deren Übertragbarkeit und Anwendbarkeit beurteilt werden.^{5,76,87,159}

Die angemessene Verwendung des Qualitätskataloges zur Beurteilung der methodisch-inhaltlichen Qualität einer Studie setzt spezifische, auch länderübergreifende Fachkenntnisse auf dem Gebiet der Medizin, der Gesundheitsökonomie, der klinischen Epidemiologie, der Biometrie sowie gegebenenfalls der formalen Entscheidungsanalyse und der Lebensqualitäts-

messung voraus. Eine Qualitätsbewertung von ökonomischen Studien auf diesem Anspruchsniveau ist damit grundsätzlich nur in interdisziplinärer Zusammenarbeit eines Expertenteams möglich und damit ressourcenintensiv.

Bewertungsprozess

Das in diesem Bericht untersuchte Anwendungsbeispiel "PSA-Screening" bestätigt die aus der allgemeinen Praxis gesundheitsökonomischer Evaluationen abgeleitete Beobachtung, dass nur wenige gesundheitsökonomische Evaluationsstudien weitgehend die im Transparenz- und Qualitätskatalog aufgestellten inhaltlichen und methodischen Kriterien erfüllen. Die Liste der hier aufgestellten Anforderungen ist somit als methodischer Goldstandard aufzufassen, anhand dessen die Stärken und Schwächen einer gesundheitsökonomischen Studie herausgefiltert und dargestellt werden sollen. Dabei mag es attraktiv erscheinen, auf aggregierter Ebene ein pauschales Transparenz- bzw. Qualitätsurteil über eine Studie in Form von "gut" oder "schlecht" allein anhand der erreichten Summenscores zu fällen. Dagegen spricht jedoch zum einen, dass ein solcher Summenscore, wie auch immer geartet, keinen Anhalt darüber gibt, welche Aspekte einer Studie für Transparenz- bzw. Qualitätsbeeinträchtigungen verantwortlich sind. Damit entfällt die Möglichkeit, die relative Bedeutung der jeweiligen Aspekte selbst zu beurteilen. Zum anderen wird bei der hier gewählten Gleichgewichtung der einzelnen Kriterien bzw. Themenbereiche nicht der dahinter stehende unterschiedliche Informationsgehalt berücksichtigt: Beispielsweise wird mit der Frage "Wurde der Typ der gesundheitsökonomischen Evaluation explizit genannt?" ein einzelner Informationsbestandteil aufgegriffen. Die Frage "Wurden angemessene Methoden zur Identifikation, Extraktion und Synthese der Effektparameter verwendet und wurden sie detailliert beschrieben?" stellt dagegen nicht nur eine Doppelfrage dar, sondern zielt auch auf eine ganze Reihe spezifischer Informationen bezüglich systematischer Reviews bzw. Meta-Analysen ab. Mit den detaillierteren Summenscores für die einzelnen Transparenz- bzw. Qualitätsdimensionen auf der aggregierten Ebene wird die Gewichtungproblematik zwar nicht ausgeräumt, doch lässt sich ein differenzierteres Bild von der Transparenz bzw. der Qualität einer Studie gewinnen. Allerdings sollte dabei im Auge behalten werden, durch welche Kriterien die jeweiligen Summenscores determiniert werden. So wird beispielsweise der Themenbereich Diskontierung u.a. durch das Kriterium der Nennung des Referenzjahres für die Kosten bestimmt ("Wurde das Referenzjahr für die Diskontierung angegeben bzw. bei fehlender Diskontierung das Referenzjahr für die

Kosten?"). Ist in der entsprechenden Studie keine Diskontierung erforderlich, wird die Ausprägung des Themenbereichs nunmehr dichotom allein durch die Nennung bzw. Nichtnennung des Referenzjahres für die Kosten determiniert, ohne mit einer Diskontierung im eigentlichen Sinne im Zusammenhang zu stehen.

Um potentielle subjektive Einflüsse bei der Studienbewertung zu kontrollieren und die Qualitätsbeurteilung zu standardisieren, sollte bei den einzelnen Bewertern ein möglichst einheitliches Verständnis der einzelnen Fragen in den Kriterienkatalogen und den dahinter stehenden Anforderungen an die Transparenz bzw. Qualität von Studien vorliegen. Als begriffliche und inhaltliche Referenz kann zunächst auf den Deskriptionskatalog verwiesen werden. Bei der Bearbeitung der Beispielthemen in den gesundheitsökonomischen Projektgruppen hat sich gezeigt, dass bei der Anwendung der Katalogfragen Rückbestätigungen zwischen den Bewertern nötig waren, um eine einheitliche und reliable Bewertung zu gewährleisten. Voraussetzung für den reliablen Einsatz der Kriterienkataloge über die drei gesundheitsökonomischen Projektgruppen hinaus ist zusätzliche Schaffung eines Begleitmanuals mit Spezifikationen zu Begriffen, Inhalten und gegebenenfalls zu Gewichtungen einzelner Fragen bzw. Frageteile. Die Erfahrungen der Bewerter aus den drei gesundheitsökonomischen Projektgruppen wurden bereits während der Entwicklungs- und Anwendungsphase der Kataloge systematisch gesammelt. Die Entwicklung eines Manuals ist in München in Zusammenarbeit mit der Hannoveraner und Ulmer Projektgruppe vorgesehen.

Insbesondere bei der Beurteilung der methodisch-inhaltlichen Qualität von ökonomischen Evaluationsstudien anhand des Qualitätskataloges kommt zwangsweise eine in Abhängigkeit von der fachlichen Kompetenz des jeweiligen Bewerter stehende subjektive Interpretationskomponente zum Tragen. Es ist deshalb anzuraten, die Bewertung von mindestens zwei Bewertern unabhängig durchführen zu lassen; bei divergierenden Einschätzungen wird eine Bewertung im Konsens gefunden und/oder ein dritter Reviewer hinzugezogen, dessen Beurteilung dann den Ausschlag gibt.¹⁷ Als weitere Maßnahme zur Minimierung von Verzerrungen bei der Studieninterpretation kann eine Verblindung der Bewerter gegenüber Angaben zu Autoren, durchführenden Institutionen und Auftraggebern der jeweiligen Studie in Erwägung gezogen werden.¹⁷ Diese Vorgehensweise hat sich bei der Bewertung des Beispielthemas "PSA-Screening beim Prostatakarzinom" bewährt.

C.6.2 PSA-Screening beim Prostatakarzinom

Zunächst wird ein Überblick über die gesundheitsökonomischen Studientypen der berücksichtigten Studien gegeben. Anschließend werden die Ergebnisse der Bewertung studienübergreifend anhand der einzelnen Dimensionen der Studienqualität diskutiert.

C.6.2.1 Überblick über die gesundheitsökonomischen Studientypen

Im Vergleich zur Menge an Publikationen, die sich in unterschiedlichster Weise mit der Wirksamkeit und Effektivität der PSA-Bestimmung zur Früherkennung des Prostatakarzinoms beschäftigen, ist nur in wenigen Untersuchungen versucht worden, auch die ökonomische Dimension dieser medizinischen Technologie zu beleuchten. Dabei ist in den 16 hierzu vorgefundenen und im Rahmen der vorliegenden Arbeit betrachteten Studien im allgemeinen nicht die PSA-Bestimmung allein Objekt der Analyse, vielmehr werden als medizinische Technologien Screening-Test-Kombinationen aus PSA-Bestimmung und anderen beim Prostatakarzinom gebräuchlichen Screening- und Diagnosemethoden untersucht. In Abhängigkeit von der übergeordneten Zielsetzung der Untersuchungen lassen sich zwei Arten von Studien unterscheiden:

- Im Mittelpunkt des Interesses der einen Studienkategorie steht die Entdeckungsrate bezüglich Prostatakarzinomen im asymptomatischen Stadium.^{1,11,25,29,34,52,74,84} Hier wird die diagnostische Wertigkeit verschiedener Screening-Teststrategien untersucht und die Evaluierung von Indikationskriterien zur Biopsie steht im Vordergrund. Dabei beschränken sich diese Studien weitgehend auf den Zeitraum des reinen Screening-Vorganges und lassen die sich ergebenden gesundheitsbezogenen und ökonomischen Langzeitfolgen außer Betracht.
- Im Mittelpunkt des Interesses der anderen Studienkategorie steht die Mortalität bzw. die Restlebenserwartung und/oder die Lebensqualität.^{8,37,49,81,83,110} Meist werden langfristige gesundheitsbezogene und ökonomische Auswirkungen der Früherkennung von Prostatakarzinom in die Untersuchung eingeschlossen. Neben den positiven Konsequenzen werden auch die Komplikationen und langfristigen Risiken des Screening-Verfahrens und der durch das Screening induzierten Folgeprozesse berücksichtigt.

- Die Studien von Littrup et al. (1993,⁹¹ 1994c⁹³) nehmen diesbezüglich eine Zwitterstellung ein.

Bei dem Versuch, die vorliegende Evidenz aus den Einzelstudien zu synthetisieren, stößt man auf die methodische Problematik, dass sich die einzelnen Studien in wichtigen Rahmenbedingungen unterscheiden, die eine Vergleichbarkeit oder Synthese im engeren Sinne verhindern. Konkret sind Differenzen in folgenden Punkten festzustellen:

- Wahl der zu vergleichenden Screening-Programme (Screening-Test-Kombinationen)
- Festlegung der Basis-Annahmen
- Mess- und Bewertungskonzepte von Ressourcenverbrauch und gesundheitsbezogenen Effekten
- Zeitpunkte der Datenerhebung und damit Stadien im Lebenszyklus (Entwicklungsstadien) der Technologien (insbesondere die PSA-Bestimmung, die TRUS und die Biopsie)
- regionale und nationale Strukturen der Gesundheitsversorgung
- Darstellung der Ergebnisse in Form von Durchschnittswerten oder inkrementellen Werten

Die ausländischen HTA-Berichte zur Früherkennung beim Prostatakarzinom haben sich vornehmlich auf die klinische Bewertung der Technologien konzentriert,^{4,26,51,129,139,149} wobei der Bericht des CETS³² mit ausführlicheren Kostenberechnungen für die Provinz Québec (Kanada) eine Sonderstellung einnimmt. In den HTA-Berichten, in welchen die gesundheitsökonomische Literatur mit berücksichtigt wurde,^{4,129,139} beschränkt sich die diesbezügliche Informationsaufbereitung auf eine meist kritiklose zusammenfassende Darstellung einzelner Studien. Diese steht im Gegensatz zu den expliziten und standardisierten Bewertungsaktivitäten bei der Wirksamkeitsanalyse. Bewertungskriterien werden lediglich in einem Bericht genannt,¹²⁹ doch ist hier der Bewertungsprozess nicht transparent dokumentiert. Soweit nachvollziehbar erschöpft er sich in der Prüfung der formalen Erfüllung der Kriterien. Die Anwendung des entwickelten Bewertungsinstruments bei den verfügbaren gesundheitsökonomischen Untersuchungen legt qualitätseinschränkende und damit auch Aussagekraft

mindernde Aspekte offen. Diese sollen nachfolgend studienübergreifend nach Themenbereichen umrissen werden, wobei der grundlegende Unterschied der Untersuchungen hinsichtlich der Fragestellung zu berücksichtigen ist. Die Dokumentationsstruktur, in welche die Qualitätsbewertung eingebunden ist, ist so angelegt, dass die Übersicht der Tabelle 15 die Datenlage zur entsprechenden Fragestellung dokumentiert, während sich die Aussagekraft dieser Informationen insbesondere in den Transparenz- bzw. Qualitätsprofilen der Bewertungsmatrizen (Tabelle 16 bis Tabelle 19) widerspiegelt. Für Detailfragen können die standardisierten Berichte einschließlich der Kommentare zu den Studien herangezogen werden, so dass der Nutzer bei Bedarf die Bewertung detailliert nachvollziehen kann.

Es sei darauf hingewiesen, dass das Nicht-Dokumentieren von relevanten methodischen und inhaltlichen Studienaspekten gleich bewertet wurde wie das offensichtliche Nicht-Erfüllen von Kriterien. Dieser Ansatz kann zu einer übergreifenden Qualitätsabwertung geführt haben. Im speziellen davon betroffen zu sein scheint die Studie von Gottlieb et al. (1994).

Die zwischen den Transparenz- und Qualitätsprofilen zu beobachtenden Abweichungen in den Einzelbewertungen wie auch in den Summen- und Gesamtscores verdeutlichen den Unterschied im Wert für den entscheidungsorientierten Nutzen und wissenschaftlichen Anspruch der beiden Kriterienkataloge. Bei keiner der Studien liegt eine Einzelbewertung im Qualitätskatalog über dem entsprechenden Punktwert im Transparenzkatalog. Auf aggregierter Ebene liegen die Summenscores bei der methodisch-inhaltlichen Qualität im Schnitt um 31% unter denjenigen bei der Transparenz. Dabei ist die Übereinstimmung zwischen Transparenz und methodisch-inhaltlicher Qualität der Studien in denjenigen Bereichen am höchsten, in welchen sich eine Bewertung hinsichtlich der Angemessenheit bzw. korrekten Durchführung von Ansätzen, Methoden, Studiengrößen oder Parametern größtenteils erübrigt, wie beispielsweise im Bereich "Diskontierung" und zum Teil auch im Bereich "Ergebnispräsentation". Andererseits ist die größte Abweichung im Bereich der "Gesundheitseffekte" festzustellen, hier gibt insbesondere das Anlegen des Maßstabes der Evidence Based Medicine zur Qualitätsbeurteilung von diagnostischen Studien bzw. systematischen Reviews den Ausschlag für die Abwertung der methodisch-inhaltlichen Qualität gegenüber der Transparenz der Studiendokumentation. Die beobachteten Unterschiede bestätigen, dass die Prüfung der Transparenz und Nachvollziehbarkeit von ökonomischen Evaluationsstudien zwar notwendig, jedoch nicht hinreichend ist, um die für eine Entscheidungsfindung relevanten Informationen

herauszuarbeiten. Dafür sind die Verfahren der ökonomischen Evaluation zu komplex und die in Abhängigkeit von Fragestellung, Evaluationsrahmen und Studiendesign stehenden Möglichkeiten, zu unterschiedlichen Ergebnissen zu gelangen, zu vielfältig.

Im folgenden werden studienübergreifend die einzelnen Teilaspekte der gesundheitsökonomischen Studien diskutiert, die gleichzeitig die Dimensionen des Transparenz- bzw. Qualitätskataloges widerspiegeln.

C.6.2.2 Fragestellung und Evaluationsrahmen

Die Fragestellung wird in den meisten Studien zumindest bezüglich der untersuchten Technologien und der Zielgrößen klar definiert. Eine explizite Definition der Zielpopulation fehlt teilweise.

Von der Einführung eines Studienberichts wird in der Regel eine Darstellung des medizinischen und ökonomischen Hintergrunds der untersuchten Thematik sowie eine Beschreibung des situativen Kontextes und des Designs der Studie erwartet. Dargelegt werden müssen eine eindeutige Fragestellung sowie die evaluierte Technologie.^{41,48,138} Dabei sollte die Technologiebeschreibung ausreichend detailliert sein, um eine Einschätzung der Übertragbarkeit der Studienergebnisse zu erlauben.⁴⁸

Im allgemeinen wird das medizinische Problemumfeld der Untersuchungen zum PSA-Screening hinreichend umrissen. Die ökonomische Bedeutung des Problemkontextes dagegen, d.h. die Implikationen des Ressourceneinsatzes und deren Relevanz für den potentiellen Adressaten der Untersuchung werden nur in einigen Studien dargelegt.^{8,81,83,91} Aufgrund der mangelhaften Einbindung der Studien in einen konkreten Entscheidungskontext wird in vielen Studien die Wahl der Vergleichstechnologien nicht ausreichend begründet und in den meisten Studien fehlt eine explizite und eindeutige Nennung der Perspektive. Obwohl die Nennung der Perspektive durchgehend in der Fachliteratur zur Methodik von ökonomischen Evaluationen als Bedingung erhoben wird,^{20,38,43,48,53,76,77,87,126,138} wird es von den meisten Studien versäumt, dieser Anforderung zu entsprechen.^{17,151}

In den meisten Studien werden zwar die untersuchten Technologien bzw. die sich anschließenden Therapiemaßnahmen explizit genannt, jedoch fehlen insbesondere in den Synthesestudien weiter präzisierende Angaben zu den verwendeten Assays,

zur Art der Biopsie (z.B. kontrolliert oder systematisch, Aspirations- oder Stanzbiopsie), zu den tangierten Versorgungsstrukturen, zum Qualifikationsgrad des ärztlichen Personals oder zu den Ausführungsarten der in Betracht gezogenen Therapien.

In allen Studien wird - explizit oder implizit - das in Betracht kommende Alter der Zielpopulation genannt. Auch wenn es in der vorliegenden Qualitätsbewertung deswegen zu keiner Abwertung kam, sei darauf hingewiesen, dass in den Synthesestudien generell präzisierende Angaben zu den Studienpopulationen und dem Setting der Untersuchungen fehlen, auf deren Daten zurückgegriffen wurde. Hier wären u.a. Hinweise auf das Alter, die klinische Vorgeschichte, den Rekrutierungsmodus oder die Versorgung im allgemeinen bzw. im urologischen Praxisrahmen von Interesse.^{61,68,71,72}

Mit einer Ausnahme⁹¹ wird in allen Studien lediglich das Prävalenz-Screening in Betracht gezogen, wobei der Zeithorizont kurzfristig im Rahmen der Studie einen Querschnittsmoment darstellt oder langfristig den 5-Jahres-Zeitraum nach der Karzinom-Diagnose¹¹⁰ bzw. das Absterben der Kohorte^{8,49,81,84} umfasst. Dieser Zeithorizont geht zwar im allgemeinen transparent aus den Studien hervor, die methodisch-inhaltliche Qualität wird jedoch insbesondere dadurch abgewertet, dass das Prävalenz-Screening als problematisch und hinsichtlich der Zielsetzung, das Screening im allgemeinen zu evaluieren, als unzureichend zu erachten ist. Hierzu sei auf die Screening-Kriterien von Wilson & Jungner (1968)¹⁶¹ verwiesen und das Beispiel von Neuhauser & Lewicki (1975)¹⁰⁶ im Zusammenhang mit den bei wiederholten Screening-Untersuchungen ermittelbaren inkrementellen Kosten-Effektivitäts-Quotienten in Erinnerung gerufen.

Die genannten Einschränkungen der Studienqualität in den Bereichen Fragestellung und Evaluationsrahmen erschweren Aussagen zur Übertragbarkeit der Studienergebnisse auf einen anderen Entscheidungskontext mit entsprechenden Bedingungen und Anforderungen. Auffällig ist außerdem, dass in mehr als der Hälfte der Studien der Evaluationstyp nicht ausdrücklich bezeichnet wird. Dies dürfte sich auch bei Literaturrecherchen mit enger gefasster Suchstrategie als die für diese Arbeit gewählte - beispielsweise bei einer Begrenzung der Suchstrategie auf Schlagwörter wie "cost(-)effectiveness" oder bei Verwendung der "cost-benefit analysis"-MeSH-Indexierung¹⁷ - negativ auf die Trefferquote auswirken.

C.6.2.3 Analysemethoden und Modellierungen

Untersuchungen zur Früherkennung ziehen oft Modelle heran, um optimale Teststrategien zu identifizieren sowie um Daten zu extrapolieren oder generieren.¹²⁶ Die Nachvollziehbarkeit bei den berücksichtigten Studien ist in dieser Hinsicht zum Teil eingeschränkt, dies vor allem aufgrund nicht ausreichender Kostenaufführungen und auch aufgrund der relativ zahlreich identifizierten Rechenfehler. Insbesondere irritierend sind hier u.a. die beachtlichen Inkongruenzen der Zwischen- und damit auch Endergebnisse bei den ansonsten seriös durchgeführten Analysen von Barry et al. (1995)⁸ und Coley et al. (1997)³⁰, welche auf den gleichen Daten und den gleichen Annahmen beruhen. Eine diesbezügliche Anfrage mit der Bitte um Erklärung von Seiten der Autoren blieb unbeantwortet. Der Unterschied zwischen Transparenz und methodisch-inhaltlicher Qualität der zehn ökonomischen Studien zum PSA-Screening, welche Modellierungen einbezogen haben, 8,37,49,52,81,83,84,91,93,110 wird insbesondere durch die mangelnde, sich auf die Aussagekraft der Studien negativ auswirkende Validität der für die Modelle herangezogenen Daten bedingt. Keiner der herangezogenen unsicheren Modellparameter basiert auf einem RCT oder einer systematischen Meta-Analyse bzw. alternativ auf Bayes-Ansätzen.⁴⁸

C.6.2.4 Gesundheitseffekte

Von entscheidender Bedeutung für die Beurteilung der Qualität einer gesundheitsökonomischen Evaluationsstudie und die angemessene Einschätzung der Verwendbarkeit ihrer Ergebnisse sind Angaben zur Evidenz der medizinischen Effektivität der Intervention.^{48,126,132} Auch eine Analyse, welche vor dem Vorliegen adäquater Effektivitätsstudien durchgeführt wird (sog. "was wenn"-Analyse⁴⁸), kann mit nützlichen Informationen zur Entscheidungsfindung beitragen, allerdings müssen dann die Ergebnisse im Licht der in der Evaluation getroffenen Annahmen interpretiert werden.⁴⁸ Ist die Evidenz für die Effektivität einer Intervention widersprüchlich oder ohne Beweiskraft, müssen die Studienergebnisse gleichermaßen besonnen bzw. zurückhaltend interpretiert werden. Aus diesem Grunde sollte eine eingehende Diskussion der Gesundheitseffekte wesentlicher Bestandteil der Darstellung einer Studie sein.⁴⁸

Sowohl die Originalstudien als auch die Synthesestudien zum PSA-Screening schneiden, wenn auch aus unterschiedlichen Gründen, hinsichtlich der Methodik der

Effektermittlung sehr negativ ab. Die Qualität der Studiendesigns bei den Originalstudien ist variabel, jedoch im allgemeinen gering. Qualitätsmindernd wirkt sich u.a. der selection bias bei der Rekrutierung aus: nur in einem Fall⁵² wurde eine Randomisierung vorgenommen. Insbesondere die Nichtanwendung der Biopsie als Goldstandard bei allen Studienteilnehmern oder zumindest einem randomisierten Teil der Studienpopulation^{27,69} mit dem dabei resultierenden work-up bias stellt prinzipiell alle Angaben zu Sensitivität, Spezifität, positivem und negativem prädiktiven Wert in Zweifel. In keiner dieser Studien wird eine Kontrollgruppe mitgeführt, so dass die Entdeckungsraten nur einen begrenzten Informationswert haben. Der positive prädiktive Wert in diesen diagnostischen Studien repräsentiert nicht den Stellenwert der PSA-Bestimmung allein auf, vielmehr spiegelt er die Biopsie-Kriterien, die Compliance mit einer Indikation zur Biopsie-Abklärung sowie die Genauigkeit der Biopsie-Methode wider. Damit ist kein Studiendesign hinreichend solide, um Folgerungen bezüglich der Wirksamkeit oder Effektivität der Karzinomentdeckung mittels der PSA-Bestimmung stützen zu können.

Bei den Synthesestudien, die nach den Kriterien für systematische Reviews beurteilt wurden,^{69,111,128} fehlen in den meisten Fällen Angaben zu Quellen und Suchstrategien der Literaturrecherche, Aussagen zur Validität bzw. Qualität der herangezogenen Studien bzw. zu Methoden der Relevanzbewertung. Die Datenlage, auf die zurückgegriffen werden muss, ist mangelhaft und insbesondere bezüglich der Behandlungseffektivität bei der Frühentdeckung eines Prostatakarzinoms von beschränkter Aussagekraft.^{4,26,51,129,139,149,155} Die Ergebnisse der Untersuchungen stehen damit in Abhängigkeit von den getroffenen Annahmen und es ist zu prüfen, welche Abstriche bei ihrer Interpretation zu machen sind. Das Schicksal von Individuen mit falschnegativem Testbefund wird lediglich in einer Untersuchung⁸¹ rechnerisch extrapoliert.

Bei den vier Studien, in welchen Gesundheitszustände mit Präferenzen bewertet werden, beruhen in einem Fall⁸¹ die Bewertungsdaten auf den Präferenzen von Medizinern, in den drei anderen^{49,91,93} auf Literaturangaben ohne nähere Spezifizierung. Da angesichts der mangelhaften Datenlage zur Früherkennung beim Prostatakarzinom wiederholt auf die Bedeutung der aufgeklärten individuellen Entscheidung zur Durchführung einer PSA-Bestimmung hingewiesen wird,^{32,37,116} erscheint es angebrachter, in ökonomischen Studien vermehrt auch die Präferenzen von Patienten hinsichtlich definierter Gesundheitszustände in die Evaluation einzubeziehen.

C.6.2.5 Kosten

In allen veröffentlichten ökonomischen Studien zum PSA-Screening werden Produkte zur Erstellung des Mengengerüsts verwendet, in fünf^{1,25,29,52,74} Untersuchungen erfolgt eine Kombination mit der Mengenbestimmung nach Erstellungsaufwand (z.B. Personalkosten).

Keine der Studien hat alle relevanten Kostenbereiche berücksichtigt. Kosten für die Implementierung und Organisation des Screening-Programmes wurden in keiner der evaluierten gesundheitsökonomischen Studien eingeschlossen. Diese beinhalten u.a. neues Personal, Ausstattung, Schulungskosten, Aufklärungsarbeit und Öffentlichkeitsarbeit. Bei Nicht-Berücksichtigung dieser Kosten kommt es zu einem systematischen Fehler zugunsten des Screenings. Nur zwei Studien^{52,74} bezogen die Rekrutierung der Teilnehmer für das Screening-Programm mit ein.

Alle aufgeführten Studien berücksichtigten die direkten medizinischen Kosten, die für die reine Screening-Untersuchung einschließlich der in den jeweiligen Strategien enthaltenen konfirmatorischen Diagnostik anfielen. Für die reine Screening-Untersuchung waren dies hauptsächlich personelle und Materialkosten für die Durchführung des PSA-Screening-Tests, je nach Strategie parallel oder im Rahmen der weiterführenden Diagnostik sequentiell mit der digitalen rektalen Untersuchung und/oder der transrektalen Ultraschalluntersuchung. Die konfirmatorische Diagnostik bestand aus einer ultraschallgesteuerten Biopsie einschließlich Histopathologie und teilweise einem Staging. Nur zum Teil wurde explizit angegeben, ob eine Antibiotikaphylaxe vor der Biopsie in die Berechnungen einbezogen wurde.

Die ökonomischen Konsequenzen von falschpositiven bzw. nicht eindeutig bestimmbar Testbefunden werden - mit Ausnahme der Studie von Gustafsson et al. (1995)⁵² bei den nicht eindeutig bestimmbar Tests - durchweg ignoriert. In Anbetracht von Empfehlungen wie die im Auftrag der Deutschen Krebsgesellschaft, der Deutschen Krebshilfe und der Deutschen Gesellschaft für Urologie herausgegebenen Standards¹⁵⁸, führt ein erhöhter PSA-Wert bei negativer Biopsie fast immer zu Kontrolluntersuchungen und zu Biopsie-Wiederholungen. Da die falschpositiven Testbefunde im Vergleich zu den richtigpositiven überwiegen, können die ökonomischen Folgen intensiver Interventionsweisen nach einem positiven ersten Test potentiell sehr bedeutend sein und sollten in einer vollständigen Evaluation berücksichtigt werden. Bei Nicht-Berücksichtigung dieser Kosten ist mit einem systematischen Fehler zugunsten des Screenings zu rechnen.

Die Kosten für die primäre, d.h. initiale Therapie identifizierter Karzinome (Behandlung identifizierter Karzinome, Behandlung fortgeschrittener oder metastasierender Erkrankung) ging in neun von insgesamt 16 Studien ein. Die Angabe der Behandlungsverfahren lässt in den meisten Fällen erkennen, dass dabei auch fortgeschrittene Erkrankungsstadien einbezogen wurden, bezüglich der Behandlung von Metastasen wurden allerdings im allgemeinen keine expliziten Angaben gemacht. Die beiden Studien von Littrup et al. (1993, 1994c) berücksichtigten Kosten für die Primärtherapie nur in der Kosten-Nutzen-Analyse, nicht aber in der Kosten-Effektivitäts-Analyse. Insgesamt wurden die einzelnen Behandlungsverfahren nur teilweise explizit genannt, wie z.B. radikale Prostatektomie, Strahlentherapie, endokrine Therapie, Orchiektomie.

Die Genauigkeit der Kostenerhebung für die weitergehende Behandlung ist unterschiedlich. Bei allen Studien, welche die Therapie mit einbeziehen, werden Behandlung und Behandlungskosten in Abhängigkeit vom klinischen Stadium bestimmt. In der Praxis jedoch schließen Therapieentscheidungen andere Informationen (z.B. zur Komorbidität und dem Differenzierungsgrad)^{148,158} mit ein. Im Rahmen einer RCT erhobene Kosten können hier plausiblere Schätzungen der tatsächlichen Behandlungskosten sein. Adjuvante Therapiemaßnahmen werden in keiner der vorliegenden Untersuchungen berücksichtigt.

Die Kosten der sekundären Therapie, d.h. die Behandlung von Komplikationen im Rahmen der Primärtherapie, wurde nur noch in sieben Fällen berücksichtigt, in der Studie von Abramson et al.¹ war diese Information nicht aus der Publikation zu entnehmen. Allerdings machen nur Krahn et al.⁸¹ in ihrer Publikation genauere Angaben zu diesem Bereich, sie nannten in einer tabellarischen Preisgerüstaufstellung die folgenden Komplikationen: Kurzzeitkomplikationen der primären Behandlung, Langzeitbehandlung von Komplikationen (partielle/komplette Impotenz, partielle/komplette Inkontinenz, Urethraobstruktion).

Nur zwei Untersuchungen versuchen explizit, die Kosten für die weitere Versorgung der Tumorpatienten, d.h. die Kosten für die Therapie bei Rezidiven bzw. Progression zu erfassen.^{8,81} Dabei wird der posttherapeutischen Nachsorge nur in der Studie von Krahn et al.⁸¹ umfassend Rechnung getragen. Krahn et al. bezogen Kosten für folgende medizinische Leistungen mit ein: posttherapeutische Nachsorge (nach radikaler Prostatektomie: jährliche Konsultation beim Urologen und PSA), Behandlung eines Lokalrezidivs (PSA, TRUS, Biopsie, zwei Konsultationen), Nachsorge des Lokalrezidivs (PSA, Konsultation alle drei Monate), Diagnostik bei

Verdacht auf Metastasierung (PSA, Knochenszintigramm, zwei Konsultationen), Nachsorge Metastasierung (Konsultationen, Blutabnahme, Medikation).

Keine der Studien analysiert die Kosten, die anfallen, wenn aufgrund gewonnener Lebensjahre andere zu behandelnde Erkrankungen auftreten. Inwieweit Personalkosten in den verwendeten Kosten oder Abrechnungsbeträgen enthalten sind, bleibt meist unklar. Nur in eine Studie⁵² werden Fixkosten für Ausstattung und Räumlichkeiten in die Analyse eingebracht.

Direkte nichtmedizinische Kosten werden weitgehend nicht untersucht. Nur eine Studie⁵² berücksichtigt die Transportkosten zum Ort der Screening-Untersuchung. In Abhängigkeit von der Screening-Strategie schwankt der Anteil dieser Kosten an den direkten Gesamtkosten der Karzinomdiagnose zwischen 17% und 44%.

Produktivitätsausfallskosten werden zwar in die Kosten-Nutzen-Analysen von Littrup et al. einbezogen, jedoch wird nicht spezifiziert, für welche Situationen und Zeiträume mit einem kompletten oder partiellen Arbeitsausfall gerechnet wurde. In keiner der Kosten-Effektivitäts-Analysen werden Produktivitätsausfallskosten aufgrund der Behandlung, der Langzeitmorbidity oder der Frühmortalität berücksichtigt. Lediglich in einer Studie⁵² werden die Zeitkosten für Reisen und Warten im Zusammenhang mit der Screening-Untersuchung selbst in die Analyse aufgenommen. In dieser Studie wurden anhand von Probandenfragebögen der untersuchungsbedingte Zeitaufwand und Daten zur Berufstätigkeit erhoben und daraus Einkommensverluste berechnet.

Intangible Kosten wurden nur in den beiden Kosten-Nutzen-Analysen geschätzt, in den Kosten-Nutzwert-Analysen wurde z.T. darauf verwiesen, dass intangible Kosten auf der Seite der Gesundheitseffekte berücksichtigt wurden.

Zur Bewertung des Mengengerüsts wird meist auf administrative Preise zurückgegriffen. Eine Studie⁸¹ setzt betriebliche Selbstkosten als Schattenpreise ein, eine weitere verwendet ein pauschales Verfahren, bei dem generelle Durchschnittspreise pro Leistungseinheit ermittelt werden.⁵² Weder die Methoden zur Ermittlung der Mengengerüste noch die Methoden zur Ermittlung der Preise werden in der Regel transparent dargestellt. Das Referenzjahr für die Preise wird in über der Hälfte der Studien nicht dokumentiert. Damit wird die Auslegung der Kosten-Effekt-Relation (Kosten-Effektivitäts-Quotient, Kosten-Nutzwert-Quotient bzw. Nutzen-Kosten-Differenz) wie auch der Vergleich der Ergebnisse mit denen anderer Studien erschwert. In keiner der Studien, in welchen langfristige Kosten anfallen

können, werden Preisanpassungen für Inflation bzw. Deflation vorgenommen, womit die Preisentwicklung von Technologien ignoriert wird.⁴⁸ Die Darstellung der Mengen- und Preiskomponenten ist nur in knapp der Hälfte aller berücksichtigten Studien so ausreichend detailliert ausgewiesen, dass im nachhinein mit anderen ökonomischen Parametern eine Umrechnung und Anpassung der Evaluationsresultate an andere ökonomische Gegebenheiten erfolgen könnte.

C.6.2.6 Diskontierung

Lediglich in einem Viertel^{8,81} der Studien, in welchen eine Diskontierung von Kosten und Effekten in Betracht kam, wurde auch eine Diskontierung (mit einer Diskontrate von 5%) durchgeführt. Transparenzeinschränkend wirkt sich in diesem Themenbereich auch die fehlende Nennung eines Referenzjahres für die Kosten aus.

C.6.2.7 Ergebnisse und Ergebnispräsentation

Technische Modellvalidierungen dienen der Feststellung der Korrektheit von Programmierung und Modellberechnungen.^{48,132} In keiner der Modellierungsstudien wurden solche Maßnahmen berichtet. Dies mag zunächst als nicht qualitätseinschränkend erscheinen, allerdings weisen die zahlreichen, auch in Studien ohne Modellierung festgestellten, numerischen Inkonsistenzen auf eine Problematik bei der Programmierung von Tabellenkalkulationsprogrammen bzw. beim Einsatz von Taschenrechnern hin. Dies stellt die Glaubwürdigkeit der Ergebnisse in Frage, sollte doch eine einfache Überprüfung der Berechnungen bei jeder Studiendokumentation erwartet werden können. Insofern wäre im Zusammenhang mit dem Kriterienkatalogen zu überlegen, ob das Kriterium der Modellvalidierung nicht auch gerade für Studien ohne Modellierung angewendet werden sollte.

Hervorzuheben ist, dass die Betrachtung inkrementeller Effekte und Kosten nicht die Regel ist. Werden Durchschnittswerte angegeben, sollte man sich vergegenwärtigen, dass auch bei den impliziten Strategien "Kein Screening" oder "Status quo" nur selten Kosten und Nutzen in Höhe von Null resultieren. Das Nebeneinander von Durchschnittswert- und inkrementeller Betrachtung trägt zusätzlich zur Heterogenität der Studien und Studienergebnisse bei. Inwieweit die zu beobachtende Spannweite der Ergebnisse zu den Kosten-Effekt-Relationen auf methodische und

konzeptionelle Unterschiede zurückgeführt werden kann oder tatsächlich durch Effizienzdifferenzen bedingt ist, bleibt unklar. Aus den Resultaten der Basis-Analysen lassen sich lediglich folgende übergreifende Ergebnisse ableiten:

- Die Entdeckungsrate von asymptomatischen Prostatakarzinomfällen aller Stadien schwankt für die verschiedenen Screening-Strategien zwischen 1% und 5%. Beim Inzidenz-Screening ist mit einer Abnahme der Entdeckungsraten zu rechnen.
- Der Gewinn an Restlebensdauer pro Screening-Teilnehmer im Vergleich zur Alternative "Kein Screening" liegt je nach Altersgruppe im Bereich von Tagen bis zu wenigen Wochen (ohne Diskontierung), der Gewinn an Restlebensdauer pro behandeltem Prostatakarzinom im Bereich von Jahresfraktionen. Beide Ergebnisse sind abhängig von den Annahmen zur Behandlungseffektivität, die in den vorliegenden Studien das Screening begünstigten.
- Die in DM konvertierten Kosten pro Screeningteilnehmer schwanken zwischen 28 DM und 1 806 DM bei durchschnittlicher bzw. zwischen 87 DM und 988 DM bei inkrementeller Betrachtungsweise. Pro identifiziertem Krebsfall liegen die Durchschnittskosten zwischen 1 202 DM und 18 419 DM. Die inkrementellen Kosten pro identifiziertem Krebsfall werden mit 6 545 DM bis 18 579 DM angegeben.
- Die Kosten pro gewonnenem Lebensjahr schwanken je nach durchschnittlicher und inkrementeller Analyse und je nach berücksichtigten Kostenarten über drei Zehnerpotenzen. So wird in der Studie von Optenberg & Thompson umgerechnet ein durchschnittlicher Wert von 131-325 DM¹¹⁰ pro gewonnenem Lebensjahr errechnet, während Krahn et al.⁸¹ je nach Alter und Strategie zu einem inkrementellen Ergebnis von 127 690 DM bis 823 770 DM pro gewonnenem Lebensjahr kommen.
- Bei einer Adjustierung der Restlebenserwartung nach der Lebensqualität ist insbesondere im höheren Alter mit einem Verlust an QALYs im Vergleich zur Alternative "Kein Screening" zu rechnen, so dass eine dominante Entscheidungssituation herrscht, die unabhängig von der Höhe der Mehrkosten durch das Screening gegen ein Screening ausfällt.
- Die Kosten der Früherkennung werden weniger von den Screening-Tests an sich, als von der sich anschließenden diagnostischen Abklärung (TRUS und Biopsien), dem Staging und den Behandlungsmaßnahmen verursacht. Eine

übergreifende Evidenz zu Nachsorgekosten, Produktivitätsausfallkosten und Kosten für die Implementierung von PSA-Screening-Programmen besteht nicht. Auch zu den Kosten für die Behandlung anderer Erkrankungen aufgrund zusätzlicher Lebensjahre gibt es keine Aussagen.

- Bei der Durchschnittswertbetrachtung wie bei der inkrementellen Betrachtung weisen die effektivsten Screening-Strategien jeweils den größten - kurzfristigen oder langfristigen - Kosten-Effektivitäts-Quotienten bzw. Kosten-Nutzwert-Quotient auf. Der Kosten-Effektivitäts-Quotient steigt mit zunehmendem Alter aufgrund von höheren Diagnose- und Behandlungskosten infolge höherer Prävalenz, höherer Komorbidität und höherer Komplikationsraten. Ferner sinkt die Kosten-Effektivität mit zunehmendem Alter aufgrund einer geringeren Erwartung einer klinischen relevanten Auffälligkeit bzw. Beeinträchtigung in der Restlebensdauer.

C.6.2.8 Behandlung von Unsicherheiten

Evaluationsstudien stellen immer nur Schätzungen eines Interventionsaufwands und -ergebnisses dar, welche aus erhebungstechnischen oder -ökonomischen Gründen teils mit einer erheblichen Unsicherheit behaftet sind.^{76,87} Der Informationswert einer Studie liegt vor allem darin, die Vor- und Nachteile eines Verfahrens möglichst deutlich darzustellen und diejenigen Faktoren zu identifizieren, von denen die medizinischen und ökonomischen Konsequenzen der verglichenen Alternativen maßgeblich abhängen.¹²⁶ Um als Entscheidungshilfe herangezogen werden zu können, ist eine Prüfung der Robustheit der ermittelten Aufwand-Ertrags-Relationen durch eine Sensitivitätsanalyse erforderlich. Dabei kann z.B. der Einfluss veränderlicher Preisstrukturen geprüft werden, die Bedeutung veränderter Prävalenzen und Inzidenzen, der Einfluss veränderlicher Komplikationsraten, die Auswirkungen modifizierter Sensitivität und Spezifität der verwendeten Diagnoseverfahren oder die davon ausgehenden Auswirkungen auf die Indikationsstellung. Auch die quantitative Bewertung verschiedener Gesundheitszustände bezüglich ihrer Lebensqualität ist einer Sensitivitätsanalyse zu unterziehen.

In den vorliegenden ökonomischen Studien zum PSA-Screening sind im allgemeinen keine hinreichend umfassenden Sensitivitätsanalysen durchgeführt worden. Die verbreitetste Praxis ist das Berichten von 1-Weg-Sensitivitätsanalysen für eine selektierte Gruppe von Parametern. Die Studie von Krahn et al. (1994)⁸¹ hebt sich

durch eine relevante strukturelle Sensitivitätsanalyse und klar dargestellte 1-Weg-Sensitivitätsanalysen für eine Reihe von Variablen hervor. Die Empfindlichkeit der Ergebnisse bezüglich der Annahmen zur Behandlungseffektivität insbesondere der radikalen Prostatektomie ist den Studien gemein, die hierzu Sensitivitätsanalysen durchgeführt haben.^{8,81,91,93} Im allgemeinen bleibt jedoch das Ausmaß der Unsicherheit unklar. Da keine der Untersuchungen auf einem randomisierten Design oder auf einer Meta-Analyse beruht, konnten auch keine stochastischen Methoden zur Abschätzung der Variation der Ergebnisse eingesetzt werden.^{48,108}

C.6.2.9 Diskussion der Autoren der untersuchten Publikationen

Wichtig in der Diskussion ist die Interpretation der gesundheitsökonomischen Resultate, wobei der Einfluss von wichtigen Annahmen, von der Datenqualität, von der Sensitivität der Ergebnisse gegenüber Schlüsselparametern und von Einschränkungen der Analyse auf das Ergebnis hervorgehoben und erörtert werden sollte.⁴⁸ Die von den Autoren geführte Diskussion in den vorliegenden Untersuchungen zum PSA-Screening beschränkt sich meist auf den Vergleich mit Ergebnissen anderer Untersuchungen, ohne dass auf deren Besonderheiten im Studiendesign oder Bezugsrahmen eingegangen wird. Bei Aussagen zu systematischen Fehlern oder Unsicherheiten wird in der Regel von einem Bias in Richtung Screening ausgegangen.^{8,81,110} Ethische und Verteilungsfragen werden grundsätzlich nie aufgeworfen. So wäre beispielsweise die Frage zu diskutieren, wie viele Screening-Teilnehmer, bei denen ein Prostatakarzinom nie klinisch relevant werden würde, welche Risiken eingehen, damit andererseits das Leben von Patienten mit progredientem Prostatakarzinom durch die Früherkennung verlängert wird. In den Studien, in denen die Ergebnisse in Relation zu anderen unabhängigen Gesundheitsprogrammen - in der Regel dem Screening beim Mammakarzinom - gesetzt werden, werden so gut wie keine Aussagen über den zeitlichen Bezug, die Bezugsalternativen der Effektivität oder Methoden der Datenerhebung gemacht bzw. diese kritisch betrachtet.

C.6.2.10 Schlussfolgerungen der Autoren der untersuchten Publikationen

Zwei der Autoren sprechen sich auf der Basis der jeweiligen Datenlage gegen das Screening aus.^{81,110} Ein weiterer rät von einer Biopsie allein aufgrund der PSA-Bestimmung ab.⁴⁹ Fünf Autoren legen weiteren Klärungsbedarf nahe.^{1,8,25,37,83} Vier

Autoren erachten das Screening als realisierbar bzw. empfehlen es,^{11,29,34,74} während vier weitere auf die bei Früherkennungsprogrammen zu beachtende unterschiedliche diagnostische Kosten-Effektivität verschiedener Screening-Strategien hinweisen.^{52,84,91,93} Das Spektrum der Schlussfolgerungen spiegelt damit die Standpunkte in der derzeitigen Kontroverse bezüglich der Früherkennung beim Prostatakarzinom wider. Dabei lassen sich die Folgerungen lediglich in etwas mehr als der Hälfte der Studien aus den berichteten Ergebnissen ableiten, subjektive Auffassungen gehen damit in die Ergebnisinterpretation mit ein. Zwischen der Qualität der Studie - ausgedrückt durch den Gesamtscore wie auch einzelne Summenscores der methodisch-inhaltlichen Qualität - und der Richtung der Schlussfolgerung ist kein systematischer Zusammenhang ausfindig zu machen. Die beiden vom Gesamtscore her qualitativ hochwertigsten Studien - die Studie von Barry et al. (1995)⁸ und diejenige von Krahn et al. (1994)⁸¹ - legen folgendes nahe:

- zum einen weiteren Klärungsbedarf, insbesondere bezüglich der Behandlungseffektivität, sowie eine Ablehnung der Aufnahme der PSA-Bestimmung in den Leistungskatalog von Medicare

"...This analysis of the estimated risks, benefits, and costs of DRE and PSA screening highlights the uncertainty surrounding this topic, particularly regarding the magnitude of the future hazards represented by the types of prostate cancer that will be detected with DRE and PSA and of the effectiveness of treatment at reducing these risks of future metastasis and cancer death..."

"...On the one hand, the concept of early detection of prostate cancer with DRE and PSA cannot be dismissed out of hand, because, if the multiple favorable assumptions used in this analysis are true, and if quality-of-life implications of impotence and incontinence are assumed to be inconsequential, such screening might be as cost-effective as other commonly performed strategies for early detection of cancer among Medicare patients, such as testing stool occult blood or mammography, at least for men age 65 and 70 years. On the other hand, subjecting 15% to 40% of men to a recommendation for prostate biopsy, inflicting incontinence and impotence on large numbers of men, and spending large numbers of health care dollars in an era where growth in health care costs needs to be constrained seems irresponsible in the absence of controlled data documenting that the benefits of early detection outweigh these risks and costs..."

"...From Medicare's perspective, the policy question about whether to provide a benefit for PSA screening depends on whether the large early investment needed for such a program (some of which will be consumed anyway in terms of treatment costs even of men continue to pay for their own tests) is warranted at a time when a legitimate question can be raised about whether such screening does more harm than good. We believe that it is premature to offer a Medicare benefit for PSA testing for early detection of prostate cancer." (Barry et al., 1995⁸)

- zum anderen eine Ablehnung des Screenings mittels der PSA-Bestimmung, DRU und TRUS auf der Grundlage der bestehenden Evidenz

"...Our analysis does not support using PSA, TRUS, or DRE to screen asymptomatic men for prostatic cancer. Screening may result in poorer health outcomes and will increase costs dramatically. Assessment of comorbidity, risk attitude, and valuation of sexual function may identify individuals who will benefit from screening, but selecting high-prevalence populations will not improve the benefit of screening." (Krahn et al., 1994⁸¹)

C.6.3 Überlegungen zur Übertragbarkeit der Ergebnisse nach Deutschland

Keine der einbezogenen Studien wurde in Deutschland bzw. unter Berücksichtigung von aus Deutschland stammenden Daten durchgeführt. Die berücksichtigten gesundheitsökonomischen Studien zum Thema "PSA-Screening" schließen Datenmaterial aus USA, Kanada, Großbritannien, Italien, Schweden und Frankreich ein. Dabei wurde teils multizentrisch vorgegangen und Daten aus verschiedenen Populationen wurden gemeinsam analysiert, teils wurden medizinisch-epidemiologische Daten einer Population mit ökonomischen Daten einer anderen Population in der Analyse zusammengeführt.

Die Heterogenität der den einzelnen Studien zugrunde liegenden Grundgesamtheiten und Gesundheitssystemen führt zu der Frage, inwieweit diese Ergebnisse einen relevanten Beitrag zur Entscheidungsfindung innerhalb des deutschen Gesundheitssystems leisten können. Dabei müssen zweierlei Aspekte berücksichtigt werden:

- (1) die Frage danach, ob die finalen medizinischen oder gesundheitsökonomischen Studienergebnisse bzw. Ergebnisparameter (z.B. Anzahl der gewonnenen Lebensjahre pro Screening-Teilnehmer, Kosten-Effektivitäts-Quotient) übertragen werden können oder nur Teilergebnisse bzw. intermediäre Ergebnisparameter (z.B. Anteil der zu erwarteten Biopsien, Kosten für die Rekrutierung der Screening-Teilnehmer).
- (2) die Frage nach der internen Validität als notwendige Voraussetzung für die Generalisierbarkeit und die Übertragbarkeit der Studienergebnisse nach Deutschland.

Nach O'Brien¹⁰⁷ kann ein Studienresultat nur übertragen werden, wenn die Verhältnisse bezüglich folgender sechs Bereiche weitgehend gleich sind: Demographie und Epidemiologie der Erkrankung, klinische Praxis und Konventionen, Anreizsysteme und Regulationsmechanismen für Leistungserbringer im Gesundheitssystem, relative Preise, Verbraucherpräferenzen und Opportunitätskosten.

Zunächst wären bei der Beurteilung der Übertragbarkeit einer medizinischen Technologie technische und klinische Parameter zu prüfen wie im Falle von PSA-Screening Faktoren wie Laborqualität, Einhaltung der Einflussfaktoren auf das PSA etc. Allerdings erscheinen diese Aspekte vor dem Hintergrund der globalen Übertragungsprobleme von einer Nation auf die andere nachrangig.

Bedeutende klinisch epidemiologische Faktoren, von deren Übertragbarkeit nicht ohne weiteres ausgegangen werden kann sind die Operationsqualität und die Behandlungserwartung, welche sich wiederum beide auf die Restlebenserwartung und die gesundheitsbezogene Lebensqualität auswirken können. Eine systematische Prüfung dieser Parameter erscheint jedoch zum gegebenen Zeitpunkt aufgrund mangelnden Wissens zur evidenten Abweichungen bzw. fehlender deutscher Daten nicht durchführbar. Aus diesem Grunde sollten solche Punkte nur bei berechtigten Zweifeln an der Übertragbarkeit klinisch-epidemiologischer Parameter für die Ablehnung oder Annahme einer Technologie herangezogen werden. Wie die Erfahrung zeigt, führt das Warten auf systematische Übertragbarkeitserkenntnisse bzw. gute deutsche Studien eher zu einem Entscheidungs nihilismus. Aus diesem Grunde ist beim Fehlen von inhaltlichen Gegenargumenten zu erwägen, ob bis zum Vorliegen deutscher Daten (zunächst) die aus der Literatur vorliegenden Ergebnisse zur medizinischen Effektivität als Anhaltspunkte auch für Deutschland verwendet

werden sollten. Optenberg¹¹⁰ weist allerdings darauf hin, dass die Ergebnisse des Prävalenz-Screenings nicht auf ein Follow-up-Screening übertragbar sind.

Ein im Rahmen des BMG-Projekts durchgeführter Bericht zur Effektivität des PSA-Screenings beim Prostatakarzinom¹¹⁶ kommt zu dem Schluss, dass die Übertragbarkeit der Effektivität des PSA-Screenings anders als für andere medizinische Teilbereiche für die deutsche Situation relativ gut gegeben ist, da die Voraussetzungen für ein Prostatakarzinom-Screening aus struktureller Sicht in Deutschland genauso wie in den anderen westlichen Ländern bestehen.

Anders ist dies bei den Kosten und der Kosten-Effektivitäts-Relation. Eine Beurteilung der Übertragbarkeit der ökonomischen Parameter wie z.B. der relativen Preise oder der Kosten pro Screeningteilnehmer gestaltet sich problematisch durch die Tatsache, dass in den untersuchten Studien die Erstellung des Mengengerüsts vorwiegend nach Aufwand der Leistungserstellung (Kostenarten) oder nach erstellten Zwischenprodukten (Versorgungsmengen) erfolgte und keine separate Spezifikation der einzelnen Mengen mit relativen Preisen stattfand.

Bei der Beurteilung der Übertragbarkeit der Kosten-Effektivitäts-Relation einer Technologie ist die Validität beider - der medizinischen und der ökonomischen - Ergebnisparameter bezüglich der Population, auf die übertragen werden soll, zu prüfen.

Gustafsson et al.⁵² bemerken in ihrer gesundheitsökonomischen Evaluation des PSA-Screenings, dass zwar die absoluten Werte für Kosten und Effektivität nicht ohne weiteres auf externe Populationen übertragbar sind, dass jedoch eine Übertragbarkeit der Kosten-Effektivitäts-Ergebnisse auf die westliche Welt durchaus sinnvoll ist, da die Quotienten-Maßzahlen wie die Kosten-Effektivitäts-Ratio populationsübergreifend relativ stabil sind.

Littrup et al.⁹³ kommen aufgrund einer Sensitivitätsanalyse zu dem Schluss, dass zwar die Kosten für DRU, PSA, TRUS und Biopsie zwischen verschiedenen Ländern variieren, dass jedoch eine stabile Rangfolge der Strategien bezüglich der Kosten pro entdecktem Krebsfall zu erwarten sind.

Selbst bei der Restriktion auf die zahlreichen nordamerikanischen Studien zeigt sich eine große Heterogenität der Kosten-Effektivitäts-Quotienten. Dies weist auf das große Potential systematischer Fehler (Bias) und die daraus resultierende eingeschränkte interne Validität der Studienergebnisse hin. Erst wenn von einem relativ homogenen Studienergebnis innerhalb einer bestimmten Population und einer

hohen internen Validität ausgegangen werden kann, ist es sinnvoll, in einem zweiten Schritt Überlegungen zur Generalisierbarkeit über die Studienpopulation hinaus und zur Übertragbarkeit auf eine andere Population bzw. einen anderen gesundheitsökonomischen Kontext anzustellen. Deshalb wird in der folgenden Diskussion zur Übertragbarkeit das Problem der eingeschränkten internen Validität entsprechend berücksichtigt.

Bezüglich der internen Validität der evaluierten Ergebnisse herrschen folgende Probleme:

- Oft fehlt die Spezifikation der Zielpopulation, der eingesetzten Technologie, der Perspektive, der tangierten Versorgungsstrukturen.
- Die verwendeten Daten besitzen niederes Evidenzniveau, da randomisierte klinische Studien fehlen und keine adäquat durchgeführten Meta-Analysen zu allen Parametern vorliegen.
- Meta-Analysen für die Behandlungseffektivität fehlen.
- Die medizinische Evidenz ist umstritten.
- Wenige Aussagen zur Compliance, die eine wesentliche Bedingung für die erfolgreiche Durchführung einer Screening-Maßnahme darstellen.
- Fehlende Goldstandarduntersuchungen bei PSA-testnegativen Personen.
- Mangelnde Datenlage zu den Präferenzen und der Lebensqualität.
- Es fehlen ausreichende Aussagen zu den Kosten und der Kosten-Effektivität im zeitlichen Verlauf bei wiederholten Screening-Durchführungen.
- Nicht-Berücksichtigung vieler relevanter Kosten; insbesondere fehlen zumeist Rekrutierungs- und Implementierungskosten, Kosten für die systematische und langfristige Nachsorge, Kosten zum Follow-up von PSA-positiven Personen mit negativer Biopsie und die Produktivitätsausfallkosten.

Neben den genannten Faktoren der internen Validität beeinflussen die obengenannten externen Faktoren nach O'Brien die Übertragbarkeit von Studienergebnissen auf ein anderes Entscheidungsland.¹⁰⁷ Bezüglich der Generalisierbarkeit bzw. Übertragbarkeit der Studienergebnisse zum PSA-Screening nach Deutschland sind folgende Punkte problematisch:

- Selektionsfehler bei der Rekrutierung verhindern eine Übertragung auf deutsche Massen-Screening-Verhältnisse.
- In einigen Studien wurde innerhalb bestimmter Screening-Strategien auf spezialisierte Urologen zurückgegriffen. Eine systematische Prüfung, inwieweit solche Spezialisten in Deutschland für ein Massen-Screening zur Verfügung stehen, steht noch aus.
- Da anstelle eines detaillierten Mengengerüsts mit relativen Preisen meist nur Zwischenprodukte (Versorgungsmengen) angegeben werden, kann keine Kostenrechnung mit deutschen Preisen durchgeführt werden.
- Oft fehlt die für eine Anpassung der Kosten notwendige Nennung des Referenzjahres.
- Systematisch erhobene Präferenz- und Lebensqualitätsdaten für die Zielerkrankung Prostatakarzinom im deutschen Kontext fehlen.
- Die mangelnde Dokumentation und Nachvollziehbarkeit der entscheidungsanalytischen Modelle mit deren Parametern verhindern in den meisten Fällen eine Re-Analyse unter an Deutschland angepassten Gesichtspunkten.

Externe Faktoren, die die Übertragbarkeit von Studienergebnissen von einem Untersuchungsland auf ein Entscheidungsland beeinflussen, können sich im Laufe der Zeit verändern. Damit ist die Übertragbarkeit der Ergebnisse einer intern validen Studie eine variable Eigenschaft und stets innerhalb des aktuellen Entscheidungskontextes zu beurteilen. Schon deshalb ist es angezeigt, die Übertragbarkeit von Studienergebnissen getrennt von der internen Validität der Studienergebnisse zu beurteilen. Eine hohe Studienqualität dient allerdings als notwendige, nicht aber als hinreichende Bedingung für die Übertragbarkeit der Evidenz aus dem Ausland auf deutsche Verhältnisse. Auf die Problematik bei der Beurteilung der Möglichkeit eines Transfers von Studienergebnissen auf den Bezugsrahmen bzw. Entscheidungskontext des Nutzers einer Datenbank wurde bereits mehrfach hingewiesen.^{5,21,22,39,86,107,140}

Aus diesen Gründen haben sich die drei gesundheitsökonomischen Projektgruppen darauf geeinigt, die Prüfung der Übertragbarkeit von Ergebnissen ökonomischer Evaluationsstudien, wie sie im Rahmen dieses Kapitels vorgenommenen wird, nicht in einen standardisierten Kriterienkatalog aufzunehmen, sondern kontextabhängig und aktuell in einem sich an die Beurteilung der methodisch-inhaltlichen

Studienqualität anschließenden Schritt der Technologiebewertung zu behandeln. Dabei wurden allgemeine Probleme, Methoden und Lösungsansätze in Rahmen des Themas "Übertragbarkeit ausländischer Studienergebnisse auf Deutschland" von der Ulmer Projektgruppe erarbeitet, zusammengestellt und mit den weiteren ökonomischen Projektgruppen abgestimmt, so dass an dieser Stelle auf diesen Beitrag verwiesen werden kann.¹⁵⁹ In diesem Beitrag wird insbesondere die Übertragbarkeit ausländischer Studienergebnisse aus dem Bereich der ökonomischen Evaluation auf Deutschland aufgegriffen. Hier wird untersucht, ob bzw. unter welchen Umständen der Transfer von Ergebnissen mit hoher Wahrscheinlichkeit möglich ist und wie sich die Studienresultate über die Zeit und zwischen Währungssystemen anpassen lassen. Ferner weisen die Autoren darauf hin, dass zwischen drei Übertragungsebenen unterschieden werden muss:¹⁵⁹ (1) Übertragung aus dem Studienland in das Entscheidungsland (z.B. USA - Deutschland). (2) Übertragung aus einem Studiensystem in ein Entscheidungssystem (z.B. gesetzliche Krankenversicherung - private Krankenversicherung). (3) Übertragbarkeit des Studienkontextes auf den Entscheidungskontext (z.B. Patientenselektion in der Studie - Zielpopulation im Entscheidungskontext). Je nach Ebene, auf der übertragen werden soll, kann die Entscheidung, ob eine Übertragung möglich ist, unterschiedlich ausfallen.

Ein weiteres grundsätzliches Dilemma besteht in der Interpretation von Beobachtungsdaten und den Daten aus RCTs. Obwohl RCTs bzw. daraus resultierenden Meta-Analysen die höchste Evidenzstufen bezüglich der Bestimmung der Wirksamkeit einer medizinischen Technologie (engl. efficacy) besitzen, können diese Daten ungeeignet sein, die Effektivität unter Routinebedingungen (engl. effectiveness) zu bestimmen. Dieser Aspekt betrifft im besonderen Kosten-Effektivitäts-Studien, da diese Studien valide Langzeitschätzungen von Kosten-Effektivitäts-Relationen unter realistischen Bedingungen zum Ziel haben.¹³² Gründe für Zweifel an der Geeignetheit von RCT-Daten zur Bestimmung der Effektivität sind u.a. künstliche Bedingungen bezüglich der Teilnahmeraten, der Compliance, der Patientenselektion, der Validität diagnostischer Befunde und der Therapieerfolgs- bzw. Komplikationsraten. Meist herrschen im Rahmen von RCTs Rahmenbedingungen, die bezüglich der genannten Punkte günstiger sind als die zu erwarteten Bedingungen beim routinemäßigen und flächendeckenden Einsatz einer Technologie, was zu einer Überschätzung des gesundheitsbezogenen Effektes der untersuchten Technologie führen kann. Dasselbe gilt für kleinere prospektiv durchgeführte klinische Beobachtungsstudien, die bis auf die Randomisierung ähnlich günstige Bedingungen aufweisen können wie RCTs. Aus den genannten

Gründen wird immer wieder gefordert, die Ergebnisse aus RCTs zur Wirksamkeit den Ergebnissen großer Beobachtungsstudien zu Routineprozessen (large data bases) gegenüberzustellen, um Abweichungen erkennen und kritisch interpretieren zu können. Dies gilt insbesondere für ein Screening-Verfahren wie dem PSA-Screening, da hier Studienbedingungen und Routinebedingungen erheblich voneinander abweichen können.¹³²

Die genannten Aspekte sind bei der Übertragung von Studienergebnissen auf Deutschland zu berücksichtigen. Insbesondere bei der Übertragung der Wirksamkeitsergebnisse von PSA-Screening-Studien mit klinischen Originaldaten und kleinen Fallzahlen ist vorsichtig umzugehen, da hier der notwendige Zwischenschritt in der Kette Wirksamkeit im Ausland - Routineeffektivität im Ausland - Routineeffektivität in Deutschland fehlt. Geeigneter erscheinen in diesem Zusammenhang entscheidungsanalytische Modellierungsstudien,¹³² allerdings ist auch hier darauf zu achten, aus welchem Kontext die einbezogenen Parameter stammen.

Will man die gesundheitsökonomische Evidenz der untersuchten gesundheitsökonomischen Studien zum PSA-Screening bezüglich der Entscheidungsrelevanz für das deutsche Gesundheitssystem zusammenfassen und berücksichtigt dabei die limitierenden Aspekte der internen Validität, Generalisierbarkeit und Übertragbarkeit auf Deutschland, so ist folgendes festzustellen:

- Für Deutschland liegt bislang keine gesundheitsökonomische Studie zum PSA-Screening beim Prostatakarzinom vor. Sollen die vorliegenden ökonomischen Untersuchungen aus dem Ausland, vornehmlich aus den USA, zur Unterstützung des Entscheidungsprozesses herangezogen werden, sind einige wesentliche Limitationen des Informationswertes der Studien in die Überlegungen mit einzubeziehen, die im folgenden aufgeführt werden.
- Keine der Studien basiert auf validen und sicheren Angaben zur diagnostischen Sensitivität und Spezifität der Screening-Tests im allgemeinen, des PSA-Assays im speziellen. Die diagnostische Abklärung ist durch Nicht-Anwenden der Biopsie als Goldstandard in der gesamten Studienpopulation mit einem work-up bias verbunden.^{61,71} Die Biopsie selbst als diagnostischer Goldstandard ist nicht valide.^{4,26,129,139,149,158} Die Angaben zur Sensitivität, Spezifität und prädiktiven Werten hängen auch entscheidend von der Intensität (kontrollierte vs. systematische Biopsie, Biopsie-Wiederholungen) der Biopsie-Abklärung ab.⁹⁶

- Die Studien sind zum Teil veraltet bzw. beziehen sich auf ältere nicht mehr aktuelle Daten. Damit kommt es zu einem schwer beurteilbaren Nebeneinander von verschiedenen Phasen im Lebenszyklus einzelner Technologien. In diesem Zusammenhang ist auch zu erwähnen, dass die zur Steigerung der Malignomspezifität des PSA in der Literatur diskutierten und in der Praxis auch eingesetzten Testmodifikationen - PSA-D, altersabhängige PSA-Grenzwerte, PSA-V und neuerdings auch Q-PSA - nur unzulänglich untersucht wurden.
- Zweitbiopsien und - zum Teil invasive - Kontrolluntersuchungen bei falsch-positivem Testbefund werden im Regelfall weder bei den Gesundheitseffekten noch bei den Kosten berücksichtigt.
- Gesundheitseffekte und Kosten bei falschnegativem Testbefund wird lediglich in einer Studie⁸¹ Rechnung getragen.
- Alle Studien untersuchen das Prävalenz-Screening. Hinweise zur Entwicklung bei Screening-Wiederholungen gehen lediglich aus einer Untersuchung⁹¹ bezüglich der Prostatakarzinom-Diagnose hervor und sprechen für einen starken Anstieg der Kosten-Effektivitäts-Quotienten pro jährlich zusätzlich entdecktem Karzinom. Die mit der Einführung eines Früherkennungsprogrammes verbundenen organisatorischen und Infrastrukturkosten werden nicht oder nur ungenügend^{29,74} berücksichtigt.
- Mit einer Ausnahme²⁵ werden in den Studien Massen-Screening-Programme untersucht. Kosten und Konsequenzen des case finding, zu welchem die in Deutschland angebotenen Untersuchungen zur Früherkennung von Krebserkrankungen zu zählen sind, sind ungenügend analysiert. Der Compliance mit dem Screening, der Indikation zur Biopsie und der Indikation zur Therapie wird ungenügend Rechnung getragen: den Modellierungsstudien wird, mit einer Ausnahme,⁸ eine Compliance von 100% zugrunde gelegt.
- Kosten und Nutzen des Screenings hängen in den Studien entscheidend von den Annahmen zur Behandlungseffektivität ab.
- Die durch das PSA-Screening beim Patienten in der Phase der diagnostischen Abklärung anfallenden Kosten sind auf der Grundlage der Daten aus einer Studie⁵² nicht unbeträchtlich. Zu addieren sind bei Früherkennungsmaßnahmen im Rahmen des case finding Zeitkosten in den Wartezimmern.

- Ausmaß, Kosten und gesundheitsbezogene Effekte adjuvanter Therapiemaßnahmen werden in keiner Studie berücksichtigt.
- Nur die Studien von Krahn et al.⁸¹ und Barry et al.⁶ berücksichtigen wenigstens teilweise die medizinischen und ökonomischen Langzeitkonsequenzen des Screenings und beziehen eine regelmäßige posttherapeutische Nachsorge in die Analyse mit ein. Dabei kommen beide Studien unter verschiedenen Annahmen zu einem positiven inkrementellen Effekt des Screenings auf die Restlebenserwartung, der je nach Altersgruppe und untersuchter Screening-Strategie im Bereich von 0 bis 16 Tagen pro Screening-Teilnehmer liegt. Allerdings führt eine Analyse der qualitätsadjustierten Lebenserwartung bei Krahn et al. in allen untersuchten Altersgruppen und Strategien zu dem Ergebnis, dass ein Screening insgesamt mehr Schaden als gesundheitsbezogenen Nutzen bringt und höhere Kosten verursacht, als die herkömmliche Diagnosepraxis. Barry et al. führten keine solche Analyse durch.
- Produktivitätsausfallkosten werden in keiner Studie vollständig berücksichtigt, obwohl die Männer der betrachteten Altersklassen durchaus noch erwerbs- bzw. arbeitsfähig sind. Da auch eine potentiell kurative Therapie (radikale Prostatektomie bzw. Strahlentherapie) mit beachtlicher Morbidität und Mortalität verbunden ist, die Ausfallzeiten (bei Freizeit bzw. Arbeitszeit) bedingen, kann hier mit ziemlicher Sicherheit von Produktivitätsausfallkosten ausgegangen werden. Gleichzeitig ist aufgrund der nicht gesicherten Behandlungseffektivität unklar, wie hoch die durch Therapie potentiell einsparbaren Produktivitätsausfallkosten sind. Hier sei nochmals darauf hingewiesen, dass einige Wissenschaftler bzw. Nutzer einer Evaluation aus der kritischen Diskussion um die Produktivitätsausfallkosten den Schluss ziehen, auf diese Berechnung grundsätzlich zu verzichten.⁸⁷
- Intangible Kosten bzw. Präferenzdaten werden kaum berücksichtigt.^{49,81,91,93} Die Nutzwerte, die eingesetzt wurden, sind - soweit angegeben - nicht patientenzentriert. Psychologische Effekte der Krebsdiagnose bleiben unberücksichtigt.
- Die benigne Prostatahyperplasie (BPH) und deren gesundheitsbezogene und ökonomische Auswirkungen werden nie mit berücksichtigt. Innerhalb der diagnostischen Abklärungsmaßnahmen im Rahmen des PSA-Screenings spielt die BPH auch eine - wenn auch untergeordnete - Rolle, da bei symptomatischen Patienten therapeutische Maßnahmen zur Behandlung der BPH eingeleitet werden können.

- Die Kosten für die Behandlung anderer Erkrankungen aufgrund zusätzlicher Lebensjahre bei rechtzeitig und erfolgreich therapiertem Prostatakarzinom wurden nicht berücksichtigt. Somit ist die vollständige Beurteilung aus der gesellschaftlichen Perspektive nicht möglich.

Selbst wenn in Deutschland von einer ähnlichen Demographie und Epidemiologie des Prostatakarzinoms ausgegangen wird und unterstellt wird, dass der Effekt eines PSA-Massen-Screenings eine - wenn auch geringe - Verlängerung der Restlebensdauer mit sich bringt, so bleibt auch nach der Evaluation von Studien mit entscheidungsanalytischen Modellen und lebenslangem Zeithorizont offen, wie sich letztlich die Bilanz nach Berücksichtigung der Lebensqualität darstellt.

Bezüglich der fehlenden deutschen Daten zur Lebensqualitätsbewertung ist darauf hinzuweisen, dass nur eine entsprechend hohe Akzeptanz von screening- und therapiebedingten Komplikationen (z.B. Inkontinenz, sexuelle Funktionsstörungen) zu einer positiven Bilanz an qualitätsadjustierten Lebensjahren bei Einführung eines PSA-Screenings führen könnte.

Die problematische Datenlage bzw. deren Präsentation in den Publikationen lässt die Frage offen, welche zusätzlichen Kosten durch ein routinemäßig durchgeführtes PSA-Screening in Deutschland verursacht würden. Dies betrifft verschiedene Kostenmaßzahlen wie die Kosten pro Screening-Teilnehmer, die Kosten pro entdecktem Krebsfall und die Kosten pro gewonnenem Lebensjahr. Über die Kosten, die bei der Implementierung eines PSA-Screening-Programmes und bei der Rekrutierung der Screening-Teilnehmer anfallen würden, gibt es keine hinreichenden Anhaltspunkte aus den untersuchten Studien.

Für eine eindeutigere Beurteilung der Datenlage zur medizinischen Effektivität zugunsten oder gegen ein PSA-Massen-Screening sind die derzeit laufenden RCTs abzuwarten. Entsprechend zurückhaltend formuliert ist auch die Mehrzahl der heute existierenden Guidelines, Praxisrichtlinien und Empfehlungen und rät von einer Einführung eines Routine-Screenings vor dem Vorliegen dieser Ergebnisse ab.

Bis dahin bleiben die beiden zentralen von Whitmore (1988)¹⁶⁰ gestellten Fragen offen:

"Is cure necessary in those for whom it is possible?"

"Is cure possible in those for whom it is necessary?"

Eine neue Sachlage kann sich zukünftig ergeben, wenn aufgrund spezifischerer Tests oder prognostischer Faktoren eine gezieltere Selektion von Hochrisikopatienten möglich wird oder wenn neue therapeutische Verfahren mit einer geringeren Komplikationsrate entwickelt werden.

Zusammenfassend hebt das entwickelte Bewertungsinstrument Begrenzungen des Informationswerts der vorhandenen ökonomischen Studien zum PSA-Screening durch methodisch-inhaltliche Unzulänglichkeiten sowohl bei Ermittlung der Gesundheitseffekte als auch bei der Kostenerfassung hervor. Die klinische Evidenz, auf welcher die analysierten gesundheitsökonomischen Studien aufbauen, ist sowohl hinsichtlich der diagnostischen Validität der verfügbaren Früherkennungsverfahren im allgemeinen, der PSA-Bestimmung im speziellen, wie auch hinsichtlich der Effektivität der Behandlung mangelhaft und mit einem - expliziten oder impliziten - Bias zugunsten des Screenings behaftet. Die Operationalisierung der Kostenmessung ist insgesamt stark vereinfacht und erfolgt auf einem hohem Aggregationsniveau. Viele Kostenaspekte des Screenings werden - auch hier mit einem expliziten oder impliziten Bias zugunsten der Früherkennung - übergangen. Auf der Grundlage der Studienergebnisse ist im ersten Jahr eines Massen-Screening-Programmes mit beträchtlichen Kostenimplikationen monetärer wie auch gesundheitsbezogener Art zu rechnen, welche unmittelbare und sichere Finanzwirksamkeit haben. Dem stehen fragliche und abzudiskontierende zukünftige Einsparungen aufgrund verhinderter Mortalität bzw. symptomatischer Morbidität gegenüber. Die Kosten-Effekt-Relation des in Deutschland interessierenden routinemäßigen case finding mit dabei zu berücksichtigenden Compliance-Aspekten ist nicht untersucht.

C.7 Schlussfolgerungen und Empfehlungen

Bezüglich der Effektivität des PSA-Screenings beim Prostatakarzinom sei auf den Bericht von Pientka und dessen Schlussfolgerungen verwiesen.¹¹⁶ Bezüglich der ökonomischen Komponente ist folgendes festzustellen:

- Eine Übertragung der Ergebnisse der untersuchten gesundheitsökonomischen Studien auf Deutschland ist aufgrund mangelhafter interner Validität der Studien, umstrittener medizinischer Effektivität, mangelnder Datenlage zu den Präferenzen und der Lebensqualität, nicht spezifizierter Mengengerüste, fehlender Aussagen zu Kosten und der Kosten-Effektivität eines wiederholten Screenings und der Nichtberücksichtigung relevanter Kostenarten nicht ohne weiteres möglich.
- Bevor es zum routinemäßigen Einsatz eines Prostatakarzinom-Massen-Screenings mit Aufnahme in den Leistungskatalog der gesetzlichen Krankenversicherung kommt, sollte eine stärkere Evidenz für die Abnahme der gesellschaftlichen Krankheitslast herrschen.
- Bei einem Massen-Screening wäre mit zusätzlichen Kosten für das deutsche Gesundheitssystem bei fraglichen Gewinnen an Lebensjahren und Lebensqualität zu rechnen. Aufgrund der großen Heterogenität der Kosten-Effekt-Relationen in den ausgewerteten Studien und der skizzierten erheblichen Probleme einer Übertragbarkeit auf die Bundesrepublik, lassen sich die zu erwartenden Kosten und Kosten-Effektivitäts-Relationen (einschl. der Folgekosten, auch in bezug auf falschpositive Befunde und eine systematische Nachsorge der Erkrankten etc.) gegenwärtig für Deutschland auch nicht annähernd quantifizieren.
- Zur Beurteilung von Compliance, Effektivität und Kosten von einem Massen-Screening sind die Ergebnisse aus randomisierten Studien abzuwarten.
- Der individuellen Patientenentscheidung zur Durchführung eines PSA-Tests sollte grundsätzlich ein der Deklaration von Helsinki entsprechendes Aufklärungsgespräch vorausgehen.⁶⁷ Dieses Gespräch sollte eine patientengerechte Darlegung sowohl der Mortalitätsrisiken beispielsweise bei zukünftigen Operationen beinhalten, als auch des Risikos einer verminderten Lebensqualität aufgrund der Komplikationen von Diagnose- oder Therapiemaßnahmen. Angesprochen werden sollte insbesondere die Tatsache, dass mit zunehmendem Alter die Wahrscheinlichkeit dafür zunimmt, die Diagnose eines

Prostatakarzinoms gestellt zu bekommen, welches im Falle der Nichtentdeckung keine klinischen Konsequenzen gehabt hätte. Dies setzt kritische Betrachtungen der diagnostischen, therapeutischen, prognostischen und kostenbezogenen Implikationen des PSA-Screenings seitens der Ärzteschaft voraus.

De facto ist das Screening mittels der PSA-Bestimmung auch in Deutschland bereits verbreitete Praxis. Die ökonomischen Studien zum PSA-Screening weisen darauf hin, dass die hauptsächlichsten finanziellen Kosten und gesundheitsbezogenen Risiken durch die therapeutischen Implikationen der Frühdiagnose eines Prostatakarzinoms bedingt werden. Selbst wenn Patienten für die Durchführung des PSA-Assays finanziell selbst aufkommen müssen, bleibt der Großteil der gesundheitsbezogenen Risiken für die Patienten einerseits sowie der finanziellen Ausgaben für die Kostenträger andererseits bestehen. Hier gilt es, das Bewusstsein für die potentiellen medizinischen Risiken eines Screening-Tests zu schärfen.

Bezüglich der entwickelten Kriterienkataloge zur Bewertung gesundheitsökonomischer Studien wird empfohlen,

- den Qualitätskatalog anzuwenden,
- dabei sollte der Bewertungsprozess von einem interdisziplinären Team durchgeführt werden, welches die Bereiche Medizin, Gesundheitsökonomie, klinische Epidemiologie, Biometrie, der formalen Entscheidungsanalyse und der Lebensqualitätsmessung fachlich abdecken kann.

Zum methodischen Vorgehen im Rahmen des systematischen und evidenzgestützten Entscheidungsprozesses gibt es zwei grundsätzlich entgegengesetzte Extremvarianten: (1) die Entwicklung eigener entscheidungsanalytischer Modelle anhand deutscher Daten und Strukturen. (2) die qualitative Interpretation und Zusammenfassung internationaler HTA-Berichte.

Die erste Methode ist sehr zeit- und ressourcenaufwendig und die bundesdeutsche Datenlage ist im allgemeinen schwach. Hingegen lässt die zweite Methode - wie am Beispiel "PSA-Screening" veranschaulicht - zu viele Fragen bei der Übertragbarkeit offen. Dennoch liegen in einzelnen Bereichen Ergebnisse bzw. aus vorbereitenden Meta-Analysen Parameter vor, die als übertragbar zu bewerten sind. Aus diesem Grunde halten wir in als relevant angesehenen Themenfeldern folgende Vorgehensweise für sachgerecht und notwendig:

- extern entwickelte Modelle hoher Qualität und Flexibilität werden durch Forschungsvorhaben an die bundesdeutschen Verhältnisse angepasst,
- Studienergebnisse zu Kosten und Kosten-Effektivitäts-Relationen werden mit Blick auf die bundesdeutschen Verhältnisse reanalysiert.
- bei Vorliegen eines Ergebnisses aus einer der derzeit laufenden großen randomisierten Studien soll ein Update des HTA-Reports erstellt werden.

Die Validität einer solchen Re-Analyse hängt neben der Modellgüte prinzipiell vom Zugriff auf valide epidemiologische Daten, Prozessdaten der gesetzlichen und privaten Krankenversicherungen sowie gut dokumentierte Versorgungsdaten der Leistungserbringer ab.

Hiermit zeichnet sich ein Forschungsprogramm ab, dass dem gesundheitspolitischen Entscheidungsträger anwendungsorientierte Ergebnisse zu Allokationsentscheidungen zur Verfügung stellen kann, die in eine bundesdeutschen Datenbasis "Evaluation medizinischer Verfahren und Technologien" aufgenommen werden können.

Im Rahmen der gesundheitsökonomischen Evaluation der Technologie "PSA-Screening beim Prostatakarzinom" erscheint das Modell von Krahn et al. als das geeignetste Modell zur Re-Analyse für den deutschen Entscheidungskontext, da es als einziges Modell relativ vollständig die langfristigen Konsequenzen des Screenings berücksichtigt und gleichzeitig neben der reinen Analyse der inkrementellen Restlebenserwartung die Analyse der inkrementellen lebensqualitätsadjustierten Lebenserwartung unterstützt. Ferner ist es das einzige Modell, welches mit einem ausführlichen und transparenten technischen Report versehen ist. Das Modell ist insgesamt als detailliert, valide und flexibel einzustufen. Als Alternative bietet sich das Modell von Barry et al. an, welches bereits validiert wurde, mit dem Nachteil, dass dieses Modell keine Analyse von lebensqualitätsbereinigten Daten vorsieht und die Dokumentation weniger transparent erscheint. Gegebenenfalls ist mit den entsprechenden Autoren Kontakt wegen einer Kooperation aufzunehmen.

C.8 Quellenangaben

1. Abramson N, Cotton S, Eckels R, Baldock J. Voluntary Screening Program for Prostate Cancer: Detection Rate and Cost. *South Med J* 1994;87(8):785-788.
2. Arbeitsgemeinschaft Bevölkerungsbezogener Krebsregister in Deutschland. Krebs in Deutschland: Häufigkeiten und Trends. Saarbrücken: In Zusammenarbeit mit dem Robert Koch-Institut. Statistisches Landesamt Saarland, 1997.
3. Australian Cancer Society. Prostate Cancer Screening: Guidelines for Health Professionals Approved by the National Cancer Advisory Committee of the Australian Cancer Society Inc and the Clinical Oncological Society of Australia Inc. *Cancer Forum* 1995;19(1):47-50.
4. Australian Health Technology Advisory Committee (AHTAC). Prostate Cancer Screening. Canberra: Australian Government Publishing Service, 1996.
5. Baltussen B, Ament A, Leidl R. Making Cost Assessments Based on RCTs More Useful to Decision-Makers. *Health Policy* 1996;37:163-183.
6. Barry MJ, Fleming C, Coley CM, Wasson JH, Fahs MC, Oesterling JE. Should Medicare Provide Reimbursement for Prostate-Specific Antigen Testing for Early Detection of Prostate Cancer? Part I: Framing the Debate. *Urology* 1995;46(1):2-13.
7. Barry MJ, Fleming C, Coley CM, Wasson JH, Fahs MC, Oesterling JE. Should Medicare Provide Reimbursement for Prostate-Specific Antigen Testing for Early Detection of Prostate Cancer? Part III: Management Strategies and Outcomes. *Urology* 1995;46(3):277-289.
8. Barry MJ, Fleming C, Coley CM, Wasson JH, Fahs MC, Oesterling JE. Should Medicare Provide Reimbursement for Prostate-Specific Antigen Testing for Early Detection of Prostate Cancer? Part IV: Estimating the Risks and Benefits of an Early Detection Program. *Urology* 1995;46(4):445-461.
9. Beck JR, Pauker SG. The Markov Process in Medical Prognosis. *Med Decis Making* 1983;3:419-458.
10. Benoit RM. The Economics of Prostate Cancer Screening. *Oncology-Huntington* 1997;11(10):1533-48.
11. Benoit RM, Naslund MJ. An Economic Rationale for Prostate Cancer Screening. *Urology* 1994;44(6):795-803.
12. Benoit RM, Naslund MJ. A Quantitative Analysis of the Cost per Life Year Saved for Prostate Cancer Screening and Treatment. *J Urol* 1996;155 (suppl.):452A.
13. Benoit RM, Naslund MJ. The Socioeconomic Implications of Prostate-Specific Antigen Screening. *Urol Clin North Am* 1997;24(2):451-8.

14. Bitzer EM, Busse R, Dörning H, Duda L, Köbberling J, Kohlmann T, Lühmann D, Pasche S, Perleth M, Raspe H, Reese E, K R, Röseler S, Schwartz FW. Konzeption zur Anpassung und Erstellung von Technologiebewertungen in Deutschland. In: Schwartz FW, Köbberling J, Raspe H, von der Schulenburg JM, eds. Bestandsaufnahme, Bewertung und Vorbereitung der Implementation einer Datensammlung "Evaluation medizinischer Verfahren und Technologien" in der Bundesrepublik. Health Technology Assessment. Vol. 1. Baden-Baden: Nomos, 1998.
15. Black WC. Advances in Radiology and the Real Versus Apparent Effects of Early Diagnosis. *Eur J Radiol* 1998;27:116-122.
16. Black WC, Welch HG. Advances in Diagnostic Imaging and Overestimations of Disease Prevalence and the Benefits of Therapy. *New Engl J Med* 1993;328(17):1237-1243.
17. Blackmore CC, Smith WJ. Economic Analyses of Radiological Procedures: A Methodological Evaluation of the Medical Literature. *Eur J Radiol* 1998;27:123-130.
18. Breyer F, Zweifel P. Gesundheitsökonomie. 2nd ed Springer-Verlag Berlin-Heidelberg, 1997.
19. Bundesanzeiger Nr. 243 vom 31.12.1997. S. 15232, geändert durch Beschluß vom 24. April 1998, S. 10507.
20. Canadian Coordinating Office for Health Technology Assessment. Guidelines for Economic Evaluation of Pharmaceuticals. 1st ed. Canada, Ottawa: CCOHTA, 1994.
21. Canadian Coordinating Office for Health Technology Assessment. A Guidance Document for the Costing Process (Version 1.0). Canada, Ottawa: CCOHTA, 1996.
22. Canadian Coordinating Office for Health Technology Assessment. Guidelines for Economic Evaluation of Pharmaceuticals. 2nd ed. Canada, Ottawa: CCOHTA, 1997.
23. Canadian Task Force on the Periodic Health Examination. Periodic Health Examinations, 1991 Up-to-Date: Secondary Prevention of Prostate Cancer. *Can Med Assoc J* 1991;145:413-428.
24. Canadian Urological Association. Guidelines for Early Detection of Prostate Cancer. Winnipeg: CUA, 1992.
25. Chadwick DJ, Kemple T, Astley JP, MacIver AG, Gillat DA, Abrams P, Gingell JC. Pilot Study of Screening for Prostate Cancer in General Practice. *Lancet* 1991;338:613-616.
26. Chamberlain J, Melia J, Moss S, Brown J. Report Prepared for the Health Technology Assessment Panel of the NHS Executive on the Diagnosis, Management, Treatment and Costs of Prostate Cancer in England and Wales. *Brit J Urol* 1997;79(Suppl. 3):1-32.
27. Choi BCK. Sensitivity and Specificity of a Single Diagnostic Test in the Presence of Work-Up Bias. *J Clin Epidemiol* 1992;45(6):581-586.
28. Ciatto S, Bonardi R, Mazzotta A, Lombardi C, Santoni R, Cardini S, Zappa M. Confronto tra Due Modalità di Screening per il Carcinoma Prostatico: Esplorazione Rettale ed Ecografia Transrettale vs. Antigene Prostatico Specifico. *Radiol Med Torino* 1994;88(4):453-457.

29. Ciatto S, Bonardi R, Mazzotta A, Lombardi C, Santoni R, Cardini S, Zappa M. Comparing Two Modalities of Screening for Prostate Cancer: Digital Rectal Examination and Transrectal Ultrasonography vs. Prostate-Specific Antigen. *Tumori* 1995;81:225-229.
30. Coley CM, Barry MJ, Fleming C, Fahs MC, Mulley AG. Early Detection of Prostate Cancer. Part II: Estimating the Risks, Benefits, and Costs. *Ann Intern Med* 1997;126(6):468-479.
31. Coley CM, Barry MJ, Fleming C, Wasson JH, Fahs MC, Oesterling JE. Should Medicare Provide Reimbursement for Prostate-specific Antigen Testing for Early Detection of Prostate Cancer? Part II: Early Detection Strategies. *Urology* 1995;46(2):125-141.
32. Conseil d'évaluation des technologies de la santé du Québec. Screening for Cancer of the Prostate: An Evaluation of Benefits, Unwanted Health Effects and Costs. Montreal: CETS, 1995.
33. Dawson-Saunders B, Trapp RG. *Basic and Clinical Biostatistics*. 2nd ed. London: Prentice-Hall International, 1994.
34. De Biasi F, Londero D, Vanuzzo D, Pilotto L, Praturlon S, Guerra UP. Is Serum PSA Measurement a Good Test for Early Diagnosis of Prostate Cancer? Results of a Pilot Study and Cost Analysis. *Int J Biol Markers* 1996;11(1):18-23.
35. Detsky AS, Naglie G, Krahn MD, Naimark D, Redelmeier DA. Primer on Medical Decision Analysis: Part 1 - Getting Started. *Med Decis Making* 1997;17:123-125.
36. Detsky AS, Naglie G, Krahn MD, Redelmeier DA, Naimark D. Primer on Medical Decision Analysis: Part 2 - Building a Tree. *Med Decis Making* 1997;17:126-135.
37. Dorr VJ, Williamson SK, Stephens RL. An Evaluation of Prostate-Specific Antigen as a Screening Test for Prostate Cancer. *Arch Intern Med* 1993;153:2529-2537.
38. Drummond M, Richardson WS, O'Brien BJ, Levine M, Heyland DK. Users' Guides to the Medical Literature: XIII. How to Use an Article on Economic Analysis of Clinical Practice. A: Are the Results of the Study Valid? *JAMA* 1997;277(19):1552-1557.
39. Drummond MF, Bloom BS, Carrin G, Hillman AL, Hutchings HC, Knill-Jones RP, et al. Issues in the Cross-National Assessment of Health Technology. *Int J Technol Assess Health Care* 1992;8:671-682.
40. Drummond MF, Brandt A, Luce BR, Rovira J. Standardizing Methodologies for Economic Evaluation in Health Care. *Int J Technol Assess Health Care* 1993;9(1):26-36.
41. Drummond MF, Jefferson TO. Guidelines for Authors and Peer Reviewers of Economic Submissions to the BMJ. *BMJ* 1996;313:275-283.
42. Drummond MF, Torrance G, Mason J. Cost-Effectiveness League Tables: More Harm than Good? *Soc Sci Med* 1993;37(1):33-40.
43. Eisenberg JM. *Clinical Economics. A Guide to the Economic Analysis of Clinical Practices*. *JAMA* 1989;262(20):2879-2886.

44. Feinstein AR, Sosin DM, Carolyn W. The Will Rogers Phenomenon. Stage Migration and New Diagnostic Techniques as a Source of Misleading Statistics for Survival in Cancer. *New Engl J Med* 1985;312(25):1604-1608.
45. Fenny D, Furlong W, Boyle M, Torrance GW. Multi-Attribute Health Status Classification Systems: Health Utilities Index. *Pharmaco Economics* 1995;7:490-502.
46. Glasser JH. The Aims and Methods of Technology Assessment. *Health Policy* 1988;9:241-250.
47. Gohagan JK, Prorok PC, Kramer BS, Cornett JE. Prostate Cancer Screening in the Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial of the National Cancer Institute. *J Urol.* 1994;152(5 Pt 2):1905-9.
48. Gold MR, Siegel JE, Russell LB, Weinstein MC, eds. *Cost-Effectiveness in Health and Medicine*. New York - Oxford: Oxford University Press, 1996.
49. Gottlieb RH, Mooney C, Mushlin AI, Rubens DJ, Fultz PJ. The Prostate. Decreasing Cost-Effectiveness of Biopsy with Advancing Age. *Invest Radiol* 1996;31(2):84-90.
50. Grasso Leanza F, Panella P, Pennisi M, Pepe P. Screening del Carcinoma Prostatico in Pazienti Disurici: Protocolli Diagnostici e Rapporto Costi/Benefici [Screening of Prostate Carcinoma in Dysuric Patients: Diagnostic Records and Cost/Benefit Relation]. *Arch Ital Urol Androl* 1997;69(3):143-149.
51. Green CJ, Hadorn D, Bassett K, Kazanjian A. *Prostate Specific Antigen in the Early Detection of Prostate Cancer*. Vancouver: BCOHTA, 1993.
52. Gustafsson O, Carlsson P, Norming U, Nyman CR, Svensson H. Cost-Effectiveness Analysis in Early Detection of Prostate Cancer: An Evaluation of Six Screening Strategies in a Randomly Selected Population of 2,400 Men. *Prostate* 1995;26:299-309.
53. Guyatt G, Drummond MF, Feeny D, Tugwell P, Stoddart G, Haynes RB, Bennett K, Labelle R. Guidelines for the Clinical and Economic Evaluation of Health Care Technologies. *Soc Sci Med* 1986;22(4):393-408.
54. Guyatt GH, Naylor CD, Juniper E, Heyland DK, Jaeschke R, Cook DJ. Users' Guides to the Medical Literature: XII. How to Use Articles About Health-Related Quality of Life. *JAMA* 1997;277(15):1232-1237.
55. Guyatt GH, Sackett DL, Cook DJ. Users' Guides to the Medical Literature: II. How to Use an Article About Therapy or Prevention. A: Are the Results of the Study Valid? *JAMA* 1993;270(21):2598-2601.
56. Guyatt GH, Sackett DL, Cook DJ. Users' Guide to the Medical Literature: II. How to Use an Article About Therapy or Prevention. B: What Were the Results and Will They Help Me in Caring for My Patients? *JAMA* 1994;271(21):59-63.
57. Guyatt GH, Sackett DL, Sinclair JC, Hayward R, Cook DJ, Cook RJ. Users' Guides to the Medical Literature: IX. A Method for Grading Health Care Recommendations. *JAMA* 1995;274(22):1800-1804.

58. Guyatt GH, Veldhuyzen Van Zanten SJO, Feeny DH, Patrick DL. Measuring Quality of Life in Clinical Trials: A Taxonomy and Review. *Can Med Assoc J* 1989;140:1441-1448.
59. Hammerer P, Huland H. Stellenwert der PSA-Bestimmung für die Früherkennung des Prostatakarzinoms. *Urologe (A)* 1995;34:283.
60. Harvey CM. The Reasonableness of Non-Constant Discounting. *J Publ Econom* 1994;53:31-51.
61. Haynes RB. How to Read Clinical Journals. II. To Learn About A Diagnostic Test. *Can Med Assoc J* 1981;124:703-710.
62. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt GH. Users' Guides to the Medical Literature: VIII. How to Use Clinical Practice Guidelines. A: Are the Recommendations Valid? *JAMA* 1995a;274(7):570-574.
63. Hayward RS, Wilson MC, Tunis SR, Bass EB, Guyatt GH. Users' Guides to the Medical Literature: VIII. How to Use Clinical Practice Guidelines. B: What are the Recommendations and Will They Help You in Caring for Your Patients? *JAMA* 1995b;274(20):1630-1632.
64. Hennekens CH, Buring JE. *Epidemiology in Medicine*. Boston - Toronto: Little Brown, 1987.
65. Hessel, Kohlmann, Krauth, Novy, Seitz, Siebert, Wasem. Gesundheitsökonomische Evaluation in der Rehabilitation. Teil I: Prinzipien und Empfehlungen für die Leistungserfassung. Forschungsstelle für Gesundheitsökonomie und Gesundheitssystemforschung. Diskussionspapier Nr. 19, <http://www.reha-verbund.de/oekonom.htm>. Hannover: 1998.
66. Hinman FJ. Screening for Prostatic Carcinoma. *J Urol* 1991;145:126-130.
67. Hölzel D. Prostatakarzinom. Ist die Früherkennung in einer Sackgasse? *Dtsch Arztebl* 1995b;92(25/26):B1353-B1361.
68. Hulka BS. Cancer Screening. Degrees of Proof and Practical Application. *Cancer* 1988;62(8):1776-1780.
69. Irwig L, Tosteson ANA, Gatsonis C, Lau J, Colditz G, Chalmers TC, Mosteller F. Guidelines for Meta-Analyses Evaluating Diagnostic Tests. *Ann Intern Med* 1994;120(8):667-676.
70. Jacobsen SJ, Katusic SK, Bergstrahl EJ, Oesterling JE, Ohrt D, et al. Incidence of Prostate Cancer Diagnosis in the Eras Before and After Serum Prostate-Specific Antigen Testing. *JAMA* 1995;274(18):1445-1449.
71. Jaeschke R, Guyatt GH, Sackett DL. Users' Guides to the Medical Literature: III. How to Use an Article About a Diagnostic Test. A: Are the Results of the Study Valid? *JAMA* 1994a;271(5):389-391.
72. Jaeschke R, Guyatt GH, Sackett DL. Users' Guides to the Medical Literature: III. How to Use an Article About a Diagnostic Test. B: What Are the Results and Will They Help Me in Caring for My Patients? *JAMA* 1994b;271(9):703-707.

73. John J, Hofmann U, Nagl H, Schneider M. Ökonomische Evaluation von Gesundheitsleistungen in Deutschland. DLR-Projektträger des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie. Bremerhaven: Wirtschaftsverlag NW, Verlag für neue Wissenschaft GmbH, 1996.
74. Kantrowitz W, Doyle J, Semeraro J, Krane RJ. Prostate Cancer Screening in a Large Corporation Population. *J Occup Environ Med* 1995;37(10):1193-1198.
75. Kassenärztliche Bundesvereinigung / Bundesverband der Betriebskrankenkassen. Frequenzstatistik BMAE-EGO. 1988-1996.
76. König H-H, Stratmann D, Leidl R. Wann ist eine Leistung kosteneffektiv? Leitfaden für die Beurteilung von ökonomischen Evaluationen. *Münch. med. Wschr.* 1998;140(15):216-220.
77. Konsensgruppe "Gesundheitsökonomie". Deutsche Empfehlungen zur gesundheitsökonomischen Evaluation - Hannoveraner Konsens. *Z Allg Med* 1996;72:485-490.
78. Koopmanschap MA, Rutten FFH, van Ineveld BM, von Roijen L. The Friction Cost Method for Measuring Indirect Costs of Disease. *J Health Economics* 1995;14:171-189.
79. Kovacs A, Gulacsi L, Sarosi G. Prostate Cancer Screening: Supplier Induced Demand. Annual Meeting of International Society of Technology Assessment in Health Care; 1989.
80. Krahn M, Mahoney J, Eckman M, Pauker S, Detsky AS. Supplementary Data and Results: Screening for Prostate Cancer: A decision Analytic View. National Archive Publication Service Document, 1994.
81. Krahn MD, Mahoney JE, Eckman MH, Trachtenberg J, Pauker SG, Detsky AS. Screening for Prostate Cancer. A Decision Analytic View. *JAMA* 1994;272(10):773-780.
82. Krahn MD, Naglie G, Naimark D, Redelmeier DA, Detsky AS. Primer on Medical Decision Analysis: Part 4 - Analyzing the Model and Interpreting the Results. *Med Decis Making* 1997;17:142-151.
83. Kramer BS, Brown ML, Prorok PC, Potosky AL, Gohagan JK. Prostate Cancer Screening: What We Need to Know. *Ann Intern Med* 1993;119(9):914-923.
84. Launois R. Cost-Effectiveness Analysis of Strategies for Screening Prostate Cancer. In: Zweifel P, Frech IHE, eds. *Health Economics Worldwide (bzw. Developments in Health Economics and Public Policy, vol. I)*. Dordrecht-Boston-London: Kluwer Academic Publishers, 1992.
85. Laupacis A, Wells G, Richardson WS, Tugwell P. Users' Guides to the Medical Literature: V. How to Use an Article About Prognosis. *JAMA* 1994;272(3):234-237.
86. Leidl R. Standardisierung und Flexibilisierung ökonomischer Evaluationsstudien. In: Trampisch H, Lange S, eds. *Medizinische Forschung, Ärztliches Handeln*. München: MMV Medizin Verlag, 1995.

87. Leidl R. Der Effizienz auf der Spur: Eine Einführung in die ökonomische Evaluation. In: Schwartz FW, Badura B, Leidl R, Raspe H, Siegrist J, eds. Das Public Health Buch. Gesundheit und Gesundheitswesen. München-Wien-Baltimore: Urban und Schwarzenberg, 1997;347-369.
88. Levine M, Walter S, Lee H, Haines T, Holbrook A, Moyer V. Users' Guides to the Medical Literature: IV. How to Use an Article About Harm. *JAMA* 1994;271(20):1615-1619.
89. Littrup PJ, Goodman A. Economic Considerations of Prostate Cancer. The Role of Detection Specificity and Biopsy Reduction. *Cancer* 1995;75(7):1987-1993.
90. Littrup PJ, Goodman AC. Costs and Benefits of Prostate Cancer Screening. *In Vivo* 1994a;8:423-428.
91. Littrup PJ, Goodman AC, Mettlin C. The Benefit and Cost of Prostate Cancer Early Detection. *CA Cancer J Clin* 1993;43(3):134-149.
92. Littrup PJ, Goodman AC, Mettlin CJ, Murphy GP. Cost Analyses of Prostate Cancer Screening: Frameworks for Discussion. *J Urol* 1994b;152:1873-1877.
93. Littrup PJ, Kane RA, Mettlin CJ, Murphy GP, Lee F, Toi A, Badalament R, Babaian R. Cost-Effective Prostate Cancer Detection. Reduction of Low-Yield Biopsies. *Cancer* 1994c;74(12):3146-3158.
94. Luce BR, Elixhauser A. Standards for Socioeconomic Evaluation of Health Care Products and Services. Berlin-Heidelberg: Springer-Verlag, 1990.
95. Mason J, Drummond M. Reporting Guidelines for Economic Studies. *Health Economics* 1995;4:85-94.
96. McNaughton CM, Ransohoff DF, Barry MJ. Early Detection of Prostate Cancer: Serendipity Strikes Again. *JAMA* 1997;287:1516-1519.
97. McNeil BJ, Pauker SG. Decision Analysis for Public Health: Principles and Illustrations. *Ann Rev Public Health* 1984;5:135-161.
98. Menemeyer ST, Cyr LP. A Bootstrap Approach to Medical Decision Analysis. *J Health Economics* 1997;16:741-747.
99. Mettlin CJ, Jones GW, Averette H, et al. Defining and Updating the American Cancer Society Guidelines for the Cancer-Related Check Up: Prostate and Endometrial Cancers. *CA Cancer J Clin* 1993;43:42-46.
100. Naglie G, Krahn MD, Naimark D, Redelmeier DA, Detsky AS. Primer on Medical Decision Analysis: Part 3 - Estimating Probabilities and Utilities. *Med Decis Making* 1997;17:136-141.
101. Naimark D, Krahn MD, Naglie G, Redelmeier DA, Detsky AS. Primer on Medical Decision Analysis: Part 5 - Working with Markov Processes. *Med Decis Making* 1997;17:152-159.
102. Nakagawa S, Ebisui K, Sugimoto K, Nakanishi H, Kanemitsu N, Watanabe H. [Cost-Effectiveness Analysis of Prostatic Cancer Screening]. *Jpn J Urol* 1997;88(10):892-899.

103. National Cancer Institute. Prostate Cancer Screening State-of-the-Art Statement. Physician Data Query (Database). Bethesda: NCI, 1993.
104. Naylor CD, Guyatt GH. Users' Guides to the Medical Literature: X. How to Use an Article Reporting Variations in the Outcomes of Health Services. *JAMA* 1996a;275(7):554-558.
105. Naylor CD, Guyatt GH. Users' Guides to the Medical Literature: XI. How to Use an Article About a Clinical Utilization Review. *JAMA* 1996b;275(18):1435-1439.
106. Neuhauser D, Lewicki AM. What Do We Gain from the Sixth Stool Guaiac? *New Engl J Med* 1975;293:226-228.
107. O'Brien BJ. A Tale of Two (or More) Cities: Geographic Transferability of Pharmacoeconomic Data. *Am Journal Managed Care* 1997;3(Suppl.):33-39.
108. O'Brien B, Drummond MF, Labelle R, Willan A. In Search of Power and Significance: Issues in the Design and Analysis of Stochastic Economic Appraisals. *Medical Care* 1994;32:150-163.
109. O'Brien BJ, Heyland DK, Richardson WS, Levine M, Drummond MF. Users' Guides to the Medical Literature: XIII. How to Use an Article on Economic Analysis of Clinical Practice. B: What are the Results and Will They Help Me in Caring for My Patients? *JAMA* 1997;277(22):1802-1806.
110. Optenberg SA, Thompson IM. Economics of Screening for Carcinoma of the Prostate. *Urol Clin North Am* 1990;17(4):719-737.
111. Oxmann AD, Cook DJ, Guyatt GH. Users' Guides to the Medical Literature: VI. How to Use an Overview. *JAMA* 1994;272(17):1367-1371.
112. Oxmann AD, Sackett DL, Guyatt GH. Users' Guides to the Medical Literature: I. How to Get Started. *JAMA* 1993;270(17):2093-2095.
113. Parsonage M, Neuburger H. Discounting and Health Benefits. *Health Economics* 1992;1:71-76.
114. Patrick DL, Deyo RA. Generic and Disease-specific Measures in Assessing Health Status and Quality of Life. *Med Care* 1989;27(3 Suppl):S217-S232.
115. Pauker SG, Kassirer JP. Decision Analysis. *New Engl J Med* 1987;316(5):250-258.
116. Pientka L. PSA-Screening beim Prostatakarzinom. *Health Technology Assessment*. Vol. 5. Baden-Baden: Nomos, 1999.
117. Potosky AL, Miller BA, Albertsen PS, Kramer BS. The Role of Increasing Detection in the Rising Incidence of Prostate Cancer. *JAMA* 1995;273(7):548-552.
118. Ransohoff DF, Feinstein AR. Problems of Spectrum and Bias in Evaluating the Efficacy of Diagnostic Tests. *New Engl J Med* 1978;299(17):926-930.
119. Richardson WS, Detsky AS. Users' Guides to the Medical Literature: VII. How to Use a Clinical Decision Analysis. A: Are the Results of the Study Valid? *JAMA* 1995a;273(16):1292-1295.

120. Richardson WS, Detsky AS. Users' Guides to the Medical Literature: VII. How to Use a Clinical Decision Analysis. B: What Are the Results and Will They Help Me in Caring for My Patients? *JAMA* 1995b;273(20):1620-1613.
121. Rosen MA. Analysis of the Annual Cost of a Fully Implemented Prostate Cancer Screening Program. *Urology* 1995;153 (suppl):464A.
122. Rovira J. Standardizing Economic Appraisal of Health Technology in the European Community. *Soc Sci Med* 1994;38(12):1675-1678.
123. Sackett DL. How to Read Clinical Journals: I. Why To Read Them and How to Start Reading Them Critically. *Can Med Assoc J* 1981a;124:555-558.
124. Sackett DL. How to Read Clinical Journals: V. To Distinguish Useful from Useless or Even Harmful Therapy. *Can Med Assoc J* 1981b;124:1156-1162.
125. Schmailzl KJG. *Harrisons Innere Medizin*. 13th ed. Vol. 2nd. New York: McGraw-Hill, 1995.
126. Schöffski O, Glaser P, Schulenburg Graf v. d. J-M. *Gesundheitsökonomische Evaluationen. Grundlagen und Standortbestimmung*. Berlin-Heidelberg: Springer-Verlag, 1998.
127. Schön D, Bertz J, Hoffmeister H. *Bevölkerungsbezogene Krebsregister in der Bundesrepublik Deutschland*. Band 3, RKI-Schriften 2/1995. München: Medizinischer Verlag, 1995.
128. Schwarz FW, Busse R, Raspe H, al e. Bestandsaufnahme, Bewertung und Vorbereitung der Implementation einer Datensammlung 'Evaluation medizinischer Verfahren und Technologien' in der Bundesrepublik. Vorläufiger Endbericht. . Hannover/Witten, Lübeck, Wuppertal., 1997.
129. Selley S, Donovan J, Faulkner A, Coast J, Gillat D. *Diagnosis, Management and Screening of Early Localised Prostate Cancer*. *Health Technol Assess* 1997;1(2).
130. Semjonow A, Brandt B, Oberpenning F, Hertle L. Unterschiedliche Bestimmungsverfahren erschweren die Interpretation des prostataspezifischen Antigens. *Urologe (A)* 1995;34:303-315.
131. Simpson KN, Brown RE. Cost Effectiveness of Adding Prostate Specific Antigen (PSA) Test to Digital Rectal Examination (DRE) for Early Detection of Prostate Cancer. *J Urol* 1993;149 (suppl):413A.
132. Sloan FA. *Valuing Health Care, Costs, Benefits, and Effectiveness of Pharmaceuticals and Other Medical Technologies*. Cambridge-New York- Melbourne: Cambridge University Press, 1995.
133. Sonnenberg FA, Beck JR. *Markov Model in Medical Decision Making: A Practical Guide*. *Med Decis Making* 1993;13:322-338.
134. *Sozialgesetzbuch. (SGB) V: Gesetzliche Krankenversicherung*. München: C. H. Beck, 1993.
135. Standaert B, Denis L. The European Randomized Study of Screening for Prostate Cancer: an update. *Cancer* 1997;80(9):1830-4.

136. Statistisches Bundesamt. Statistisches Jahrbuch 1996 für die Bundesrepublik Deutschland. Wiesbaden: Metzler Poeschel, 1996.
137. Stoddart GL, Drummond MF. How to Read Clinical Journals. VII. To Understand an Economic Evaluation (Part A). *Can Med Assoc J* 1984a;130:1428-1434.
138. Stoddart GL, Drummond MF. How to Read Clinical Journals: VII. To Understand an Economic Evaluation (Part B). *Can Med Assoc J* 1984b;130(1542-1549).
139. Swedish Council on Technology Assessment in Health Care (SBU). Mass Screening for Prostate Cancer. *Int J Cancer* 1996(Suppl. 9: i-iii):1-72.
140. Szucs TD. Medizinische Ökonomie. Eine Einführung. München: Urban & Vogel, 1997.
141. Task Force on Principles for Economic Analysis of Health Care Technology. Economic Analysis of Health Care Technology. A Report on Principles. *Ann Intern Med* 1995;122(1):61-70.
142. Torrance GW. Measurement of Health state Utilities for Economic Appraisal. A Review. *J Health Economics* 1986;5:1-30.
143. Torrance GW. Measurement of Health state Utilities for Economic Appraisal. *J Chron Dis* 1987;40(6):593-600.
144. Torrance GW, Feeny D. Utilities and Quality-Adjusted Life Years. *Int J Technol Assess Health Care* 1989;5:559-575.
145. Trout KS. How to Read Clinical Journals: IV. To Determine Etiology or Causation. *Can Med Assoc J* 1981;124:985-990.
146. Tugwell P. How to Read Clinical Journals: III. To Learn the Clinical Course and Prognosis of Disease. *Can Med Assoc J* 1981;124:869-872.
147. Tugwell P. How to Read Clinical Journals: VI. To Learn About the Quality of Clinical Care. *Can Med Assoc J* 1984;130:377-381.
148. Tumorzentrum München. Empfehlungen zur Diagnostik, Therapie und Nachsorge: Urogenitale Tumoren. 2nd ed, 1997.
149. U.S. Congress, Office of Technology assessment (OTA). Cost and Effectiveness of Prostate Cancer Screening in Elderly Men. Washington, D C: U. S. Government Printing Office, 1995.
150. U.S. Preventive Services Task Force. Screening for Prostate Cancer. Guide to Clinical Preventive Services: An Assessment of the Effectiveness of 169 Interventions. Baltimore: U. S. Preventive Services Task Force, 1989.
151. Udvarhelyi IS, Colditz GA, Rai A, Epstein AM. Cost-Effectiveness and Cost-Benefit Analyses in the Medical Literature. Are the Methods Being Used Correctly? *Ann Intern Med* 1992;116(3):238-244.
152. Voges G, Wulff HD, Planz C, Hohenfellner R. Tumoren des männlichen Genitales. In: Hornbostel H, Kaufmann W, Siegenthaler W, eds. *innere Medizin in Praxis und Klinik*. Vol. 2. Stuttgart-New York: Thieme Verlag, 1992.

153. von Eschenbach A, Ho R, Murphy GP, Cunningham M, Lins N. American Cancer Society Guidelines for the Early Detection of Prostate Cancer. *Cancer* 1997;80(9):1805-1807.
154. Wagener C, Hossfeld DK. Diagnostische Validität von Tumormarkern. *Urologe (B)* 1997;37:203-208.
155. Wasson JH, Cushman CC, Bruskewitz RC, Littenberg B, Mulley AG, Wennberg JE, et al. A Structured Literature Review of Treatment for Localized Prostate Cancer. *Arch Fam Med* 1993;2:487-493.
156. Weingärtner K, Riedmiller H. Prostatakarzinom. Radikale Prostatektomie, neoadjuvante und adjuvante Therapiemodalitäten. *Urologe (B)* 1998;38:186-193.
157. Weining C, Semjonow A, Hertle L. Nutzen verschiedener Verfahren zur Verbesserung der Spezifität von PSA in der Prostatakarzinomdiagnostik. *Urologe (B)* 1997;37:216-220.
158. Weissbach L, Miller K. Diagnostische und therapeutische Standards in der Urologischen Onkologie. München-Bern-Wien-New York: W. Zuckschwerdt Verlag, 1998.
159. Welte R, Leidl R. Übertragung ausländischer Studienergebnisse aus dem Bereich der ökonomischen Evaluation auf Deutschland: Probleme und Lösungsvorschläge. Projekt "Aufbau einer Datenbasis 'Evaluation medizinischer Verfahren und Technologien' in der Bundesrepublik". Bonn: BMG, 1998.
160. Whitmore WF. Overview. Historical and Contemporary. *Natl Cancer Inst Monogr* 1988;7:7-11.
161. Wilson JMG, Jungner G. Principles and Practice of Screening for Disease. Geneva: WHO, 1968.

D Anhang

D.1 Anhang 1: Dokumentationen

Doku 1: HTA-Reports, Reviews, Richtlinien	
Review: PSA-Screening beim Prostatakarzinom	
berücksichtigte Einrichtungen:	
CRD	<input type="checkbox"/>
CC	<input type="checkbox"/>
UKCHO	<input type="checkbox"/>
KEZ / ZFR	<input type="checkbox"/>
SBU	<input checked="" type="checkbox"/>
TNO-VG	<input type="checkbox"/>
AHCPR	<input checked="" type="checkbox"/>
CCOHTA	<input checked="" type="checkbox"/>
ECRI	<input type="checkbox"/>
Andere:	<input checked="" type="checkbox"/>
AHTAC	
BCOHTA	
Danish Institute for Health Technology Assessment	
ISTAHC	
CHEPA	
Ministère de la Santé et des Services sociaux du Québec	
NCCHTA	
OTA	
Angeschriebene Einrichtungen zu nationalen Leitlinien, Empfehlungen oder Stellungnahmen zur (PSA-gestützten) Früherkennung von Prostatakarzinom und zu therapeutischen Behandlungsoptionen:	
Ärztliche Zentralstelle für Qualitätssicherung (ÄZQ)	
Deutsche Krebsgesellschaft	
Deutsche Gesellschaft für Urologie	
Südostdeutsche Gesellschaft für Urologie	
Tumorzentrum München	
Berufspolitische Ärzte- und Urologenverbände	

Doku 2: Wissenschaftliche Literatur und Primärstudien	
Review: PSA-Screening beim Prostatakarzinom	
Zeitraumen:	Monate
berücksichtigte Jahrgänge von: 1986	bis: August 1998
Datenbanken:	
MEDLINE	<input checked="" type="checkbox"/>
EMBASE	<input checked="" type="checkbox"/>
DARE	<input checked="" type="checkbox"/>
CC	<input checked="" type="checkbox"/>
TA-Datenbanken	<input checked="" type="checkbox"/>
Andere:	<input checked="" type="checkbox"/>
Econ Lit	<input checked="" type="checkbox"/>
PSYINDEXplus	<input checked="" type="checkbox"/>
sociofile	<input checked="" type="checkbox"/>
HealthSTAR	<input checked="" type="checkbox"/>
CancerLit	<input checked="" type="checkbox"/>
SOMED 1.1	<input checked="" type="checkbox"/>
NEED	<input checked="" type="checkbox"/>
weitere Datenquellen:	
(falls ja, bitte gesondert spezifizieren)	
Referenzlisten	<input checked="" type="checkbox"/> *
Kongressbände	<input type="checkbox"/>
Expertenbefragung	<input checked="" type="checkbox"/>
Fachgesellschaften	<input type="checkbox"/>
Forschungseinrichtungen	<input type="checkbox"/>
Andere:	<input checked="" type="checkbox"/>
Research Activities der Agency for Health Care Policy and Research (AHCPR), 1995-08/1998	
Health Technology Assessment News des Emergency Care Research Institute (ECRI), 1995-08/1998	

* Eine Durchsicht der Referenzlisten aller zusammengetragener Publikationen - Primärstudien, HTA-Berichte und Reviews - wurde vorgenommen.

Doku 2a: Datenbank-Suchstrategie	
Review: PSA-Screening beim Prostatakarzinom	
Datenbanken: EMBASE (WebSPIRS) MEDLINE EXPRESS (WebSPIRS; expliziter Ausschluss von EMBASE-Referenzen) CC Search (WebSPIRS) EconLit (WebSPIRS) PSYNDEXplus - Lit.&AV (WebSPIRS) sociofile (WebSPIRS) HealthSTAR (OVID; expliziter Ausschluss von MEDLINE-Referenzen) CancerLit (OVID; expliziter Ausschluss von MEDLINE-Referenzen)	
Software: WebSPIRS und OVID	
#1	prostate-specific antigen or prostate specific antigen or PSA
#2	screening
#3	cost or costs
#4	#1 and #2 and #3

Ferner wurden für den gleichen Zeitraum verwendet die Datenbank SOMED 1.1 (CD-ROM: Landesinstitut für den öffentlichen Gesundheitsdienst) und die als Teil der Cochrane Library geführten Datenbanken DARE (University of York, NHS Centre for Reviews and Dissemination: <http://nhscrd.york.ac.uk/welcome.html>) und NEED (University of York, NHS Centre for Reviews and Dissemination: <http://nhscrd.york.ac.uk>). Bei der Suche in DARE und NEED wurden die Schlagwörter "prostate cancer" und "prostate specific antigen" bzw. "prostate-specific antigen" bzw. "PSA" verwendet, in SOMED das Schlagwort Prostata

Doku 3: Primärdaten	
Review: PSA-Screening beim Prostatakarzinom	
Datenquellen:	
Abrechnungsdaten GKV	<input type="checkbox"/>
Abrechnungsdaten ambulante Versorgung	<input type="checkbox"/>
Krankenhausstatistik (Statistisches Bundesamt)	<input checked="" type="checkbox"/>
MDK-Daten	<input type="checkbox"/>
Eigene Erhebungen (bitte spezifizieren):	<input type="checkbox"/>
Angeschriebene Einrichtungen:	
Zur Epidemiologie des Prostatakarzinoms:	
Statistisches Bundesamt	
Tumorregister München	
Krebsregister Saarland	
Gemeinsames Krebsregister der Länder Berlin, Brandenburg, Mecklenburg-Vorpommern, Sachsen-Anhalt und der Freistaaten Sachsen und Thüringen	
Zur Häufigkeit von präventiven, diagnostischen und therapeutischen Interventionen beim Prostatakarzinom:	
Kassenärztliche Bundesvereinigung	
Bundesverband der Betriebskrankenkassen	

D.2 Anhang 2: Deskriptionskatalog

Hinweise zum Deskriptionskatalog:

Der Deskriptionskatalog ist eine detaillierte Zusammenstellung der in gesundheitsökonomischen Evaluationen möglichen Varianten der Studienanlage, -form, -durchführung und -auswertung. In Form eines rasterartigen Leitfadens dient er dazu, die Palette der verschiedenen Ansätze einer gesundheitsökonomischen Evaluation aufzuzeigen. Wird er auf eine Studie angewendet, ergibt sich ein Abbild der Methoden, die in der jeweils vorliegenden Studie realisiert wurden. Dieser umfangreiche Katalog dient der detaillierten, umfassenden und standardisierten Erfassung von methodischen Einzelinformationen.

Er ist insbesondere als Schulungsinstrument zur Standardisierung und Systematisierung der spezifischen Methoden und Terminologie zu verstehen. Seine Kenntnis dient als Hilfe bei der Anwendung des Transparenzkataloges (s. Tabelle 9) und des Qualitätskataloges (s. Tabelle 10).

Der Deskriptionskatalog wurde von dem gesundheitsökonomischen Panel der German Scientific Working Group Technology Assessment for Health Care im Konsensusverfahren entwickelt (Projektgruppen München, Hannover, Ulm; Einzelheiten siehe Tabelle 5).

Tabelle 24: Deskriptionskatalog. Kriterienkatalog zur deskriptiven Erfassung und Dokumentation von Merkmalen und Inhalten gesundheitsökonomischer Studien (entwickelt im Konsensusverfahren von den gesundheitsökonomischen Projektgruppen München, Hannover, Ulm)

(Fortsetzung)

D.3 Anhang 3: Erläuterungen zu den gesundheitsökonomischen Kriterienkatalogen

Es folgen Ausführungen zu den Teilbereichen des Deskriptionskataloges, die ferner für die praktische Anwendung aller drei gesundheitsökonomischen Kriterienkataloge (Deskriptionskatalog, Transparenzkatalog und Qualitätskatalog) hilfreich sind.

Publikation

Neben den Angaben zu Titel, Autor(en), Publikationsort und -jahr ist hier von Interesse, sich die Beziehungen zwischen Auftraggeber und durchführender Institution zu vergegenwärtigen, falls diese offengelegt wurden.^{20,126} Angaben bezüglich des Auftraggebers geben oftmals auch schon einen Hinweis darauf, wer als Adressat der Studie angesehen werden kann und aus welcher Perspektive diese ausgeführt wurde. Manchen Autoren ist es unmöglich, alle Daten, Begründungen und Annahmen einer gesundheitsökonomischen Evaluation auf der in Zeitschriften begrenzten Seitenzahl adäquat und nachvollziehbar zu komprimieren. Ein zusätzlicher, auf Anfrage verfügbarer technischer Bericht erleichtert es dem Leser der Studie, die Analyse eingehend zu überprüfen und zu beurteilen, sein Verständnis spezifischer Daten und Methoden zu verbessern, die Arbeit zu reproduzieren oder die Resultate an neue Entscheidungskontexte - etwa ein anderes Land - anzupassen.⁴⁸

Fragestellung und Evaluationsrahmen

Zu Konzeption und Aufbau gesundheitsökonomischer Studien gehören

- a) die Problemdefinition und eine klare, bündige Darstellung des Studienzwecks einschließlich der interessierenden Zielgrößen.^{15,33,48,132,138}
- b) die Definition der Perspektive.^{20,43,48,53,76,77,87,126,132,138,141,151} Als Perspektive wird der Standpunkt bezeichnet, aus dessen Sicht Outcomes und Ressourcenverbrauch erfasst und bewertet werden. Die Wahl der Perspektive ist von methodischer Bedeutung: Sie hat insofern einen entscheidenden Einfluss auf die Ergebnisse einer ökonomischen Evaluation, als sie festlegt, welche Outcomes

und welcher Ressourcenverbrauch berücksichtigt und wie diese bewertet werden sollen. Zur adäquaten Interpretation der Resultate ist die Offenlegung der eingenommenen Perspektive deshalb unabdingbar. Werden verschiedene Perspektiven gewählt, müssen auch getrennte Berechnungen aufgestellt werden. Die umfassendste der Perspektiven ist die gesellschaftliche, bei der alle relevanten Outcomes und jeglicher relevante Ressourcenverbrauch einbezogen werden müssen. Oft ist der Adressat der Studie für die Wahl der Perspektive entscheidend: Voraussetzung für die Akzeptanz der Ergebnisse ist, dass die gewählten Endpunkte auch für ihn von Bedeutung sind.

- c) die Bestimmung des analytischen Zeithorizonts. Der Begriff "Zeithorizont" lässt sich in zweierlei Hinsicht auslegen: Zum einen kann er den zeitlichen Rahmen einer gesundheitsökonomischen Evaluation bezeichnen und umfasst die durch den Einsatz einer Technologie bzw. Vergleichsalternative bedingten Gesundheitszustandspfade bis hin zu den im Rahmen der Evaluation interessierenden klinischen und ökonomischen Zielgrößen.^{20,48,77,126,132} Die Münchner Projektgruppe hat sich darauf verständigt, diesen analysebezogenen Aspekt des Zeithorizonts als Bestandteil des Evaluationsansatzes zu betrachten. Zum anderen kann sich "Zeithorizont" auch auf die Zeiträume der Erhebung bzw. Erfassung von Daten zu Gesundheitseffekten und Ressourcenverbrauch sowie auf die Zeiträume der Bewertung von Gesundheitseffekten und Ressourcenverbrauch beziehen. Dieser auf den Zeitraum der Datenerhebung und -bewertung bezogene Zeithorizont wurde von der Münchner Projektgruppe als die Bestimmung des Zeitpunkts der Datenerhebung ergänzende Information definiert und an den entsprechenden Stellen des Deskriptionskataloges aufgenommen.
- d) Angaben zur Methodik. Die Wahl der Analyseart hängt vom Studiengegenstand und dem Zweck der Studie ab. Zu unterscheiden sind (vgl. dazu auch oben Kapitel C.2.1.2) dabei:
- nicht vergleichende Studien zur ausschließlichen Ermittlung von Kosten, die von einer Krankheit (Krankheitskosten-Analyse) verursacht werden oder bei bestimmten Gesundheitsleistungen (Kosten-Analyse) anfallen;
 - vergleichende Studien zur Abschätzung der Kosten (Kosten-Kosten-/Kosten-Minimierungs-Analyse) bzw. der Kosten und Konsequenzen (Kosten-Konsequenzen-, Kosten-Effektivitäts-, Kosten-Nutzwert- bzw. Kosten-Nutzen-Analyse) verschiedener Maßnahmen.^{20,43,48,53,77,87,126,132,137,141,151} Kann beim Vergleich von alternativen Technologien implizit von der Gleichheit der

Konsequenzen ausgegangen werden, ist eine Kosten-Kosten-Analyse adäquat. Andernfalls ist zumindest eine Kosten-Konsequenzen-Analyse erforderlich.

Die Studien vergleichenden Charakters unterscheiden sich hinsichtlich der Bewertung der Konsequenzen, die bei Kosten-Konsequenzen-/Kosten-Effektivitäts-Analysen in naheliegenden natürlichen Einheiten, bei Kosten-Nutzwert-Analysen in Lebensqualität und Überlebenszeit aggregierenden Nutzwerten und bei Kosten-Nutzen-Analysen in Geldeinheiten erfolgt. Kosten-Nutzwert-Analysen werden oft auch als Kosten-Effektivitäts-Analysen bezeichnet.^{48,87}

- e) die Definition der Zielpopulation(en). Die Zielpopulation muss in Abhängigkeit vom Anwendungszweck der interessierenden Technologie(n) klar durch relevante demographische, indikationsbezogene, risikoprofilassoziierte oder auch geographische Angaben bestimmt werden.^{20,48,132} Auch kann es erforderlich sein, Untergruppen bezüglich der Effektivität, der Kosten oder der Präferenzen getrennt zu analysieren; auch diese müssen klar definiert werden.^{20,48}
- f) eine umfassende Beschreibung der Technologie(n) und der Vergleichstechnologie(n).^{20,43,48,53,76,77,87,126,132,138,141,151} In ökonomischen Evaluationen werden oft die Auswirkungen und Kosten von Verfahrensalternativen gegenübergestellt. Als Technologiealternative kann die interessierende Technik (als eng umgrenztes Verfahren) in unterschiedlicher Kombination mit anderen Techniken oder in unterschiedlicher Intensität bzw. Dosierung in Betracht gezogen werden, oder aber die interessierende Technik wird einer grundlegend andersartigen Technik als Vergleichstechnologie gegenübergestellt. Auf jeden Fall sollte der Vergleich vor dem Hintergrund einer möglichst vollständigen Beschreibung der medizinischen Handlungspfade vorgenommen werden. Im Hinblick auf die Vergleichbarkeit der Analysen oder die Übertragbarkeit der Ergebnisse ist es dabei auch wichtig, nähere Angaben bezüglich des Materialeinsatzes, des Personalbedarfs und der von der Anwendung der Technologie(n) und Vergleichstechnologie(n) tangierten Versorgungsstrukturen zu haben.

Eng verbunden mit dem Vergleich von Technologie- und Referenzalternativen ist das Konzept der Marginal- bzw. inkrementellen Betrachtung (vgl. dazu auch oben Kapitel C.2.1.1). Im Gegensatz zur Analyse von durchschnittlichen Konsequenzen und Kosten (z.B. Kosten pro diagnostiziertem Karzinom oder pro Operation, Lebensqualitätsgewinn pro Patient) werden bei der Marginalbetrachtung

zusätzlicher Nutzen und zusätzliche Kosten pro zusätzlich erbrachter Intensitäts- bzw. Dosierungseinheit einer Technik erfasst.^{20,43} Bei der inkrementellen Analyse ergeben sich Zusatznutzen und Zusatzkosten aus dem Vergleich der interessierenden Technologie mit der nächst effektiveren oder nächst teureren Referenzalternative^{20,43,48,87,126,132,141,151} Die Unterscheidung von marginaler und inkrementeller Betrachtung ist nicht einheitlich. Meist werden beide Vergleichsarten entweder unter inkrementeller Analyse^{48,87,132,141,151} oder unter Marginalbetrachtung^{77,126} zusammengefasst.

Gesundheitseffekte (gesundheitsbezogenes Outcome)

Bei der Datenerhebung sind hinsichtlich des zeitlichen Bezugs retrospektiv und prospektiv erhobene Daten zu unterscheiden. Daten zu klinischen Gesundheitseffekten und lebensqualitätsbezogenen Bewertungen - wie auch Daten zum Ressourcenverbrauch und dessen monetäre Bewertung - können innerhalb der gleichen Primärerhebung erfasst oder aus verschiedenen Primär- und Sekundärquellen aggregiert worden sein.^{48,126,132} Weiterhin ist die Entwicklung von Modellen von Vorteil, wenn bezüglich Gesundheitseffekten und Ressourcenverbrauch keine oder nur wenige Daten verfügbar sind (Generierung von Daten), Daten extrapoliert werden müssen oder Übertragungen der Resultate in andere Bereiche bzw. deren Anpassung an neue Entscheidungskontexte (beispielsweise unterschiedliche Behandlungsmethoden, ein anderes Land oder unterschiedliche Populationen) notwendig sind.^{48,126,132}

Durch den Einsatz von Technologien bzw. Referenzalternativen werden unterschiedliche Ereignispfade (event pathways) vorgegeben.^{48,132} Die Entwicklung dieser Pfade ist abhängig von den Eintrittswahrscheinlichkeiten sich gegenseitig bedingender Gesundheitszustände und medizinischer Handlungsreaktionen (vgl. dazu auch oben Kapitel C.2.1.4). Die korrekte Ermittlung der Effektivität und Kosten einer Technologie bzw. Referenzalternative setzt eine vollständige Beschreibung der durch die Technologieanwendung bedingten Ereignispfade und Ereigniswahrscheinlichkeiten voraus. Demnach ist es wesentlich, in einer Analyse alle relevanten Geschehnisse, die Veränderungen des Gesundheitszustandes verursachen oder mit einem Verbrauch an Ressourcen einhergehen, zu berücksichtigen. Die unterschiedlichsten medizinischen und epidemiologischen Outcome-Parameter, wie gewonnene symptomfreie Tage, gewonnene Arbeitstage, Anzahl identifizierter Tumore, Veränderungen klinischer Parameter, krankheitsfreie Überlebenszeit oder

auch zusätzliche Lebensjahre, können die Effektivität einer Technologie ausdrücken.^{20,48,87,126,132} Unterschieden werden kann dabei zwischen finalen bzw. globalen Outcome-Parametern (beispielsweise gewonnene Lebensjahre oder, falls eine Bewertung erfolgt, gesundheitsbezogene Lebensqualität) und intermediären Outcome-Parametern, wie etwa der Anzahl erfolgreich behandelter Patienten.^{48,126} Welche Outcome-Parameter im konkreten Fall angemessen sind, leitet sich aus den medizinischen Zielsetzungen oder den gesundheitspolitischen und ökonomischen Vorgaben der Anwendung einer Technologie ab.

Datenquellen für Outcome-Parameter können Primärdaten, Sekundärprodukte oder Modellierungen sein, die im Hinblick auf ihre Verwendung in gesundheitsökonomischen Studien unterschiedliche Vor- und Nachteile aufweisen.^{48,126,132} Auf jeden Fall sollten alle in einer gesundheitsökonomischen Evaluation verwendeten Datenquellen genau beschrieben sein, so dass sie hinsichtlich ihrer Eignung und ihre Validität bewertet werden können.^{48,77,126,132,138,141} Für die Beurteilung der methodischen Qualität von randomisierten kontrollierten Studien, Kohortenstudien, Fall-Kontroll-Studien, Querschnittstudien bzw. Fallserien und Diagnosestudien als Primärdatenquellen sowie von systematischen (narrativen) Reviews, Meta-Analysen, (echten) gepoolten Analysen und HTA-Berichten als Sekundärdatenquellen wurden als in den Deskriptionskatalog einzubindende Instrumente die Checklisten der HTA-Projektgruppen (German Scientific Working Group - Technology Assessment for Health Care¹⁴) herangezogen. Diese sind im Rahmen des Projekts "Aufbau einer Datenbasis zur Evaluation medizinischer Verfahren und Technologien in Deutschland" zur systematischen Beurteilung der methodischen Qualität von Effektivitätsanalysen zusammengestellt worden. Für Screening-Studien wie auch für entscheidungsanalytische Modellierungsstudien bestehen noch keine solchen Checklisten. Im Falle von Studien zu diagnostischen Aspekten beim Screening kann behelfsmäßig die Checkliste für Diagnosestudien herangezogen werden.

In Abhängigkeit von der Analyseart können Outcome-Parameter bewertet werden (vgl. dazu auch oben Kapitel C.2.1.1 und C.2.1.2): In Kosten-Nutzwert-Analysen werden die Gesundheitszustände in lebensqualitätsbezogenen Nutzeinheiten bewertet, die über die Zeit zu Nutzwerten zusammengefasst werden. Damit können Aufwendungen im Gesundheitsbereich untereinander verglichen werden.^{48,87,126,132} Kosten-Nutzen-Analysen bewerten alle Auswirkungen einer Technologie in Geldeinheiten. Damit können Aufwendungen im Gesundheitsbereich auch Investitionen in anderen Bereichen gegenübergestellt werden.^{48,87,126,132} Da klassische Outcome-Parameter, wie beispielsweise die Mortalität, die Qualität der

medizinischen Versorgung nur unzureichend widerspiegeln, gewinnt die Lebensqualität bei der Outcome-Messung im Gesundheitswesen immer stärker an Bedeutung. Theoretisch sind zwei verschiedene Verfahren zur Erfassung von gesundheitsbezogener Lebensqualität denkbar: Von einer psychometrischen Skalierung wird gesprochen, wenn Individuen die Lebensqualitätsbewertung anhand einer Klasseneinordnung oder einer einfachen numerischen Bewertung vornehmen. Vom Patienten selbst auszufüllende, standardisierte und psychometrisch geprüfte Fragebögen haben dabei große Bedeutung erlangt. Diese können:

- nach dem Grad der Aggregation der Outcome-Daten in Profil- und Indexinstrumente
- nach dem Krankheitsbezug in krankheitsspezifische und generische (krankheitsübergreifende) Lebensqualitätsmessinstrumente

klassifiziert werden.^{20,48,58,87,126} Das zweite Verfahren zur Lebensqualitätsbewertung basiert auf der Nutzenmessung. Die wichtigsten hierbei zur Verfügung stehenden Methoden sind das rating scale-Verfahren, das standard gamble-Verfahren und die time-trade-off-Methode.^{20,48,87,126,132,142} Den Gold-Standard bei der Lebensqualitätsmessung gibt es nicht, vielmehr ist die Auswahl des Messinstrumentariums von der jeweiligen Fragestellung abhängig. Damit die Ergebnisse des Messvorgangs auch verwertbar sind, müssen die verschiedenen Messinstrumente valide, sensitiv, zuverlässig, konsistent, verständlich, relevant und praktikabel sein.^{48,87,126,132} Daten zu Lebensqualitätsbewertungen können durch eigenes, direktes Messen, durch Entnahme von bereits existierenden Werten aus der Literatur oder durch Plausibilitätsüberlegungen (judgements) gewonnen werden.^{48,87}

Das Problem von Kosten-Nutzen-Analysen liegt in der Monetarisierung von Gesundheitszuständen. Oftmals werden Versorgungskosten, teilweise auch noch Produktivitätsausfallkosten von Technologie- bzw. Referenzalternativen erfasst. Eine monetäre Bewertung von Gesundheitszuständen über die Zahlungs- oder Akzeptanzbereitschaft bzw. über den Humankapitalansatz ist jedoch mit Ausnahme von wenigen, versuchsweisen Ansätzen nicht zu finden. Grundsätzlich wäre es möglich, die Zahlungs- oder Akzeptanzbereitschaft durch Beobachtung des Verhaltens von Individuen und daraus gezogenen Rückschlüssen auf ihre Präferenzen zu ermitteln, wobei von aufgedeckten Präferenzen (revealed preferences) gesprochen wird. Oder aber es wird in Befragungen bzw. Experimenten die monetäre Bewertung des Mortalitätsrisikos durch Abwägung zwischen Kosten

und Mortalitätsrisiko ermittelt.^{48,87} Standardisierte Befragungskonzepte oder experimentelle Messverfahren liegen allerdings noch nicht vor.

Kosten (ökonomisches Outcome)

Unter Kosten (bzw. Nutzen) wird der in Geldeinheiten bewertete Ressourcenverbrauch, der mit der Anwendung einer Technologie bzw. Referenzalternative verbunden ist (bzw. vermieden werden kann) verstanden.^{48,87,126,132} Der Ressourcenverbrauch wird in physischen Einheiten gemessen, etwa als Dauer der ärztlichen Beratung. Zur Berechnung der Kosten müssen die Kategorien des Ressourcenverbrauchs identifiziert, der Verbrauch gemessen und monetär bewertet werden. Die gewählte Perspektive, der festgelegte Zeithorizont und der berücksichtigte Ereignispfad einer Evaluation bestimmen im wesentlichen, welche Kosten definiert, gemessen und bewertet werden.^{48,53,87,126,132,141} Zur Prüfung der methodischen Qualität von Kostenmessungen gehören auf jeden Fall Kontrollen von Doppelzählungen und Lücken oder Kontrollen der Konsistenz bei Kombination mehrerer Datenquellen.

Direkte Kosten. Zur Identifikation und Beschreibung des Mengengerüsts der Kosten, die direkt durch die Anwendung einer Technologie bzw. Referenzalternative bedingt werden, gibt es verschiedene Konzepte. Hauptsächlich kann zwischen medizinischen und nicht-medizinischen Kosten unterschieden werden: Zu den medizinischen Kosten werden Kosten verschiedener Versorgungssektoren und -arten gezählt, wie Kosten für diagnostische Tests, Arznei-, Heil- und Hilfsmittel, für Versorgungsgüter, Pflegepersonal, medizintechnische Ausstattung oder Kosten für Gebäude und Räumlichkeiten. Auch die Gemeinkosten (Overhead-Kosten für Verwaltung, Leitung, Abschreibung etc.) gehören in diese Kostenkategorie.^{20,48,77,87,126,132} Zu den nicht-medizinischen Kosten werden Eigenleistungen des Patienten wie Zusatzkäufe, Transportkosten und Zeitkosten, aber auch alle Formen der Unterstützung durch Familienangehörige, Freunde und ehrenamtliche Helfer, durch professionelle pflegerische bzw. Sozialdienste oder Haushaltshilfen gerechnet.^{20,48,76,77,87,126,132}

Ein anderes Konzept, das auf der Kostenrechnung in Betrieben beruht und für entsprechende Messungen geeignet ist, geht vom Aufwand der Leistungserstellung aus und ermittelt die Kostenarten. Ihre Hauptkategorien sind die Personal- und die Sachkosten. Die Ermittlung des Mengengerüsts ist konzeptionell einfach, birgt aber das Problem, dass bei der Herstellung mehrerer Güter oder Dienstleistungen in

einem Betrieb die Kostenarten zu den zu evaluierenden Leistungen hinzugerechnet werden müssen. Wird nicht vom Erstellungsaufwand, sondern von erstellten Produkten und Zwischenprodukten ausgegangen, wird eine geringere Differenzierung der Kosten erreicht. Die Verwendung von Produkten zur Erstellung des Mengengerüsts vereinfacht die Kostenrechnung besonders dann, wenn die Bewertungen, also Preise oder Kostenschätzungen, für die einzelnen Produkte bereits vorliegen, etwa in der Form von Gebührenordnungen. Entsprechend können Versorgungsmengen definiert werden über Versorgungskontakte, die Versorgungszeit, Einzelleistungen, Leistungskomplexe sowie über Falltypen. In der Praxis finden sich auch Kombinationen der Mengenbestimmung nach Erstellungsaufwand und nach Produkten, welche aufgrund möglicher Überschneidungen oder Auslassungen konzeptionell nicht unproblematisch sind. Schließlich setzt bezüglich der direkten Kosten eine Unterscheidung bei der Produktionsbindung der Kosten an: Variable Kosten sind Kosten, die mit der erstellten Leistungsmenge variieren, Fixkosten hingegen bleiben auf kurze Sicht konstant und sind erst auf lange Sicht oder beim Unter- und Überschreiten von Kapazitätsgrenzen variabel.^{48,87}

Bei der Bewertung des Mengengerüsts von verbrauchten Ressourcen spiegeln die Opportunitätskosten den Wert der besten anderweitigen Verwendung der Ressourcen und somit deren (gesellschaftliche) Knappheit wider.^{48,87,126,132} In einem funktionierenden Wettbewerbsmarkt geben die Marktpreise diesen Wert mehr oder weniger wieder. Wo wegen staatlicher Regulierung die Opportunitätskosten für einen Großteil des Ressourcenverbrauchs nicht über Marktpreise bestimmt werden können, müssen Schattenpreise (administrative Preise, betriebliche Selbstkosten, pauschale Durchschnittskosten pro Leistungseinheit) ermittelt werden.^{48,87,132} Die Wahl der Methode zur Messung und Bewertung des Ressourcenverbrauchs ist abhängig von der Zielsetzung der Studie, der Verfügbarkeit und Zugänglichkeit von Daten und den Ressourcenbeschränkungen der Evaluationsstudie selbst.^{87,126}

Produktivitätsausfallkosten. Bei den Produktivitätsausfallkosten werden Ressourcenveränderungen erfasst und bewertet, die sich (meist aus gesamtgesellschaftlicher Sicht) aus Krankheit und vorzeitigem Tod ergeben.^{48,53,76,87,126,132,141} Morbiditätskosten sind solche Kosten, die mit dem krankheitsbedingten vollständigen oder partiellen Verlust der Arbeitsfähigkeit bzw. der Beschäftigung mit Freizeitaktivitäten in Zusammenhang stehen, wie beispielsweise Zeit für Erholung und Konvaleszenz. Mortalitätskosten werden durch Veränderungen der Lebenserwartung bzw. durch vorzeitigen Tod bedingt.⁴⁸ Auch

Produktivitätsausfallkosten, die im Rahmen der Primär-, Sekundär- und Tertiärprävention anfallen, sind dieser Kategorie zuzuordnen.

Zur Berechnung der Produktivitätsausfallkosten kann nach dem Humankapitalansatz oder dem Friktionskostenansatz vorgegangen worden sein.^{48,78,87,126} Beim Humankapitalansatz wird unterstellt, dass Gesundheitsausgaben aus gesamtwirtschaftlicher Sicht immer auch Investitionen in die Erhaltung der Berufsausübungsfähigkeit der Patienten sind. Demnach sind die indirekten Kosten einer Krankheit oder vorzeitigen Todes gerade so groß wie das weggefallene Produktionspotential, wobei oft auf Arbeitsmarktlöhne zurückgegriffen wird. Die Bewertung der Produktivitätsausfallkosten für diejenigen, die nicht auf dem Arbeitsmarkt aktiv sind, wird dadurch erschwert oder verhindert. Mit dem Friktionskostenansatz soll die Überschätzung von Produktivitätsverlusten vermieden werden: Pro Patient und Krankheitsperiode wird ein Produktivitätsverlust höchstens für die Dauer der durchschnittlichen Vakanz unbesetzter Stellen angenommen.

Bei der Beurteilung von Kosten-Nutzwert-Analysen ist darauf zu achten, dass Morbiditäts- wie auch Mortalitätskosten nicht zwingend auf der Kostenseite aufgeführt sein müssen. Vielmehr können sie auch bei der Lebensqualitätsmessung berücksichtigt worden sein, womit sie im Nenner des Kosten-Nutzwert-Quotienten enthalten wären.⁴⁸

Intangible Kosten. Mit intangiblen Kosten werden Folgen von Krankheit oder medizinischen Interventionen wie Schmerz, Freude oder physische Beschränkung bezeichnet. Sie sind monetär schwer zu bewerten. Manche Autoren empfehlen, die Arten der intangiblen Kosten aufzuzählen. Sie können jedoch ebenfalls auch auf der Outcome-Seite bei der Lebensqualitätsmessung erfasst werden.^{48,87,126,132}

Folgekosten gewonnener Lebensjahre. Der Einbezug von direkten und indirekten Folgekosten gewonnener Lebensjahre, die sich aus der Beeinflussung des Gesundheitszustandes durch den Einsatz einer Technologie ergeben, wird kontrovers diskutiert.^{48,87}

Kosten können parallel zu klinischen Studien erhoben worden sein, auf Sekundärdaten aus verschiedenen Quellen basieren oder durch Modellierungen generiert worden sein. Kosten aus verschiedenen Perioden sollten inflations- bzw. deflationsbereinigt in den realen Preisen eines Basis-/Referenzjahres ausgedrückt werden.^{20,48,87,126,132}

Diskontierung

Zukünftige Kosten wie auch in der Zukunft auftretende Outcome-Parameter sollten diskontiert werden.^{20,48,53,76,87,126,132,138,141,151} Der Einzelne wie auch die Gesellschaft hat im allgemeinen eine sog. positive Zeitpräferenz: Erwünschte Konsequenzen - beispielsweise positive Auswirkungen auf den Gesundheitszustand bei der Anwendung einer Technologie - sollten zum frühest möglichen Zeitpunkt verfügbar sein; während für unerwünschte Konsequenzen - beispielsweise Kosten - eine Verlagerung möglichst weit in die Zukunft bevorzugt wird. In der Zukunft auftretende Outcomes werden diskontiert, weil sie durch ihre verzögerte Verfügbarkeit an Wert verlieren. Zukünftige Kosten werden diskontiert, da bis zu ihrem verzögerten Anfallen Kapital anderweitig genutzt werden kann. Die Wahl der Diskontrate hat in Abhängigkeit von der Länge des betrachteten Zeitraumes einen zunehmenden Einfluss auf das Gesamt-Outcome und die Gesamtkosten. In einigen Ländern, in denen explizite Richtlinien für die Durchführung von gesundheitsökonomischen Analysen vorliegen, wird der Diskontierungssatz verbindlich vorgegeben (z.B. 5% in Australien und Kanada).^{20,126}

Ergebnispräsentation

Gesundheitsökonomische Analysen sollten detailliert, klar und transparent sein. Resultate sollten am besten entsprechend den einzelnen Schritten der Analyse dargestellt worden sein, wobei die Entwicklung hin zu aggregierten Werten und Bewertungen von Outcome-Parametern und verbrauchten Ressourcen nachvollziehbar sein sollte.^{20,48,77}

Durch Subgruppenanalysen können Unterschiede zwischen Untergruppen einer gegebenen Zielpopulation erfasst werden.^{20,48} Die differenzierte Betrachtung einer in bezug auf die Effektivität oder Kosten einer Technologie/Referenzalternative bzw. hinsichtlich der Gesundheitszustandspräferenzen heterogenen Population kann im Zusammenhang mit Risikoselektion und Mittelzuweisung entscheidungsrelevante Information liefern.

Behandlung von Unsicherheiten

Die meisten Daten, die in eine ökonomische Evaluationsstudie einfließen, müssen als unsicher gelten. Sensitivitätsanalysen sind ein wesentliches Instrument zur

Offenlegung des Einflusses unsicherer Daten bzw. Annahmen auf das Endergebnis und geben an, wie robust dieses ist.^{20,43,48,53,76,77,87,126,132,138,141,151} Hierbei werden durch eine Variation der Annahmen alternative Gesamtergebnisse ermittelt. In welchem Ausmaß die unsicheren Werte variiert werden sollen, steht nicht eindeutig fest. Es kann unter Konstanthaltung der übrigen Daten immer nur eine Variable verändert werden (univariable oder Ein-Weg-Sensitivitätsanalyse) oder mehrere Variablen werden gleichzeitig verändert (multivariable oder Mehr-Weg-Sensitivitätsanalyse). Auch statistische Methoden können angewandt werden, um Unsicherheiten zu quantifizieren.^{48,87,98,132} Hierbei wird die Robustheit des Gesamtergebnisses durch gleichzeitige Variation mehrerer Parameter untersucht und mittels Konfidenzintervallen bzw. Wahrscheinlichkeitsverteilungen dargestellt.

D.4 Anhang 4: Standardisierte Berichte

Vorbemerkungen zu den standardisierten Berichten:

Im folgenden werden unter Punkt "1-2 Fragestellung" jedes standardisierten Berichts die untersuchten Screening-Strategien in Kurzschreibweise aufgeführt. Diese sei an einem Beispiel erläutert.

Beispiel:

Screening-Strategie: PSA+DRU \rightarrow [$>4\text{ng/ml}$ und/oder \oplus] \rightarrow TRUS+Biopsie

Die Screening-Strategie beginnt mit PSA+DRUS. Dies bedeutet, dass in einer ersten Screening-Stufe an alle Patienten ein PSA-Test und auch ein TRUS durchgeführt wird. Der Pfeil von PSA+DRU nach TRUS+Biopsie wird durch eine eckige Klammer unterbrochen. Der Inhalt der eckigen Klammer bestimmt, wann im Anschluss an die erste Screening-Stufe PSA+DRU die zweite Screening-Stufe TRUS+Biopsie durchgeführt wird. Im vorliegenden Beispiel geschieht dies dann und nur dann, wenn der PSA-Wert $>4\text{ng/ml}$ ist und/oder der Befund der DRU positiv ist.